Part IV: Weakly-Supervised Information Extraction

Automated Mining of Structured Knowledge from Text in the Era of Large Language Models

Yunyi Zhang, Ming Zhong, Siru Ouyang, Yizhu Jiao, Sizhe Zhou, Linyi Ding, Jiawei Han

Computer Science, University of Illinois Urbana-Champaign

KDD 2024 Tutorial, Aug 25, 2024

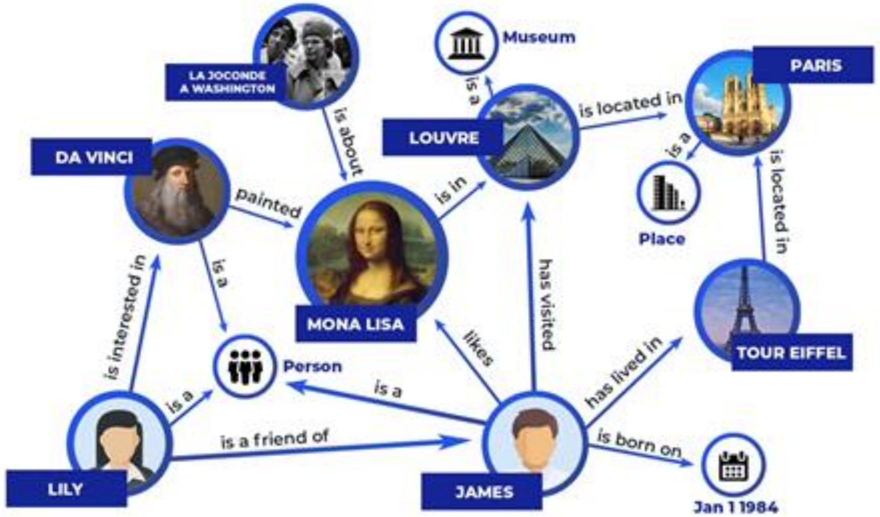
Tutorial Website:

Outline

- Overview
- Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring

Overview of Knowledge Graphs

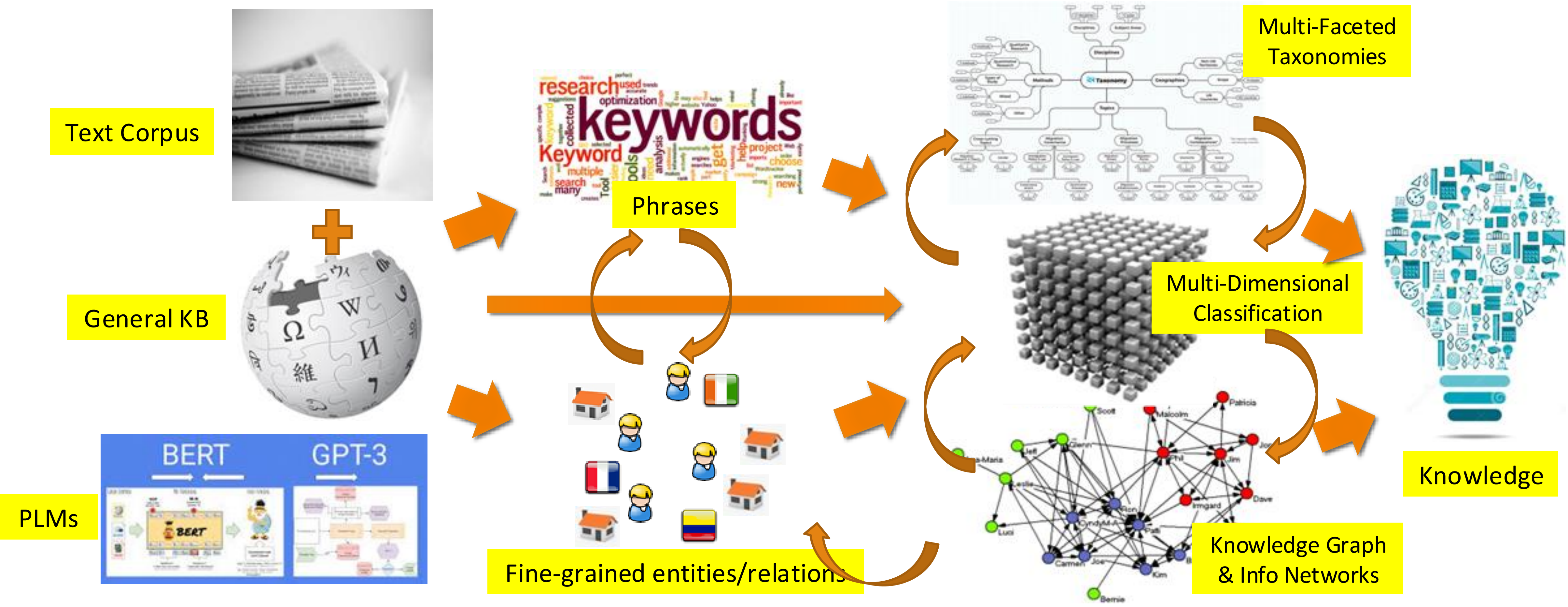
- General knowledge graphs
  - Ex. Wikidata, DBPedia, Freebase, Yago …
  - Cover a wide range of topics and domains for broad applicability
- Domain-specific knowledge graphs
  - Ex. UMLS (medical), PubChem (chemical)…
  - Cover specialized knowledge for domain-specific need
- Theme-specific knowledge graphs
  - Ex. Ukraine War, EV battery or LLM
  - Cover highly detailed information on a particular theme for specialized applications.



Ack. Figures are from Google images

Knowledge Graph Construction

- It can be comprehensive task that may involve several subtasks like phrase mining, entity typing, relation extraction…

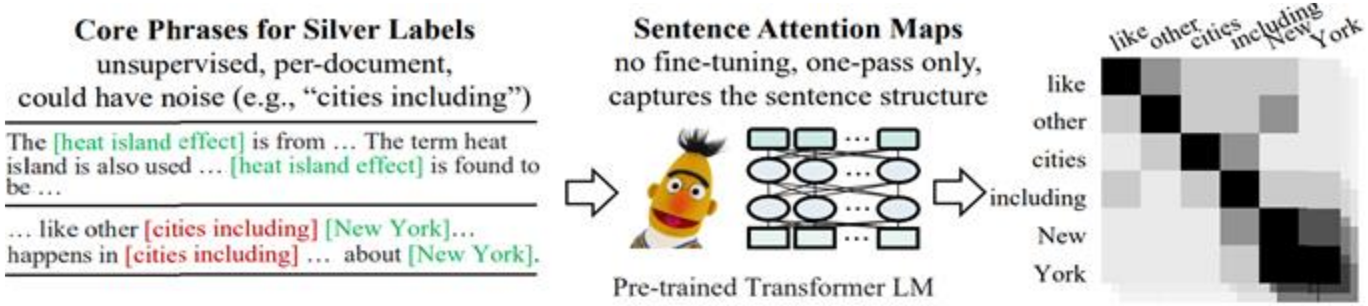


Phrase Mining Tools

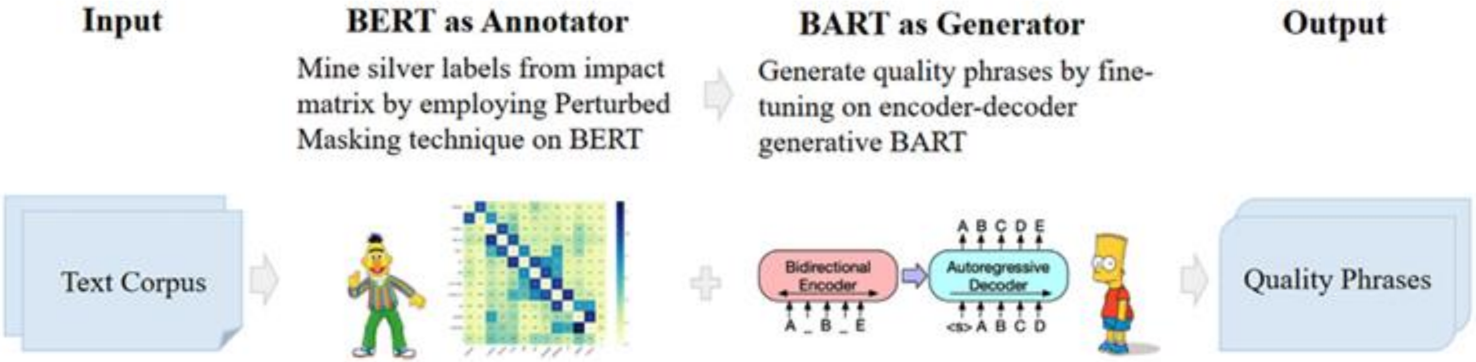
- **Pre-trained off-the-shelf toolkits**
  - PKE: chunking model based on supervised POS tagging model from NLTK
  - Spacy: industrial-grade library based on pre-trained phrase chunking model
  - StanfordCoreNLP: NLP library pre-trained on dependency parsing data
- **KB-based distantly-supervised methods**
  - AutoPhrase: uses distantly supervision for silver labels generation and statistic-based binary phrase classifier with POS-guided segmentation

Phrase Mining Tools

- **Unsupervised methods**
  - ToPMine: builds upon statistical features to discover arbitrary length phrases
  - UCPhrase: uses silver label generation with attention map-based span prediction



  - LMPhrase: Constructs silver labels from annotator (BERT) and fine-tunes seq2seq generator (BART) for phrase generation



Quantitative Evaluation

Evaluation results (%) of sentence-level phrase tagging for all compared methods. The best results are in bold, and the second best are underlined.

Model type	Model name	KP20k			KPTimes		
		Precision	Recall	F1	Precision	Recall	F1
Off-the-shelf toolkit	PKE	54.1	63.9	58.6	56.1	62.2	59.0
	Spacy	56.3	68.7	61.9	61.9	62.9	62.4
	StanfordNLP	48.3	60.7	53.8	56.9	60.3	58.6
Distantly supervised	AutoPhrase	55.2	45.2	49.7	44.2	47.7	45.9
	Wiki+RoBERTa	58.1	64.2	61.0	60.9	65.6	63.2
Unsupervised	TopMine	39.8	41.4	40.6	32.0	36.3	34.0
	UCPhrase	<u>69.9</u>	<u>78.3</u>	<u>73.9</u>	<u>69.1</u>	<u>78.9</u>	<u>73.5</u>
	LMPhrase (Ours)	**71.9**	**79.2**	**75.3**	**73.2**	**79.2**	**76.1**

Gu, Xiaotao, et al., "Ucphrase: Unsupervised context-aware quality phrase tagging." KDD'21

Zhang, Zhihao, et al., "Language models as an Annotator: Unsupervised context-aware quality phrase generation." KBS'24

Outline

- Overview
- Entity Typing
  - Knowledge-based Entity Typing
  - Ontology-guided Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring

Motivation

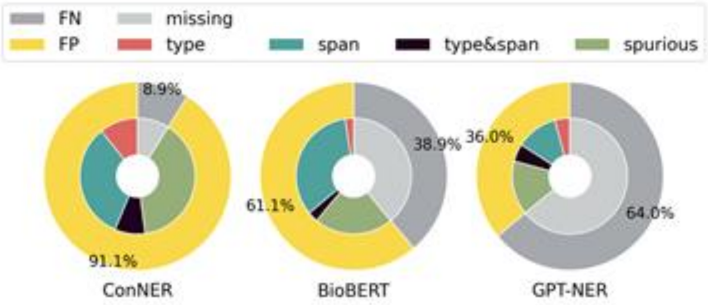
- Entity typing is a fundamental task in text mining with a wide spectrum of applications
  - question answering
  - knowledge base construction
  - dialog systems
  - …
- Deep neural models have achieved enormous success for entity typing
- However, a common bottleneck of training deep learning models is the acquisition of abundant high-quality human annotations (every entity in the sequence needs to be labeled!)

VerifiNER: Verification-augmented

- **Fine-tuned models:** false positive errors are the majority
- **Prompting-based methods:** low precision

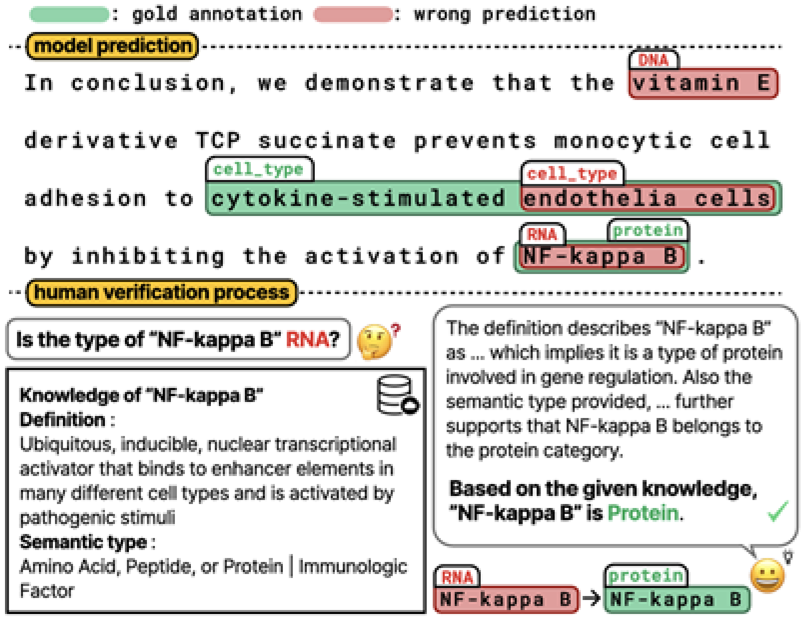
  → correcting FP cases from initial predictions
- Models resort to plausible yet in-correct predictions <u>with a small margin</u>

→ **verification** as post-hoc



Gold annotation: "… cell adhesion to cytokine-stimulated [endothelial cells]cell_type by .."

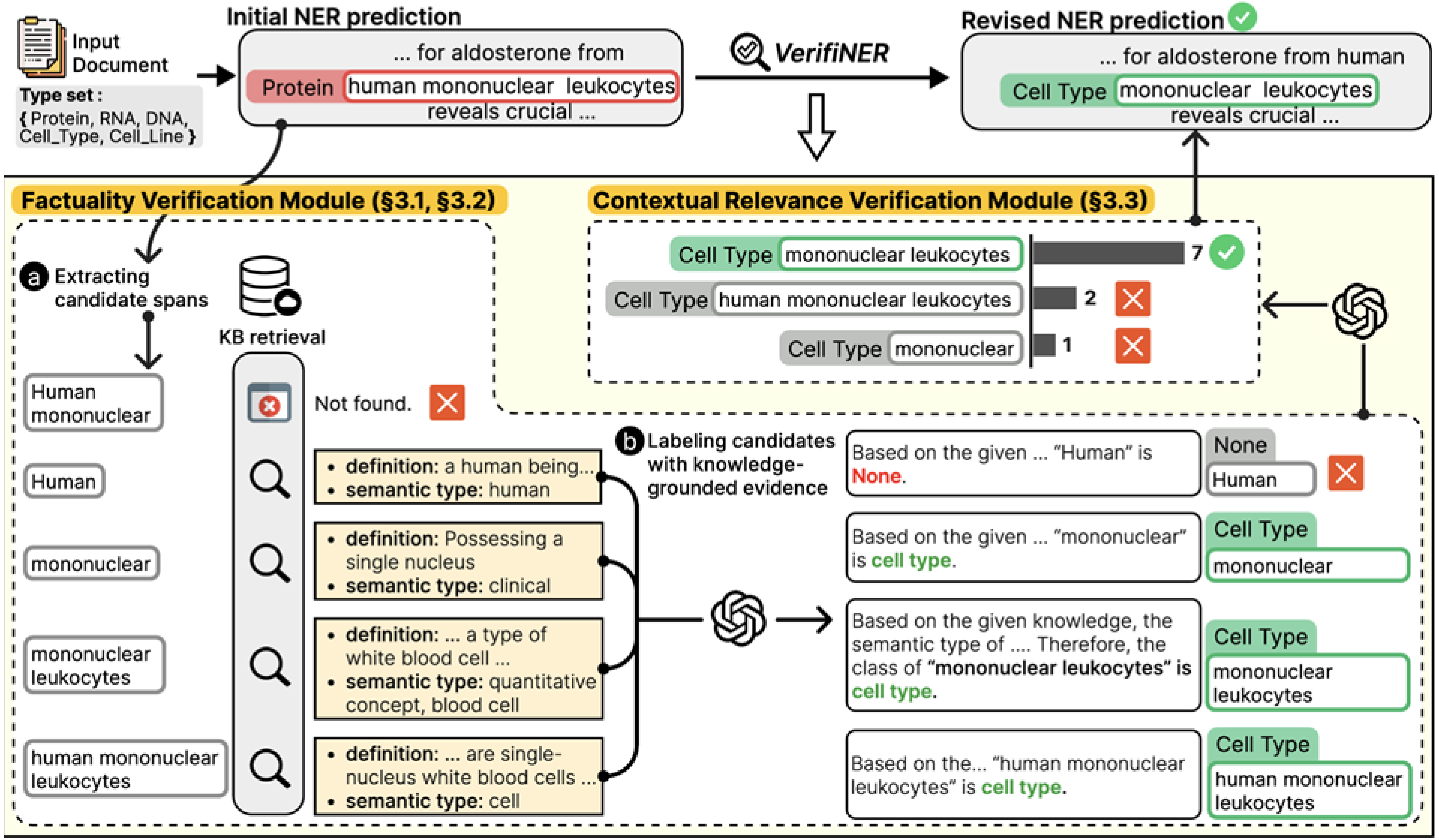
Category	Error type	Definition	Example
FP	Type	Wrong type is assigned to an entity	"… cell adhesion to cytokine-stimulated [endothelial cells] cell_line by .."
	Span	Predicted span partially overlaps with ground truth, but incorrect	"… cell adhesion to [cytokine-stimulated endothelial cells] cell_type by .."
	T&S	Both type and span are predicted incorrectly	"… cell adhesion to [cytokine-stimulated endothelial cells] cell_line by .."
	Spurious	A completely incorrect entity is predicted where gold annotation does not exist	"… [cell adhesion] cell_line to cytokine-stimulated endothelial cells by .."
FN	Missing	A gold annotation for entity exists but not predicted by a model	"… cell adhesion to cytokine-stimulated endothelial cells by .."



- External knowledge could be a valuable and effective source for verification.

Seoyeon Kim, et al., "VerifiNER: Verification-augmented NER via Knowledge-grounded Reasoning with Large Language Models", ACL'24

VerifiNER: An Overview



VerifiNER: Span & Type Factuality Verification

- Extract and verify all the candidate spans with
  - If the entity exist in the KB, extract all the related information from the KB
- Re-assign entity types
  - knowledge does not guarantee explicit implications to entity types
    - leverage the reasoning ability of LLMs

Knowledge-grounded Evidence

Entity candidate: mononulcear leukocytes
Source knowledge: [Definition] A white blood cell that lacks cytoplasmic granules with an affinity for specific biological stains
[Semantic type] Quantitative Concept, Blood Cell
Generated evidence: Based on the given knowledge, the definition of "mononulcear leukocytes" refers to a white blood cell that lacks cytoplasmic granules ... Also the semantic type of mononulcear leukocytes is associated with `cell_type`.

VerifiNER: Context Relevance Verification

- Select a final type based on the context
  - semantically relevant to the input context
  - knowledge-grounded evidence is aligned well with the context
- Sample multiple reasoning paths to reflect the related context
  - employ self-consistency to find the most plausible votes

Contextual Relevance Verification

Entity candidates:
mononuclear leukocytes, **human mononuclear leukocytes**, mononuclear

Reasoning Path 1:
Based on the provided evidence,"mononuclear leukocytes" refers to ... This aligns with the context, that it plays a crucial role in immune responses... Therefore, **"mononuclear leukocytes"** is the correct entity in this context as it provides both generality and comprehensiveness, accurately identifying the type of immune cells.

...

Reasoning Path N:
...the context is focused on human biology and immune responses. By specifying "human mononuclear leukocytes," it clarifies that the experiment pertains to leukocytes from humans ... Therefore, "human mononuclear leukocytes" is the correct entity in this context as it provides both specificity and avoidance of ambiguity.

Consistency Voting:
mononuclear leukocytes : 7 *(final selection)*
human mononuclear leukocytes : 2
mononuclear : 1

Main Results

Methods	GENIA				BC5CDR			
	P	R	F1	ΔF1	P	R	F1	ΔF1
GPT-NER (Wang et al., 2023a)	56.44	42.15	48.26	-	79.84	**47.48**	59.55	-
+ Manual Mapping	37.53	32.65	34.93	-13.33	51.82	36.98	43.16	-16.39
+ LLM-revision	52.97	**46.77**	49.68	+1.42	77.21	44.53	56.48	-3.07
+ LLM-revision w/ CoT	53.57	44.54	48.64	+0.38	76.49	44.91	56.59	-2.96
+ VERIFINER (Ours)	**72.37**	44.95	**55.46**	**+7.20**	**91.01**	46.92	**61.92**	**+2.37**
ConNER (Jeong and Kang, 2023)	74.13	**96.69**	83.92	-	84.90	**96.47**	90.32	-
+ Manual Mapping	43.62	94.50	59.69	-24.23	53.98	94.52	68.71	-21.61
+ LLM-revision	63.64	86.64	73.38	-10.54	80.35	93.07	86.25	-4.07
+ LLM-revision w/ CoT	64.85	86.92	74.28	-9.64	78.14	92.99	84.92	-5.40
+ VERIFINER (Ours)	**79.07**	91.82	**84.97**	**+1.05**	**94.77**	91.61	**93.16**	**+2.84**
BioBERT (Lee et al., 2019)	54.51	65.30	59.42	-	79.93	**95.98**	87.22	-
+ Manual Mapping	30.57	24.39	27.14	-32.28	38.65	65.78	48.69	-38.53
+ LLM-revision	52.63	65.01	58.17	-1.25	60.79	77.74	68.23	-18.99
+ LLM-revision w/ CoT	52.21	63.49	57.30	-2.12	59.43	78.66	67.71	-19.51
+ VERIFINER (Ours)	**77.45**	**67.75**	**72.31**	**+12.89**	**94.02**	91.17	**92.57**	**+5.35**

- Consistently outperforms baseline methods
- intermediate reasoning process is important to bridge the gap between the retrieved knowledge and model predictions
- Recall is a little lower

Source → Target	BC5CDR → GENIA			GENIA → BC5CDR		
	P	R	F1	P	R	F1
GPT-NER	56.44	42.15	48.26	79.84	47.48	59.55
ConNER	N/A	N/A	N/A	N/A	N/A	N/A
+ VERIFINER	**58.15**	**77.42**	**66.42**	**76.74**	**57.42**	**65.69**
BioBERT	N/A	N/A	N/A	N/A	N/A	N/A
+ VERIFINER	**66.49**	**87.25**	**75.47**	**77.64**	**71.17**	**74.27**

Source → Target	GENIA′ → GENIA			BC5CDR′ → BC5CDR		
	P	R	F	P	R	F
GPT-NER	56.44	42.15	48.26	79.84	47.48	59.55
ConNER	69.97	94.10	80.26	81.46	89.47	85.28
+ VERIFINER	**74.16**	90.03	**81.48**	**94.28**	85.42	**89.63**
BioBERT	35.36	64.50	45.68	61.75	69.95	65.59
+ VERIFINER	**78.91**	**69.68**	**74.01**	**94.16**	**71.65**	**81.38**

- VerifiNER can further improve the performance of fine-tuned models

Outline

- Overview
- Entity Typing
  - Knowledge-based Entity Typing
  - Ontology-guided Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring

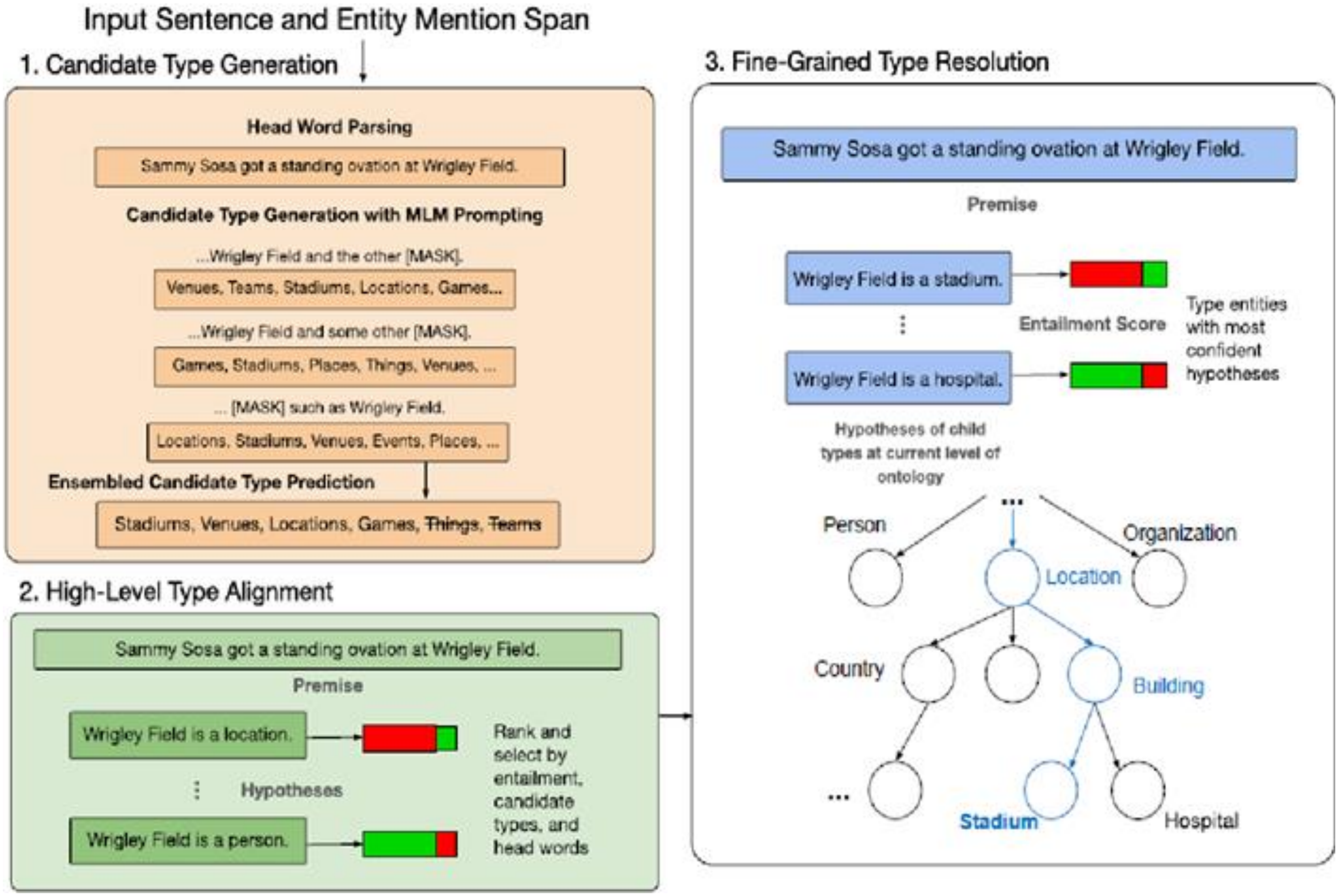
OntoType: Ontology-guided Typing

- Zero-shot entity typing: Assigns fine-grained semantic types to entities without any annotations
  - Ex. *Sammy Sosa* [Person/Player] got a standing ovation at *Wrigley Field* [Location/Building/Stadium]
- Challenges of weak supervision based on masked language model (MLM) prompting
  - A prompt generates a set of tokens, some likely vague or inaccurate, leading to erroneous typing
  - Not incorporate the rich structural information in a given, fine-grained type ontology
- OntoType: Ontology-guided, Annotation-Free, Fine-Grained Entity Typing
  - Ensemble multiple MLM prompting results to generate a set of type candidates
  - Progressively refine type resolution, from coarse to fine, following the type ontology, under the local context with a natural language inference model

Tanay Komarlu, et al., "OntoType: Ontology-Guided and Pre-Trained Language Model Assisted Fine-Grained Entity Typing", KDD'24

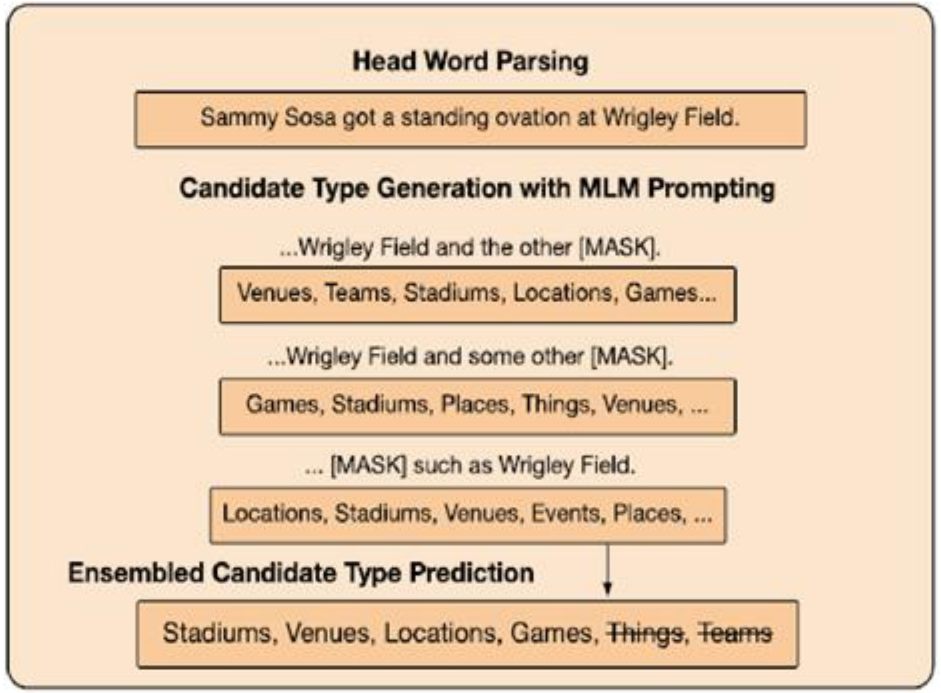
Overall Framework of OntoType : Three Steps

- Candidate type generation
  - Candidate type generation with multiple MLM prompting
  - Ensembled candidate type prediction
  - Ex. Stadium, venue, location, games, ~~things, teams~~
- High-level type alignment by entailment (local context + NLI)
- Progressively refine type resolution, from coarse to fine, following the type ontology
- Type ontology used at every step



OntoType: Step 1 - Candidate Type Generation

- Head Word Parsing
  - Mention's head word in the input text is often the cue that explicitly matches a mention to its type
  - Ex. "<u>Governor</u> Arnold Schwarzenegger gives a speech …"
  - Use the Stanford Dependency Parser to extract head word
  - Leverage the head words of the input entity to select an initial context-sensitive coarse-grained type
- Ensembled MLM Prompting
  - Leverage a BERT MLM and Hearst patterns to generate candidate types for the target mentions
  - Ensemble n patterns to generate the best candidate types
  - Consolidated candidates are generated by a majority of Hearst patterns
    - Ex. For e_1, "Stadiums, Venues, Locations, Games" retain, but the noisy types "Things" and "Teams" are removed

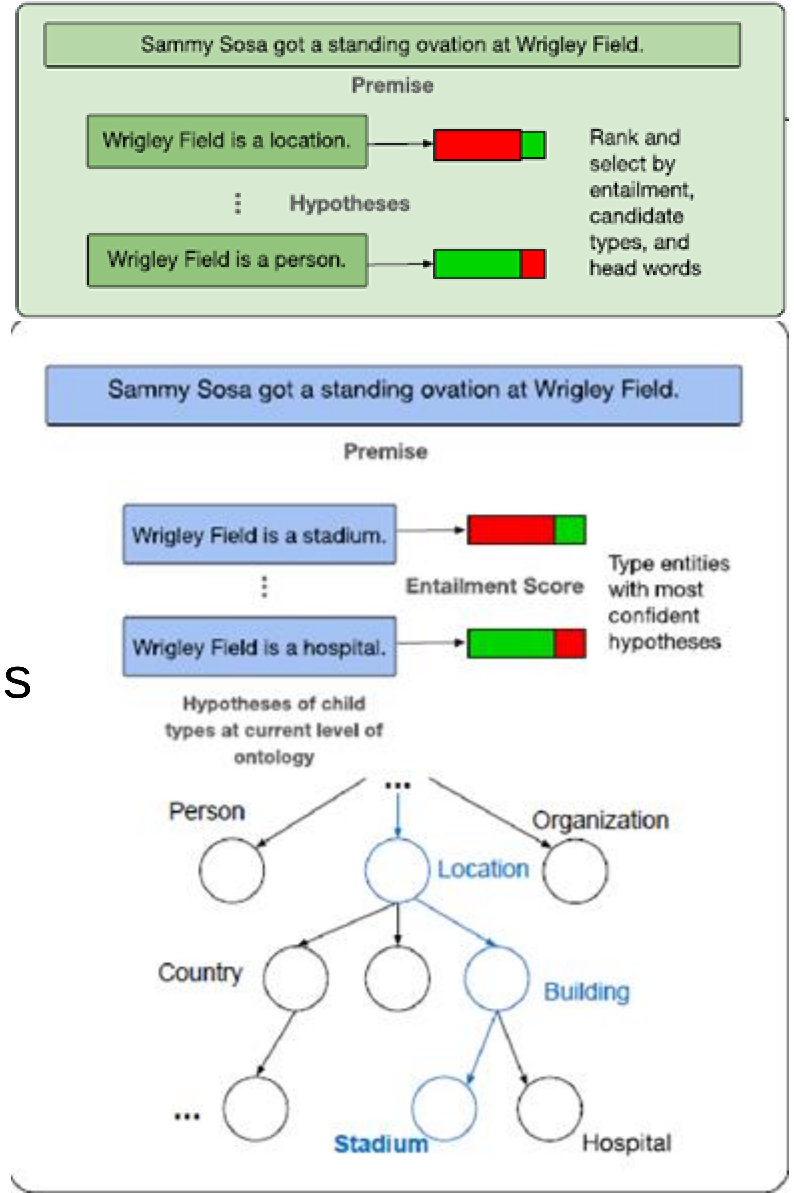


Four Hearst Patterns give the highest quality hypernyms with simple type mapping on the OntoNotes dataset

Hearst Pattern	Prec	Rec	F1
[MASK] such as	53.3	72.4	61.4
such [MASK] as	47.9	68.7	56.5
and some other [MASK]	48.8	66.6	56.4
and the other [MASK]	47.6	68.3	56.1

OntoType: Steps 2 & 3- High-Level Type Resolution & Progressive Type Refinement

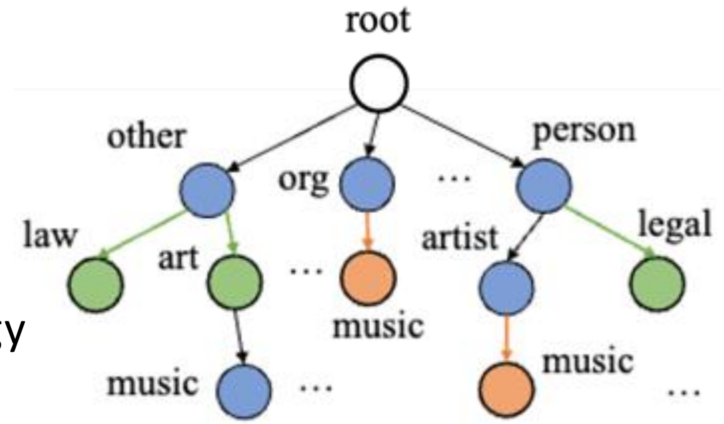
- High-level type alignment by entailment
  - Align generated candidate types to several high-level types in the type ontology by Word2Vec+ cosine similarity
  - Then select the most accurate high-level types with a pre-trained entailment language model (NLI)
- Progressively refine type resolution, from coarse to fine, following the type ontology
  - Ex. At the 2nd level of ontology, it generates the hypotheses and ranks all child types of "location“
    - This consolidates and selects "building" as the highest ranked label
    - At a deeper level, it selects the final type "stadium”
- Type ontology is used at every step

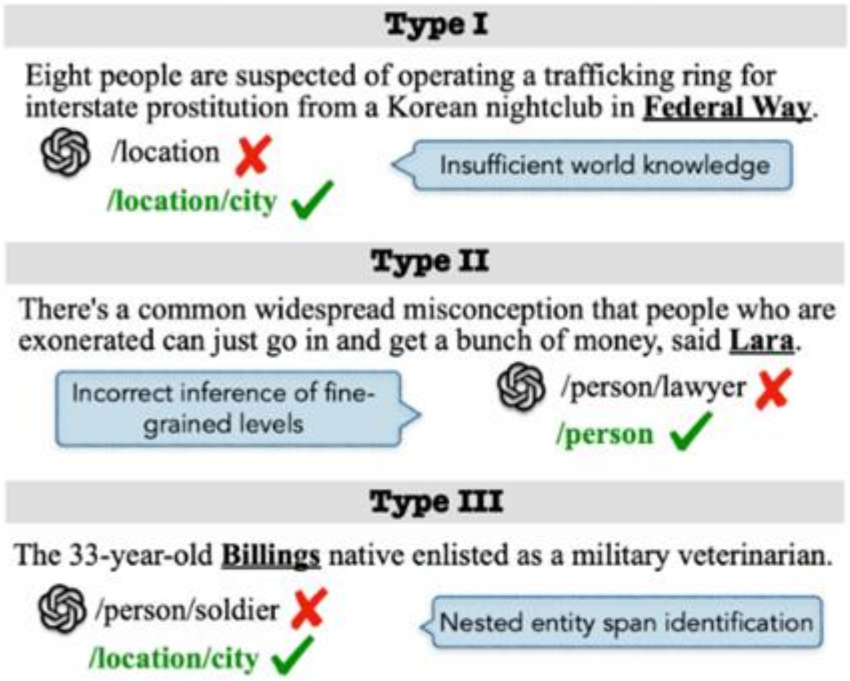


OnEFET: Ontology Enrichment for FET

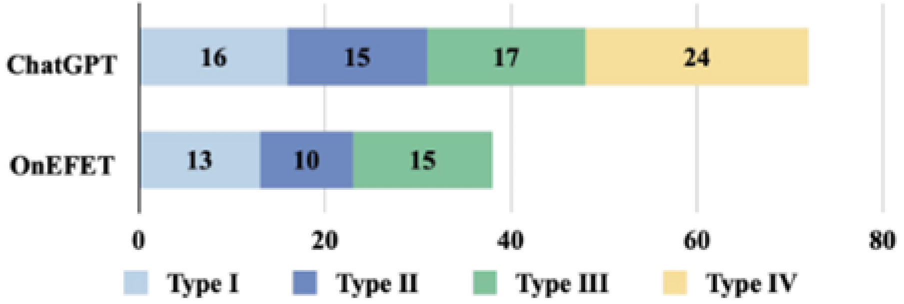
- **Task:** Zero-shot fine-grained entity typing (Zero-shot FET)
  - ❑ **Input:** a sentence with a given entity mention
  - ❑ **Output:** the entity type label from a predefined set of types

→ Usually organized as a structure — ontology





- ❑ Directly prompting LLMs (GPT) cannot handle Zero-shot FET
  - ❑ Nuanced semantic relations as ontology goes deeper and types become more fine-grained
  - ❑ Contextualized information



Siru Ouyang, et al., "Ontology Enrichment for Effective Fine-grained Entity Typing", KDD'24

Overall Framework of OnEFET : Three Steps

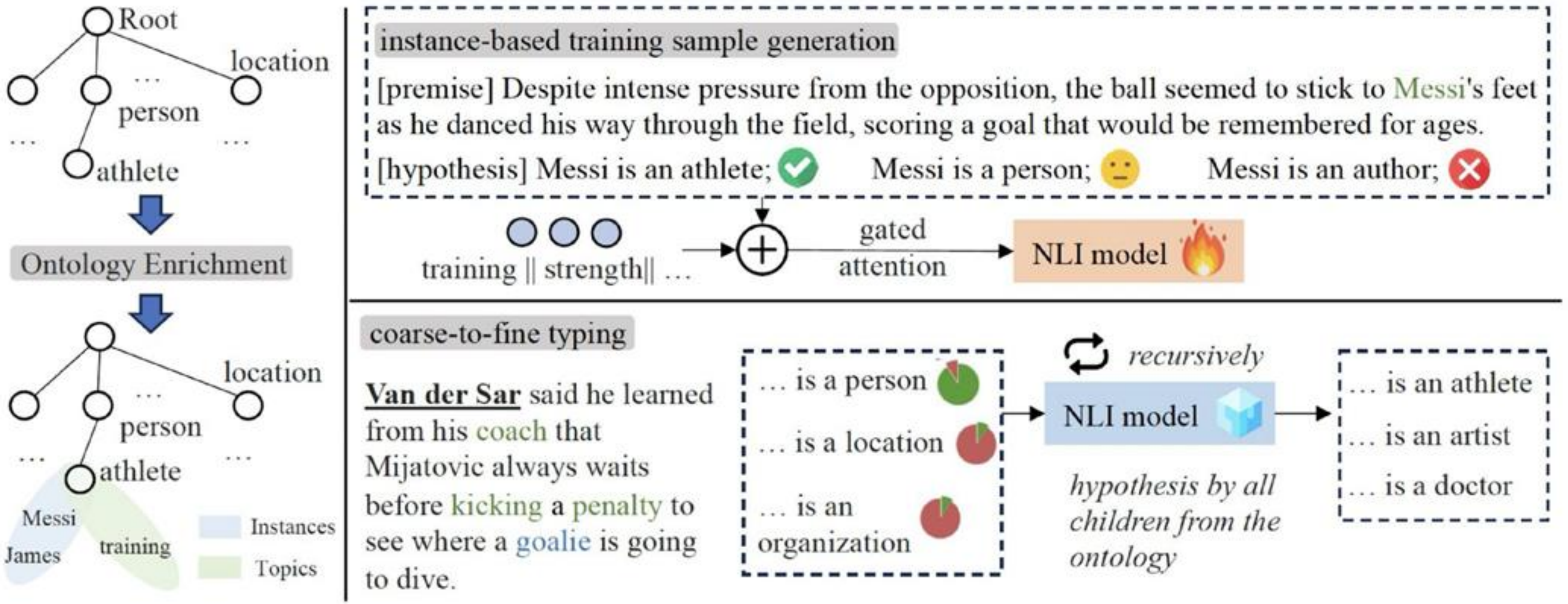
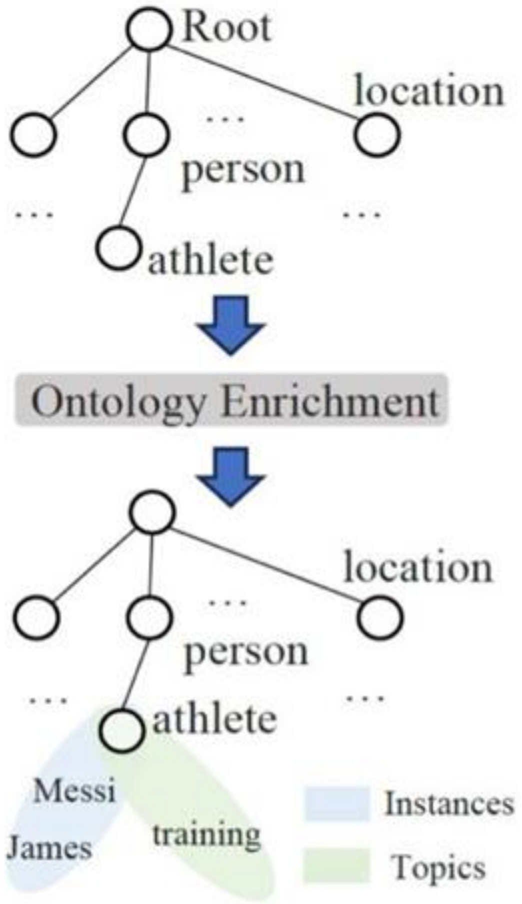


Figure 1: OnEFET framework. The left part is the ontology enrichment for instance and topic information. Then we leverage instances to generate pseudo-training data. Together with topics, we train an NLI model, which will support coarse-to-fine typing during the inference stage.

OnEFET: Step 1 – Ontology Enrichment

- **Enrichment for topic information**
  - Providing entity type T, first select 20 related documents in Wikipedia using Elastic search.
    - Filtering out noisy documents and reduce memory usage
  - SeeTopic [1] for out-of-vocabulary topic words/phrases mining

- **Enrichment for instance information**
  - LM-based instance seeds curation in question-answering style [2]
    - Providing entity type T, first retrieve related sentences in Wikipedia.
    - QA template: [CLS] What is the instance of <T> in this sentence? [SEP] <Sentence> [SEP]
  - SECoExpan [3] for instance expansion



[1] Zhang, Yu, et al. "Seed-guided topic discovery with out-of-vocabulary seeds." NAACL 2022.
[2] Jiao, Yizhu, et al. "Open-vocabulary argument role prediction for event extraction." EMNLP 2022 Findings.
[3] Zhang, Yu, et al. "Entity Set Co-Expansion in StackOverflow." Big Data 2022.

OnEFET: Steps 2 & 3 – Coarse-to-fine Typing

- **Instance information** → contextualized training samples for each fine-grained types
  - Leverage language model to generate a sentence that contain instance e of a specific type
  - Rewarding and paneling mechanism in LM decoding for diversity

$$p_g(x_i|\boldsymbol{x}_{<i}) = \frac{\exp(\boldsymbol{l}_i/\omega)}{\sum_{j=1}^{|V|}\exp(\boldsymbol{l}_j/\omega)}, \quad \omega = \begin{cases} \tau\alpha & x_i \in \boldsymbol{e} \wedge x_i \notin \boldsymbol{x}_{<i} \\ \tau\beta & x_i \in \boldsymbol{x}_{<i} \\ \tau & \text{else} \end{cases}$$

Artist	**Instance**: Leonardo Da Vinci **Topic information**: creativity, art history, style, etc. **Generated Context**: The painting depicts Christ on his way to Calvary, surrounded by angels who are carrying him up into heaven. It has been dated between 1475 and 1480. According to art historian Jos Mareda, it shows "the influence of Leonardo da Vinci", but also that "of Giotto".
Cemetery	**Instance**: Cenotaph **Topic information**: tombstones, grave markers, cremation, etc. **Generated Context**: During her visit to Paris, Emily made sure to explore the famous Père Lachaise, where notable figures such as Oscar Wilde and Jim Morrison were laid to rest.
Sports team	**Instance**: The New York Yankees **Topic information**: training, games, rivalries, etc. **Generated Context**: October 5, 2013: In the final game played at Yankee Stadium, the New York Yankees defeat the Oakland Athletics 7-3 behind four home runs from Alex Rodriguez and six RBI from Mark Teixeira.

OnEFET: Performance Study

- Use 3 benchmark FET datasets: BBN, Ontonotes, and FIGER:

Models	FIGER			OntoNotes			BBN		
	Acc.	Micro-F1	Macro-F1	Acc.	Micro-F1	Macro-F1	Acc.	Micro-F1	Macro-F1
Fully / distantly-supervised Setting									
AFET (Ren et al., 2016)	55.3	66.4	69.3	55.1	64.7	71.1	67.0	73.5	72.7
UFET (Choi et al., 2018)	-	-	-	59.5	71.8	76.8	-	-	-
BERT-MLMET (Dai et al., 2021)	-	-	-	67.4	80.4	85.4	-	-	-
LITE (Li et al., 2022)	-	**83.3**	**86.7**	-	**80.9**	**86.4**	-	-	-
Zero-Shot Setting									
ProtoZET (Ma et al., 2016b)	29.6	56.4	55.1	28.1	34.5	33.7	25.1	63.1	58.2
OTyper (Yuan and Downey, 2018)	47.2	67.2	69.1	31.8	36.0	39.1	29.0	48.8	54.4
ZOE (Zhou et al., 2019)	**58.8**	71.3	74.8	50.7	60.8	66.9	61.8	74.9	74.6
DZET (Obeidat et al., 2019)	28.5	56.0	55.1	23.1	28.1	27.6	-	-	-
MZET (Zhang et al., 2020)	31.9	57.9	55.5	33.7	43.7	42.3	29.4	68.7	60.6
ChatGPT* (Ouyang et al., 2022)	51.7	65.3	58.3	27.7	37.5	32.6	25.1	55.9	50.7
OntoType (Komarlu et al., 2023)	49.1	67.4	75.1	65.7	73.4	81.5	-	-	-
OnEFET	56.3	**72.7**	**78.6**	**68.6**	**76.3**	**83.4**	**68.5**	**80.1**	**81.7**

- OnEFET significantly outperform previous baseline models; on par with supervised settings.
- OnEFET could be smoothly transferred to unseen settings, even with more entity types

- Transferability test on UFET

Model	P	R	F1
supervised setting			
MLMET [8]	**53.6**	45.3	49.1
LITE [20]	52.4	48.8	**50.6**
zero-shot setting			
direct NLI	1.5	7.1	2.5
direct OnEFET	7.2	17.5	10.2
OnEFET	31.4	**53.1**	39.5

- Ablation Study

Model	Acc.	Micro-F1	Macro-F1
OnEFET	68.6	76.3	83.4
w/o topics	67.0	74.1	81.9
w/ 3 instances	59.8	70.5	75.6
w/o coarse-to-fine	67.5	75.7	82.1
w/o GCE loss	66.4	74.5	81.6

Outline

- Overview
- Entity Typing
- Relation Extraction
  - Supervised Relation Extraction
  - Weakly-Supervised Relation Extraction
  - Synthesis-Based Relation Extraction
  - Reasoning-Based Relation Extraction
- Comprehensive Knowledge Structuring

Relation Extraction

- **Relationship extraction** (RE): A subtask of **information extraction**
  - Detect and classify semantic relationship mentions within a set of artifacts from text (e.g., <Bill Clinton, LocatedIn, New York>)
  - Typically, determine a relationship between two distinct entities from text, producing fact triples in the form [head, relation, tail] or [subject, predicate, object]
- Datasets: Researchers have constructed multiple datasets for benchmarking relation extraction methods, e.g., infoboxes, RDF trips, freebase (WikiData), and WordNet
  - Ex. UMLS (Unified Medical Language System) (NLM) defines 134 subject categories, entity types and 54 relations
  - Ex. TACRED (Zhang, 2017): 106K+ examples of relation triples (41 relations)
- Relation extraction benefits natural language understanding in many ways
  - Ex. Knowledge graph construction, question answering, text classification, event discovery, coreference, …

Outline

- Span Detection
- Entity Typing
- Relation Extraction
  - Supervised Relation Extraction
  - Weakly-Supervised Relation Extraction
  - Synthesis-Based Relation Extraction
  - Reasoning-Based Relation Extraction
- Comprehensive Knowledge Structuring

Document-Level Relation Extraction

- Document-level relation extraction (DocRE)
  - Extract semantic relations among entity pairs in a document
- Blindly considering the full document?
  - A subset of the sentences in the doc (“evidence”) should often be sufficient to identify the relation
- An evidence-enhanced DocRE framework: EIDER
  - Efficiently extracts evidence and effectively leverages the extracted evidence to improve DocRE
- Using a document-level relationship extraction dataset DocRED (2019)
- Relation extraction benefits natural language understanding in many ways
  - Ex. Knowledge graph construction

Head:Hero of the Day Tail:the United States Rel:[country of origin]
GT evidence sentences: [1,10] Extracted evidence: [1,10]

Original document as input: [1] Load is the sixth studio album by the American heavy metal band Metallica, released on June 4, 1996 by Elektra Records in the United States ... [9] It was certified 5×platinum ... for shipping five million copies in the United States. [10] Four singles—"Hero of the Day", "Until It Sleeps", "Mama Said", and "King Nothing" — were released as part of the marketing campaign for the album.
Prediction scores: NA: 17.63 country of origin: 14.79

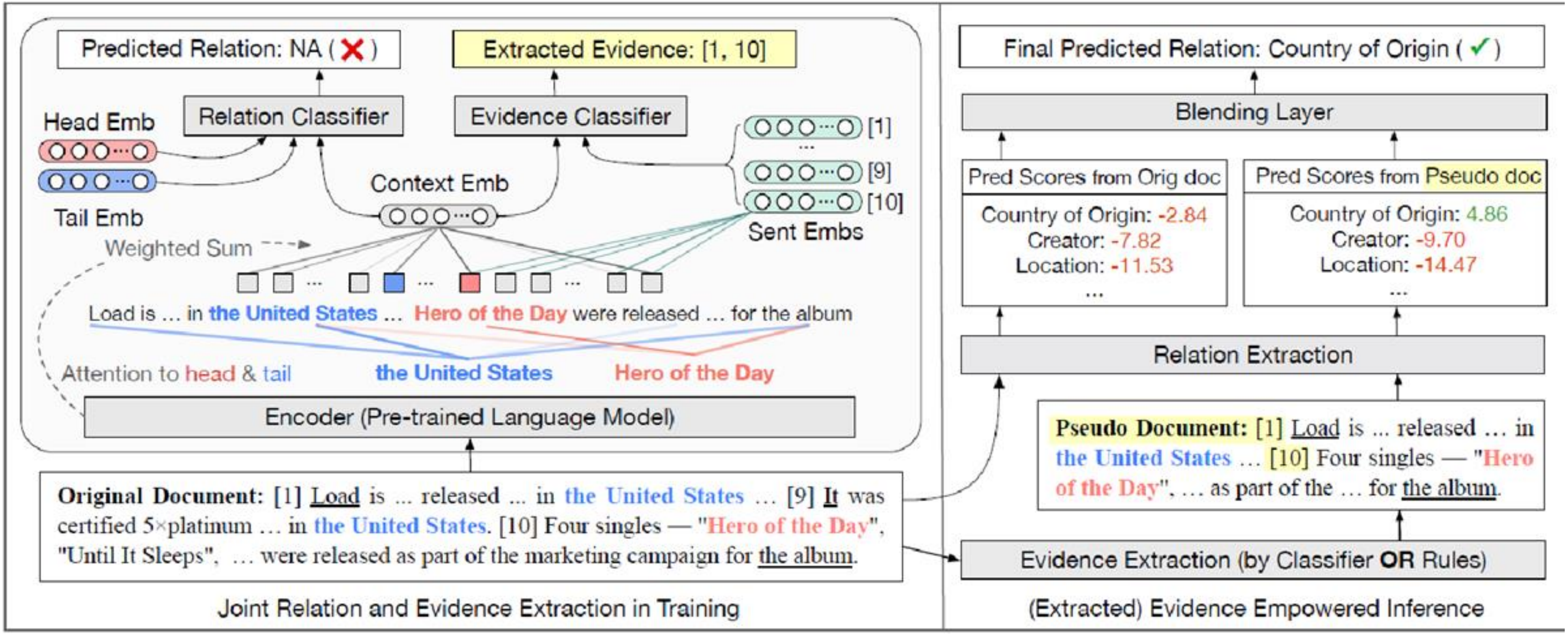
Extracted evidence as input: [1] Load is the sixth studio album ... released ... in the United States ... [10] Four singles — "Hero of the Day", ... were released ... for the album.
Prediction scores: country of origin: 18.31 NA: 13.45

Final prediction of our model: country of origin (✓)

Only need [1]+[10] to identify [head, relation, tail]

Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, Jiawei Han, “EIDER: Evidence-enhanced Document-level Relation Extraction”, ACL’22 Findings

EIDER Architecture



The left part (the training stage), we jointly extract relation and evidence using multi-task learning, where the two tasks have their own classifier and share the base encoder

The right part (the inference stage), we fuse the predictions on the original document and the extracted evidence using a blending layer

Typed Entity Marker (punct)

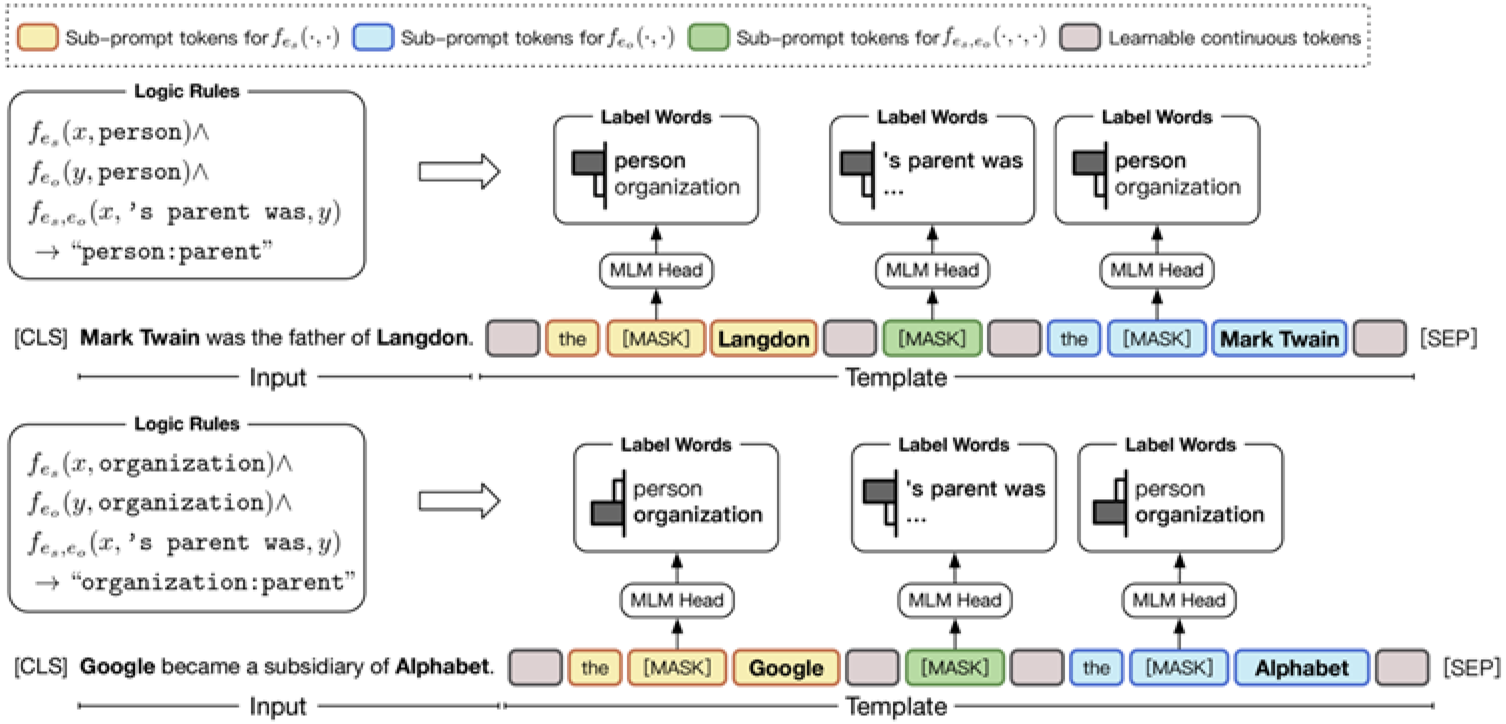
- Focused on sentence-level binary RE
- Compared different methods to mark the head and tail entities and inject the entity type knowledge for better RE

Method	Input Example	$\text{BERT}_{\text{BASE}}$	$\text{BERT}_{\text{LARGE}}$	$\text{RoBERTa}_{\text{LARGE}}$
Entity mask	[SUBJ-PERSON] was born in [OBJ-CITY].	69.6	70.6	60.9
Entity marker	[E1] Bill [/E1] was born in [E2] Seattle [/E2].	68.4	69.7	70.7
Entity marker (punct)	@ Bill @ was born in # Seattle #.	68.7	69.8	71.4
Typed entity marker	⟨S:PERSON⟩ Bill ⟨/S:PERSON⟩ was born in ⟨O:CITY⟩ Seattle ⟨/O:CITY⟩.	**71.5**	**72.9**	71.0
Typed entity marker (punct)	@ * person * Bill @ was born in # ∧ city ∧ Seattle #.	70.9	72.7	**74.6**

Wenxuan Zhou, Muhao Chen, "An Improved Baseline for Sentence-level Relation Extraction", AACL'22

PTR

- Specialized and semantically meaningful templates for each class for prompt tuning
- For binary RE, one sub-prompt for each entity and one sub-prompt for relation

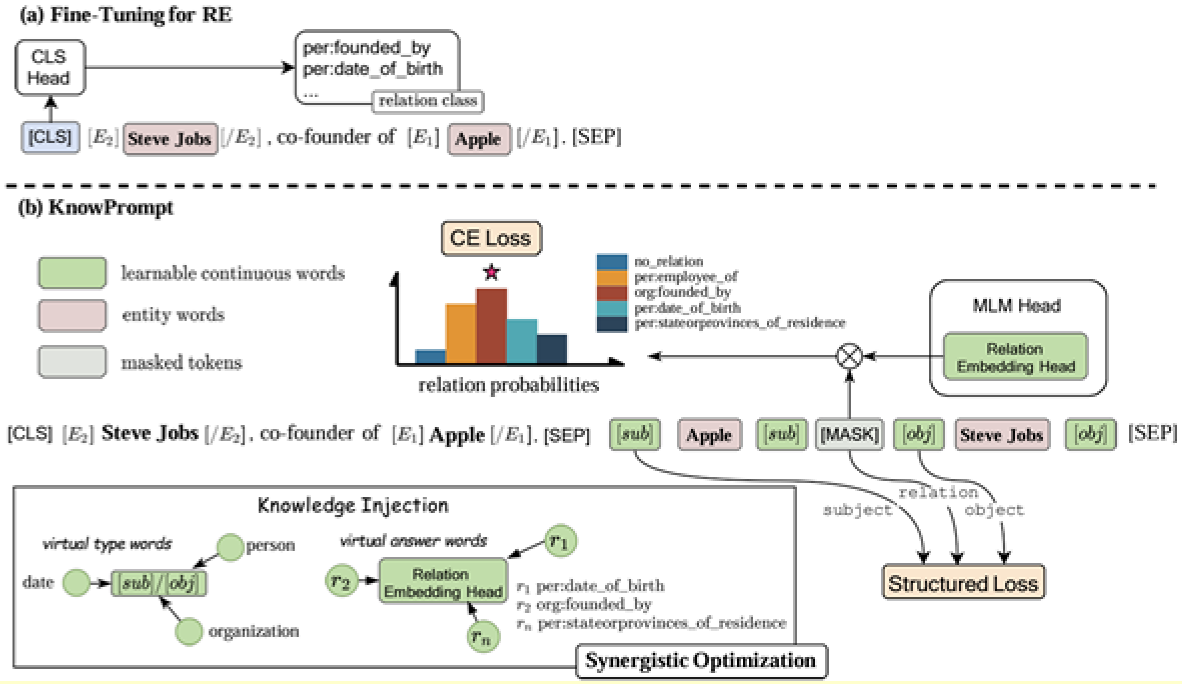


Class Label	[MASK]$_1$	[MASK]$_2$	[MASK]$_3$
per:country_of_birth	person	was born in	country
per:stateorprovince_of_birth	person	was born in	state
per:city_of_birth	person	was born in	city
per:employee_of	person	's employee was	organization
per:parents	person	's parent was	person
per:age	person	's age was	number
org:founded_by	organization	was founded by	person
org:country_of_headquarters	organization	was located in	country
org:stateorprovince_of_headquarters	organization	was located in	state
org:city_of_headquarters	organization	was located in	city
org:number_of_employees/members	organization	's employer has	number
org:members	organization	's member was	organization
org:parents	organization	's parent was	organization
no_relation	entity	is irrelevant to	entity

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, Maosong Sun, "PTR: Prompt Tuning with Rules for Text Classification", AI Open

KnowPrompt

- Leveraged virtual type words and virtual answer words for prompt tuning
- Injected prior knowledge about the subjects, objects, and relations for virtual token initialization
- Utilized implicit structure constrained loss for prompt optimization



Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, Huajun Chen, "KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction", WWW'22

Outline

- Span Detection
- Entity Typing
- Relation Extraction
  - Supervised Relation Extraction
  - Weakly-Supervised Relation Extraction
  - Synthesis-Based Relation Extraction
  - Reasoning-Based Relation Extraction
- Comprehensive Knowledge Structuring

Corpus-Level Relation Extraction

- Utilized additional representation of relation triple for initial weak supervision extraction and latent clustering for further denoising
- Applied further prompt tuning for context understanding and pattern generalization

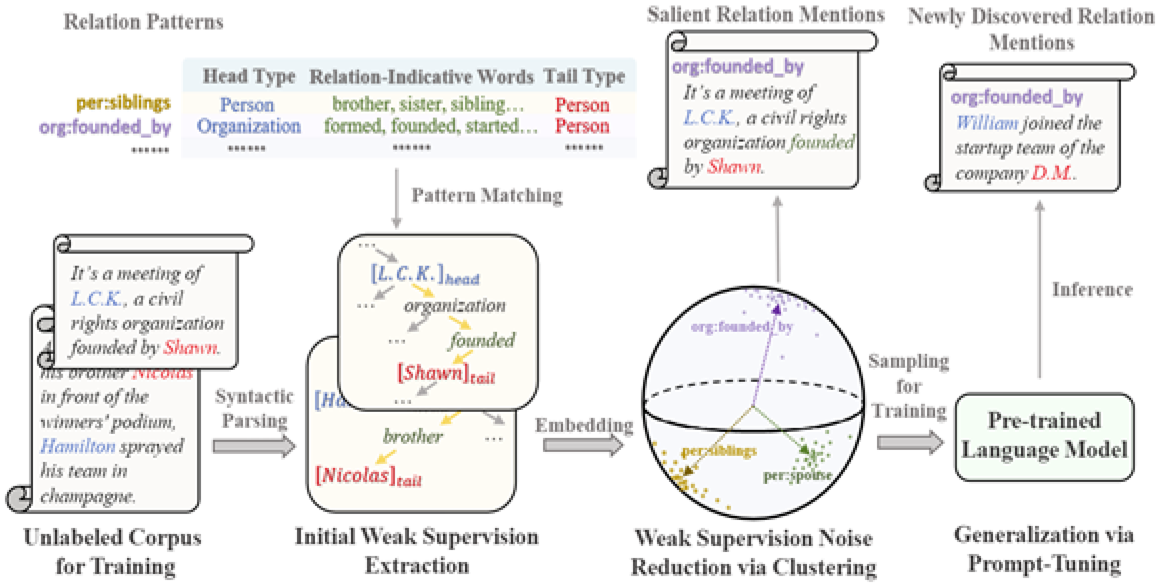


Fig. 2. Framework overview. Our model mainly consists of three steps: (1) relation triple representation extraction, (2) latent space clustering, and (3) prompt-tuning with sub-prompts.

Sizhe Zhou, Suyu Ge, Jiawei Han, "Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns", ECMLPKDD'23

Corpus-Level Relation Extraction: Experiment

- Leading low-resource performance
- Each component is indispensable
  - Weak supervision provides relatively high recall
  - Clustering provides relatively high precision
  - Prompt-tuning is important for boosting recall

Model	TACRED				TACREV				ReTACRED			
K	4	8	16	Mean	4	8	16	Mean	4	8	16	Mean
w/ weak supervision												
EXACT MATCHING*	-	-	-	48.87	-	-	-	53.67	-	-	-	54.86
COSINE	23.28	26.60	37.16	29.01	21.43	30.85	41.21	31.16	28.12	35.00	44.54	35.89
COSINE*	-	-	-	58.88	-	-	-	60.80	-	-	-	68.59
RCLUS NOISY	45.35	50.94	55.73	50.67	50.41	61.67	**66.85**	59.64	56.89	65.81	71.09	64.60
RCLUS BALANCED	45.19	55.71	59.33	53.41	55.36	58.74	64.56	59.55	53.84	65.27	71.03	63.38
RCLUS	**49.89**	**56.65**	**60.26**	55.60	**56.94**	**63.75**	66.50	62.40	**61.03**	**68.78**	**72.23**	67.35
w/ ground truth supervision												
FINE-TUNING	13.62	26.09	32.07	23.93	18.75	25.21	35.12	26.36	17.36	31.77	42.63	30.59
GDPNET	13.79	28.42	43.11	28.44	15.61	24.59	42.12	27.44	19.20	35.79	52.84	35.94
PTR	39.16	49.46	54.67	47.76	47.18	51.58	59.17	52.64	51.27	62.60	71.11	61.66

Table 1. F_1 scores (%) on full test set with different sizes ($K = 4, 8, 16$) for each relation label.

Model	TACREV		
	Precision	Recall	F_1
RCLUS			
w/ Weak	59.30	49.02	53.67
w/ Prompt	48.25	75.73	58.95
w/ Weak + Prompt	58.80	72.07	64.76
w/ Weak + Cluster	63.62	40.61	49.57
w/ Weak + Cluster + Prompt	60.76	74.29	66.85
w/ Weak + Cluster + Prompt*	57.85	78.47	66.61

Table 3. Ablation study of RCLUS

Chemical Reaction Extraction with Weak Supervision

- Task: Chemical Reaction Extraction
  - Goal: Extract chemical reactions from a scientific paper
  - Input: A scientific paper
  - Output: Multiple structured chemical reactions

Scientific Paper

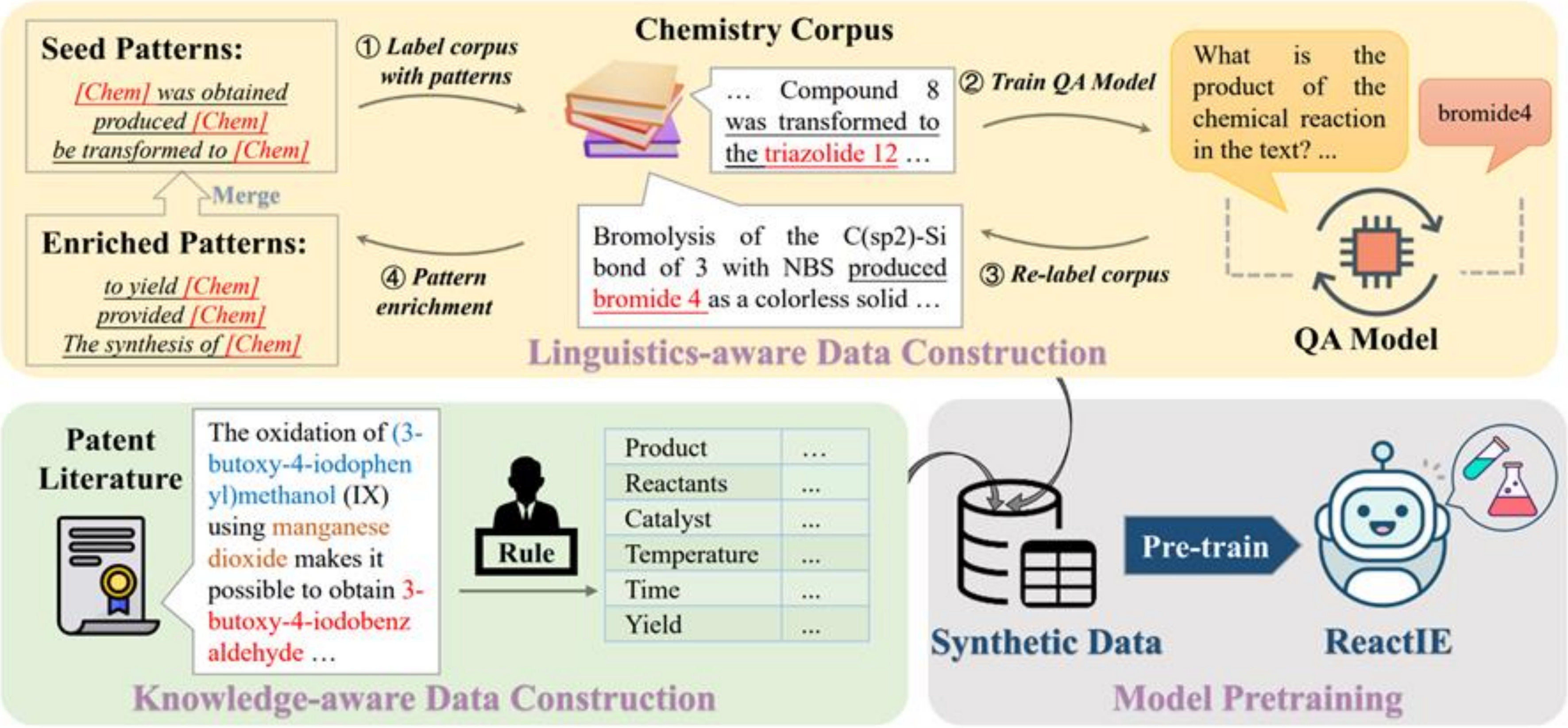
… The methyl-substituted porphyrinogens (7e and 7f) were oxidized with chloranil, and meso-unsubstituted porphyrinogens (7g and 7h) were oxidized with 0.1% aqueous $FeCl_3$ in $CHCl_3$ at room temperature to obtain 16π-conjugated systems 5e in 6%, 5f in 7%, 5g in 5%, and 5h in 4% yields. …

	Chemical Reaction 1		Chemical Reaction 2
Product	5e	Product	5g
Reactants	7e	Reactants	7g
Reaction Type	oxidation	Reaction Type	oxidation
Catalyst	chloranil	Catalyst	$FeCl_3$
Solvent	$CHCl_3$	Solvent	$CHCl_3$
Temperature	room	Temperature	room
Yield	6%	Yield	5%

Ming Zhong, Siru Ouyang, Minhao Jiang, Vivian Hu, Yizhu Jiao, Xuan Wang, Jiawei Han, "ReactIE: Enhancing Chemical Reaction Extraction with Weak Supervision", ACL'23 Findings

Chemical Reaction Extraction with Weak Supervision

- Method: ReactIE
  - Linguistic cues
  - Domain Knowledge



Chemical Reaction Extraction with Weak Supervision

❑ Result for Product Extraction

Models	P (%)	R (%)	F (%)
Unsupervised			
OPSIN	18.8	5.4	8.4
REACTIE	**69.7**	**53.5**	**60.5**
Supervised			
BiLSTM	52.4	46.7	49.4
BiLSTM (w/ CRF)	54.3	49.1	51.6
BERT	78.8	56.8	66.0
BioBERT	76.4	61.3	68.0
ChemBERT	84.6	69.4	76.2
FlanT5	88.0	83.2	85.5
ReactIE	**94.2**	**88.2**	**91.1**
- *linguistics cues*	89.8	84.7	87.2
- *domain knowledge*	92.6	87.1	89.8

❑ Result for Role Extraction

Models	P (%)	R (%)	F (%)
BERT	69.2	69.2	69.2
BioBERT	73.3	75.5	74.3
ChemBERT	77.0	76.4	76.7
ChemRxnBERT	79.3	78.1	78.7
FlanT5	76.1	75.4	75.8
ReactIE	**80.8**	**82.5**	**81.6**
- *linguistics cues*	78.1	83.3	80.6
- *domain knowledge*	74.8	79.8	77.2

❑ Case Study

Input Text: It is worth mentioning that when the reaction of 2-cyano-N,N-dimethylethanethioamide (1s) with 3-phenylpropiolaldehyde (2a) was carried out at room temperature under nitrogen for 5 h, an aldol condensation product 5a was isolated in 73% yield, which can be further converted to 2-aminothienyl ether 3sa in 85% yield under the optimized reaction conditions.

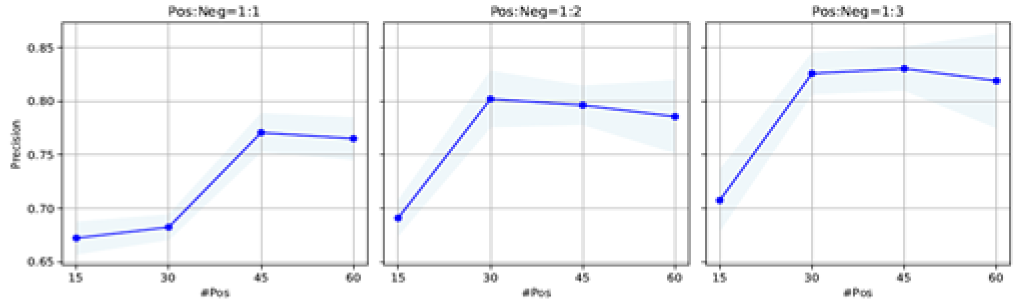
ReactIE (Significant Error)	**GPT-4** (Minor Issue)	**Reaction Miner** (Perfect)
Reaction 1: Product: 5a Reactants: (1s), (2a), … Reaction type: aldol condensation Catalyst: nitrogen Temperature: room Time: 5 h Yield: 73%	Reaction 1 Product: aldol condensation product 5a Reactant: 2-cyano-… (1s) Reactant: 3-… (2a) Atmosphere: nitrogen Temperature: room temperature Time: 5 h Yield: 73% (aldol condensation product 5a)	Reaction 1 Product: 5a Reactant: 2-cyano-… (1s), 3-… (2a) Atmosphere: nitrogen Reaction type: aldol condensation Temperature: room temperature Time: 5 h Yield: 73% (5a)
Reaction 2: Product: 3sa Reactants: (1s), (2a), … Yield: 85%	Reaction 2 Product: 2-aminothienyl ether 3sa Reaction type: optimized reaction Yield: 85% (2-aminothienyl ether 3sa)	Reaction 2 Product: 2-aminothienyl ether 3sa Reactant: 5a Yield: 85% (2-aminothienyl ether 3sa)

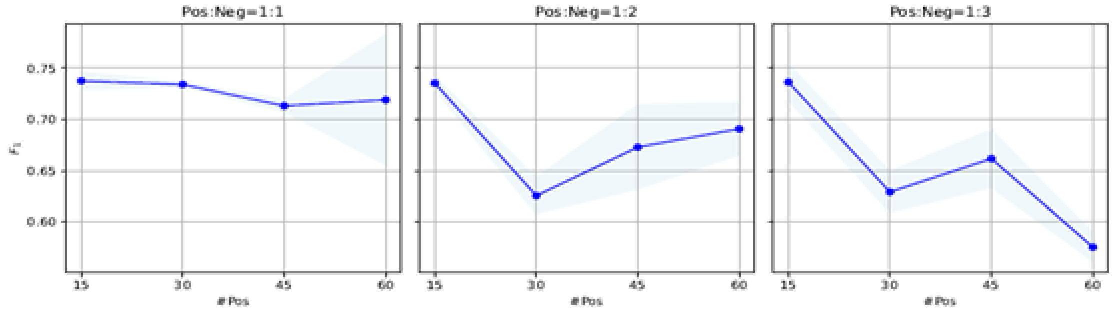
Outline

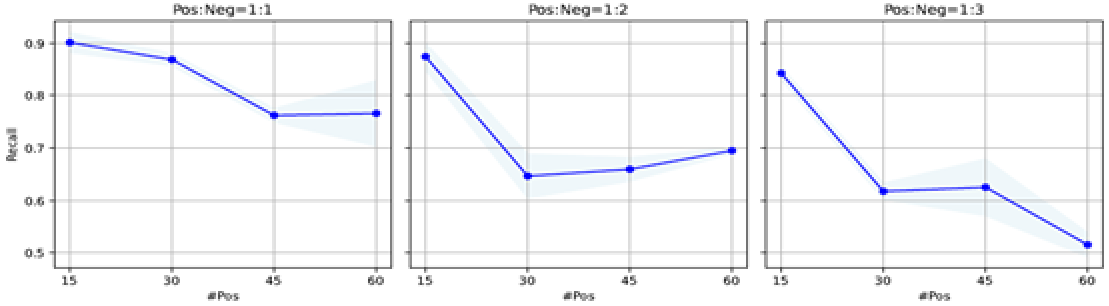
- Span Detection
- Entity Typing
- Relation Extraction
  - Supervised Relation Extraction
  - Weakly-Supervised Relation Extraction
  - Synthesis-Based Relation Extraction
  - Reasoning-Based Relation Extraction
- Comprehensive Knowledge Structuring

LLM for Zero-Shot Relation Extraction

- Prompting for more synthetic data
  - Larger #p&n or larger n:p ratio → higher precision and lower recall
  - More positive seeds does not bring proportionally larger positive pattern coverage compared to randomly sampled negative seeds → Positive predictions concentrated on dominant positive patterns



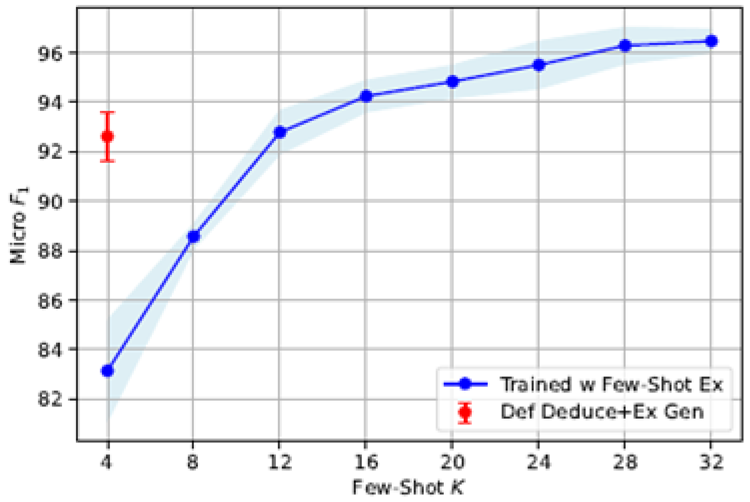




Sizhe Zhou, Yu Meng, Bowen Jin, Jiawei Han, "Grasping the Essentials: Tailoring Large Language Models for Zero-Shot Relation Extraction", arXiv preprint arXiv:2402.11142 (2024)

LLM for Zero-Shot Relation Extraction

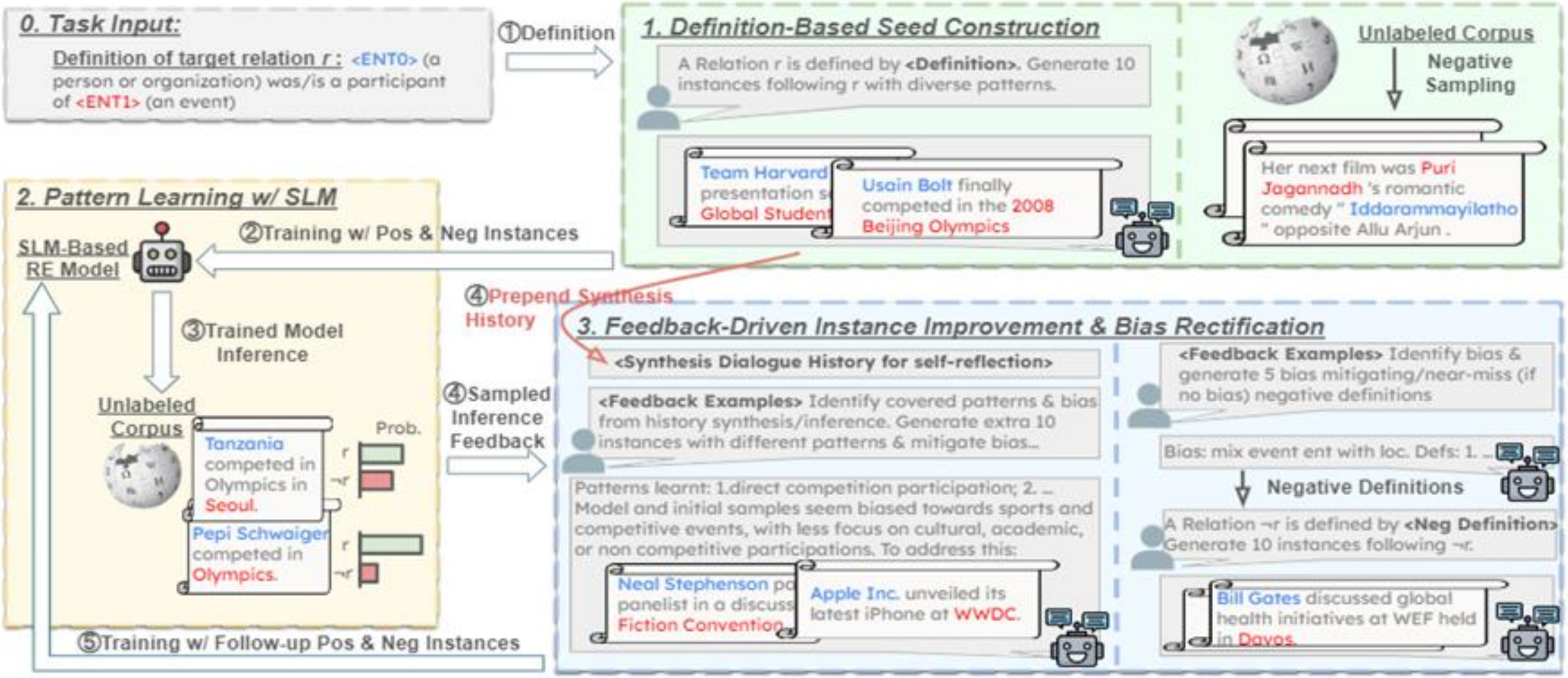
- Commonly adopted RE few-shot setup can be converted to definition-oriented setup while achieving much better results
  - Derive-and-then-generate extends the patterns conveyed by few-shot instances
  - Coverage of derived definitions is limited by the coverage of few-shot instances



Gold Definition	Gold Few-Shot Instances For Derivation	Derived Definition
<ENT1> was/is the occupation of <ENT0> (a person)	1. <ENT0>Pierre Maudru</ENT0> (1892\u20131992) was a French <ENT1>screenwriter</ENT1> . Goble p.189 He also directed three films . 2. WWF Hall of Famer Bob Backlund and Extreme Championship Wrestling <ENT1>manager</ENT1> <ENT0>Bill Alfonso</ENT0> also made surprise appearances during the event . 3. In May 2010 , Paratici moved from Sampdoria to Juventus , along with Director General Giuseppe Marotta and <ENT1>Manager</ENT1> <ENT0>Luigi Delneri</ENT0> . 4. <ENT0>Else Reval</ENT0> (14 June 1893 \u2013 25 January 1978) was a German <ENT1>film actress</ENT1> . Giesen p.210	<ENT1> is the profession in which <ENT0> (a person) works or has worked.
<ENT0> (a person or organization) was/is a participant of <ENT1> (an event)	1. He only saw limited action in <ENT1>Euro 2000</ENT1> as cover for left - back <ENT0>Arthur Numan</ENT0> . 2. <ENT0>Francesco Cameli</ENT0> was a sailor from Italy , who represented his country at the <ENT1>1928 Summer Olympics</ENT1> in Amsterdam , Netherlands . 3. <ENT0>Giannin Andreossi</ENT0> (born July 2 , 1902 , date of death unknown) was a Swiss ice hockey player who competed in the <ENT1>1928 Winter Olympics</ENT1> . 4. <ENT0>Ren\u00e9 Sch\u00f6fisch</ENT0> (born February 3 , 1962) is a German speed skater who competed for East Germany in the <ENT1>1984 Winter Olympics</ENT1> .	<ENT1> is the major international sports competition in which <ENT0> (an athlete) has competed.

RePAL

- Leverage relation definitions for LLM-based instance synthesis
- Train SLM as relation extractors with its inference on unlabeled corpus as feedback for iterative instance synthesis (LLM's self-reflection on synthesis history and both the inference feedback)



Experimental Results

- Synthesis or pattern knowledge distillation from LLMs is a promising way to boost the RE performance under zero-shot setting
- For better synthesis, LLMs should be fully comprehend the complicated semantics of target relations and LLMs should be guided to explore different relation patterns (especially long-tail patterns)

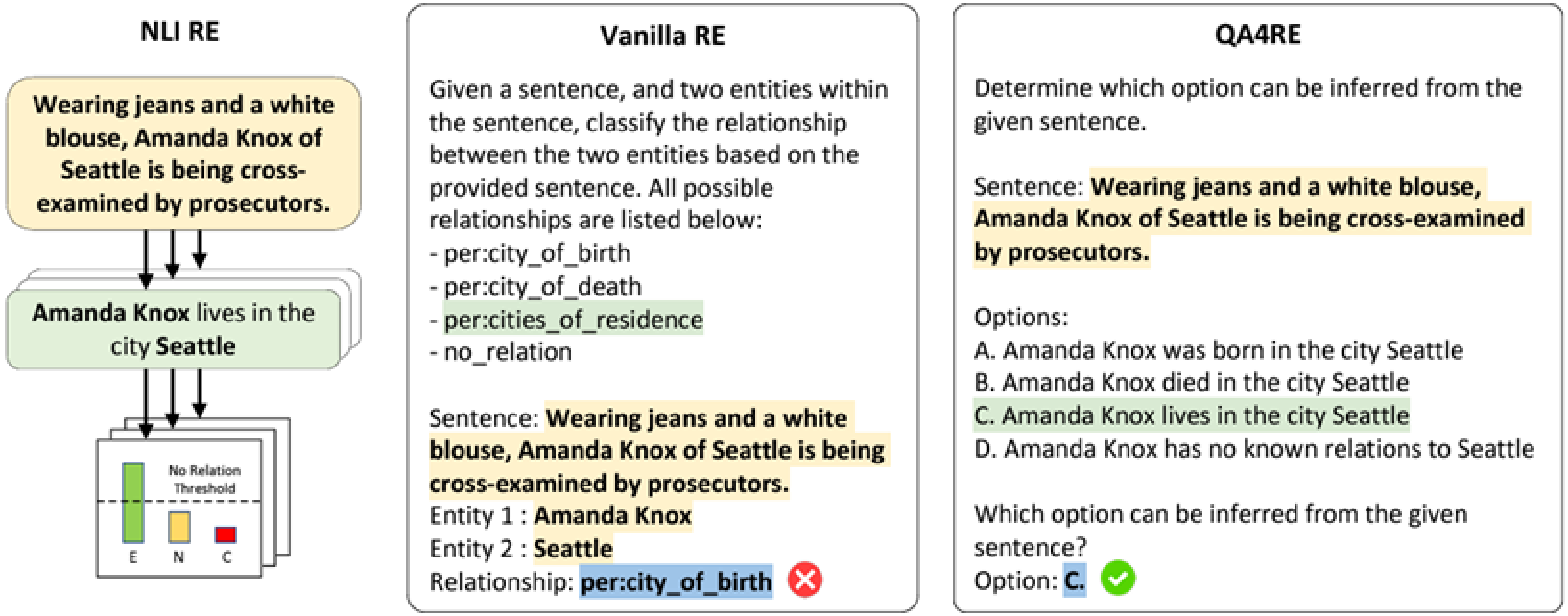
Model	DefOn-FewRel				DefOn-Wiki-ZSL			
	Precision	Recall	F_1	Macro-F_1	Precision	Recall	F_1	Macro-F_1
Fully-Supervised								
RoBERTa NLI	79.36	98.46	86.99	-	68.66	97.14	78.46	-
Zero-Shot								
Random Guess	7.14	50.77	12.52	-	6.67	51.01	11.67	-
GPT-3.5	55.09	61.81	53.94	-	42.64	45.70	39.60	-
RoBERTa NLI	48.79	81.17	55.07	-	36.23	62.65	35.28	-
ZS-BERT	35.91	35.70	-	35.78	37.11	33.85	-	35.26
RelationPrompt	74.39	66.89	-	67.78	71.89	60.50	-	61.08
RE-Matching	77.49	72.95	-	75.11	73.10	68.99	-	70.97
REPaL (Ours)	71.17	82.34	**73.02**	**88.44**	62.12	52.99	**48.17**	**81.25**

Outline

- Span Detection
- Entity Typing
- Relation Extraction and Event Extraction
  - Supervised Relation Extraction
  - Weakly-Supervised Relation Extraction
  - Synthesis-Based Relation Extraction
  - Reasoning-Based Relation Extraction
- Comprehensive Knowledge Structuring

Aligning Instruction Tasks w/ Relation Extraction

- Hypothesized that instruction-tuning has been unable to elicit strong RE capabilities in LLMs due to RE’s low incidence in instruction-tuning datasets, making up less than 1% of all tasks
- Formulating RE as multiple-choice QA and applying entity types as constraints



Kai Zhang, Bernal Jiménez Gutiérrez, Yu Su, “Aligning Instruction Tasks Unlocks Large Language Models as Zero-Shot Relation Extractors”, ACL'23 Findings

Experimental Results

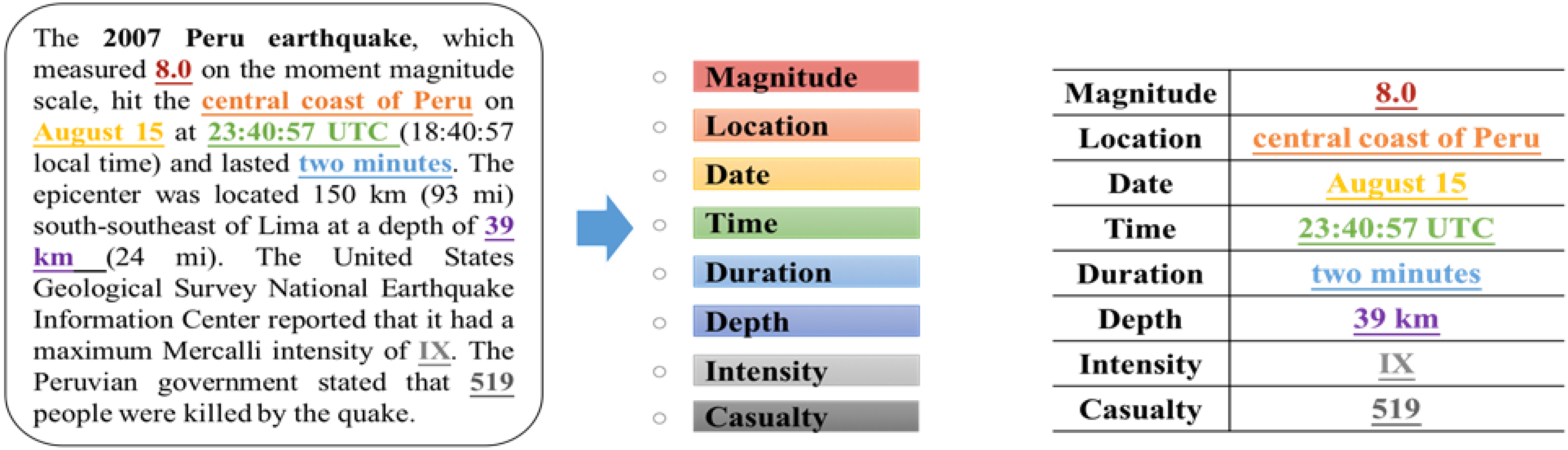
- QA4RE Achieved consistently better performance in few-shot and zero-shot settings
- Illustrated a promising way of adapting LLMs to underrepresented tasks by alignments

Methods		TACRED P	TACRED R	TACRED F1	RETACRED P	RETACRED R	RETACRED F1	TACREV P	TACREV R	TACREV F1	SemEval P	SemEval R	SemEval F1	Avg. F1
Baselines														
NLI_{BART}		42.6	65.0	51.4	59.5	34.9	44.0	44.0	74.6	55.3	21.6	23.7	22.6	43.3
$NLI_{RoBERTa}$		37.1	76.9	50.1	52.3	67.0	58.7	37.1	83.6	51.4	17.6	20.9	19.1	44.8
$NLI_{DeBERTa}$		42.9	76.9	55.1	71.7	58.3	64.3	43.3	84.6	57.2	22.0	25.7	23.7	50.1
$SuRE_{BART}$		13.1	45.7	20.4	17.9	34.6	23.6	14.1	52.3	22.2	0.0	0.0	0.0	16.5
$SuRE_{PEGASUS}$		13.8	51.7	21.8	16.6	34.6	22.4	13.5	54.1	21.6	0.0	0.0	0.0	16.4
GPT-3.5 Series														
ChatGPT	Vanilla	32.1	74.8	44.9	45.4	61.3	52.1	30.3	79.6	43.9	18.2	20.8	19.4	40.1
	QA4RE	32.8	68.0	44.2 (−0.7)	48.3	76.8	59.3 (+7.2)	34.7	79.1	48.2 (+4.3)	29.9	35.2	32.3 (+12.9)	46.0 (+5.9)
code-002	Vanilla	27.2	70.1	39.2	42.7	70.4	53.1	27.5	77.7	40.6	27.2	25.6	26.4	39.8
	QA4RE	37.7	65.4	47.8 (+8.6)	48.0	74.0	58.2 (+5.1)	31.7	65.5	42.7 (+2.1)	25.2	29.2	27.0 (+0.6)	43.9 (+4.1)
text-002	Vanilla	31.2	73.1	43.7	44.1	76.3	55.9	30.2	76.8	43.3	31.4	28.8	30.1	43.2
	QA4RE	35.6	68.4	46.8 (+3.1)	46.4	72.4	56.5 (+0.6)	35.7	76.8	48.8 (+5.4)	29.4	34.3	31.6 (+1.5)	45.9 (+2.7)
text-003	Vanilla	36.9	68.8	48.1	49.7	62.2	55.3	38.2	76.8	51.0	33.2	39.3	36.0	47.6
	QA4RE	47.7	78.6	**59.4** (+11.3)	56.2	67.2	61.2 (+5.9)	46.0	83.6	**59.4** (+8.4)	41.7	45.0	43.3 (+7.3)	**55.8** (+8.2)
FLAN-T5 Series														
XLarge	Vanilla	51.6	49.1	50.3	54.3	40.3	46.3	56.0	59.1	57.5	35.6	29.8	32.4	46.6
	QA4RE	40.0	78.2	53.0 (+2.7)	57.1	79.7	66.5 (+20.2)	40.7	85.9	55.3 (−2.2)	45.1	40.1	42.5 (+10.1)	54.3 (+7.7)
XXLarge	Vanilla	52.1	47.9	49.9	56.6	54.0	55.2	52.6	50.9	51.7	29.6	28.8	29.2	46.5
	QA4RE	40.6	82.9	54.5 (+4.6)	56.6	82.9	**67.3** (+12.1)	39.6	86.4	54.3 (+2.6)	41.0	47.8	**44.1** (+14.9)	55.1 (+8.6)

Methods	K=0	K=4	K=8	K=16	K=32
Fine-Tuning	-	9.0	21.2	29.3	33.9
PTR	-	26.8	30.0	32.9	36.8
KnowPrompt	-	30.2	33.7	34.9	35.0
$NLI_{DeBERTa}$-TEMP1	55.0	**64.2**	**64.7**	**58.7**	**65.7**
$NLI_{DeBERTa}$-TEMP2	49.4	**51.2**	47.3	**50.5**	48.1
Vanilla	48.1	46.2		-	
QA4RE	59.4	**62.0**		-	

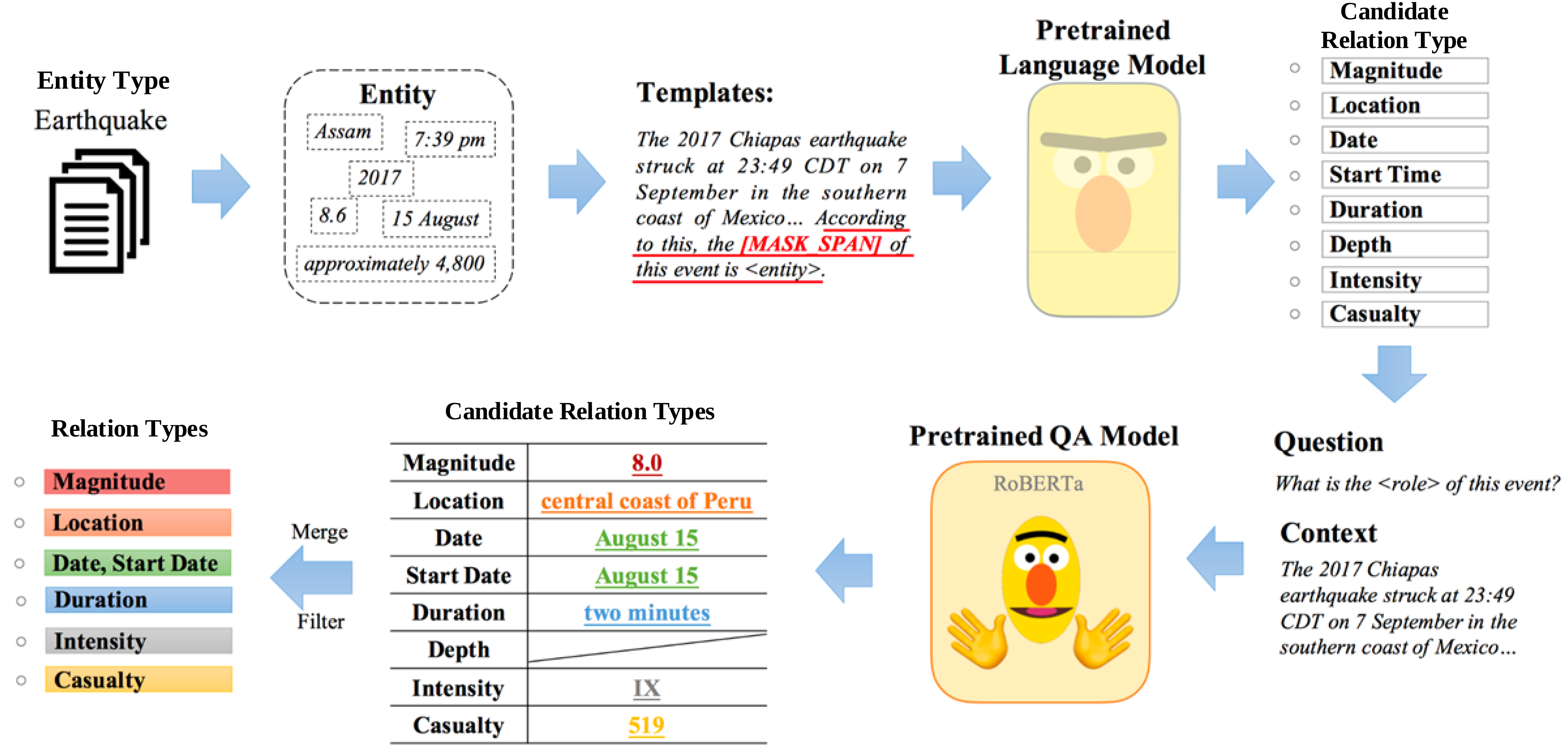
Open-Vocabulary Relation Type Discovery

- Related Work:
  - Most of existing studies rely on hand-crafted ontologies (costly, cannot generalize)
  - A few studies try to automatically induce argument roles (limited pre-defined glossary)
- **New Task**: Infer a set of relation type names for a given entity type to describe the crucial relations between the entity type and its related entities



Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji and Jiawei Han "Open-Vocabulary Argument Role Prediction for Event Extraction", EMNLP'22

Framework for RolePred



Experiment: Relation Type Discovery

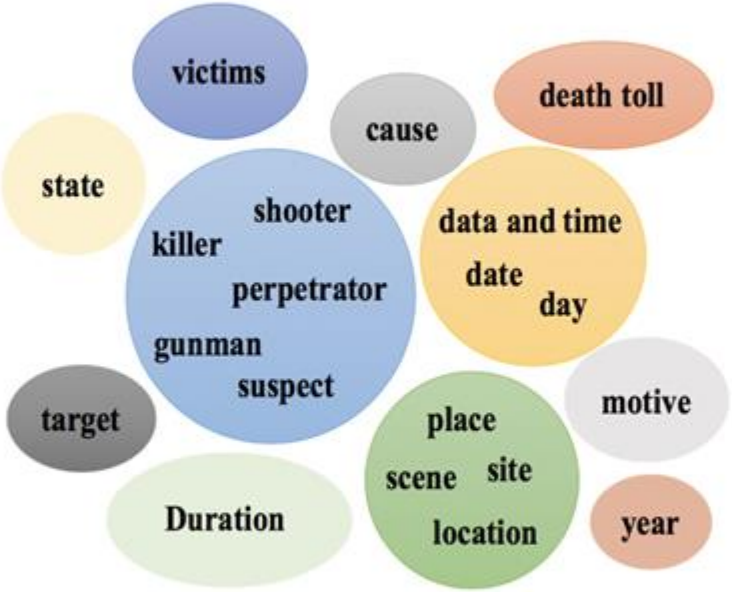
Relation Type Prediction

Models	Hard Matching Precision	Hard Matching Recall	Hard Matching F1	Soft Matching Precision	Soft Matching Recall	Soft Matching F1
LiberalEE	0.1342	0.2613	0.1773	0.3474	0.5340	0.4209
VASE	0.0926	0.1436	0.1125	0.2581	0.4274	0.3218
ODEE	0.1241	0.3076	0.1768	0.3204	0.4862	0.3862
CLEVE	0.1363	0.2716	0.1815	0.3599	0.5712	0.4415
ROLEPRED (BERT)	0.2128	0.4582	0.2906	0.4188	0.6896	0.5211
ROLEPRED (T5)	**0.2552**	**0.6461**	**0.3659**	**0.4591**	**0.7079**	**0.5570**
- RoleMerge	0.2233	0.6962	0.3381	0.4234	0.7677	0.5457
- RoleMerge - RoleFilter	0.1928	0.6582	0.2983	0.4188	0.7084	0.5264
Human	0.6098	0.8270	0.7020	0.7365	0.8732	0.7990

Relation Extraction w/o Golden Roles

Models	P	R	F1
LiberalEE	0.2009	0.2941	0.2387
VASE	0.2123	0.3257	0.2570
ODEE	0.2402	0.3712	0.2917
CLEVE	0.3529	0.3890	0.3701
ROLEPRED (BERT)	0.4170	0.4333	0.4250
ROLEPRED (Roberta)	**0.4131**	**0.5774**	**0.4817**
- RoleMerge	0.3855	0.6187	0.4750
- RoleMerge - RoleFilter	0.4397	0.5001	0.4679
ROLEPRED (Gold Roles)	0.6664	0.4948	0.5679

Example of the generated relation types



Extracted Results by RolePred and baselines

Output of RolePred

Victims	Maura Binkley and Nancy Van Vessem
State	Florida
Date	November 2, 2018
Killer	Scott Paul Beierle
Place	The yoga studio
Time	5:37 p.m. EDT
Duration	three and a half minutes
Motive	hatred of women
Target	Tallahassee Hot Yoga, a yoga studio
Year	2018

Output of ODEE

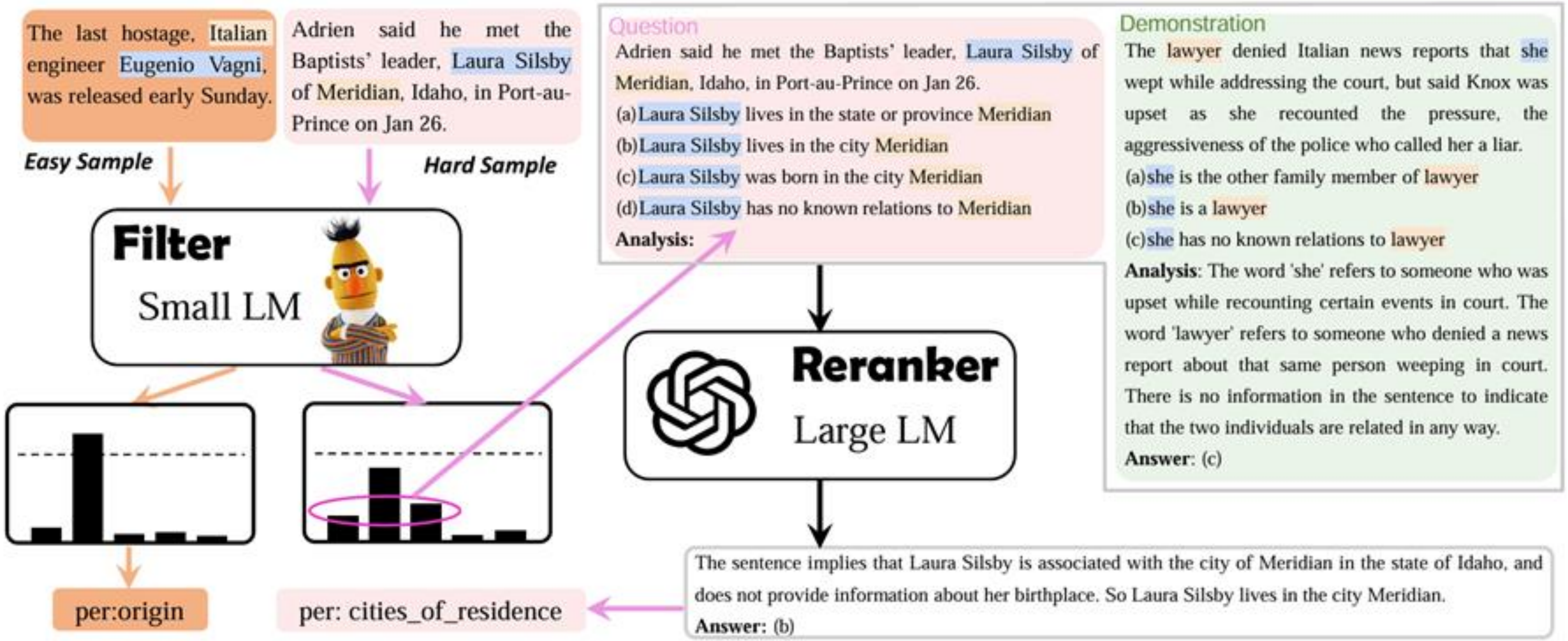
Agent	The gunman
Patient	six women

Output of CLEVE

Agent	Scott Paul Beierle
Patient	six women
Time	2018

LLM as Rerankers for Hard Samples

- Proposed a framework where LLMs complement SLMs by handling challenging samples that SLMs struggle with (based on confidence) SLMs as filters and LLMs as rerankers



Yubo Ma, Yixin Cao, YongChing Hong, Aixin Sun, "Large Language Model Is Not a Good Few-shot Information Extractor, but a Good Reranker for Hard Samples!", EMNLP'23 Findings

Experimental Results

- Effectiveness of the proposed filter-then-rerank paradigm
- Performance gains are nearly orthogonal to SLM ensemble

		FewNERD (NER)			TACREV (RE)			ACE (ED)		
		5-shot	10-shot	20-shot	20-shot	50-shot	100-shot	5-shot	10-shot	20-shot
LLM	CODEX	53.8(0.5)	54.0(1.4)	55.9(0.5)	59.1(1.4)	60.3(2.4)	62.4(2.6)	47.1(1.2)	47.7(2.8)	47.9(0.5)
	InstructGPT	53.6(–)	54.6(–)	57.2(–)	60.1(–)	58.3(–)	62.7(–)	52.9(–)	52.1(–)	49.3(–)
	GPT-4	-	-	57.8(–)	-	-	59.3(–)	-	-	52.1(–)
SLM	Previous SoTA	59.4(1.5)	61.4(0.8)	61.9(1.2)	62.4(3.8)	68.5(1.6)	72.6(1.5)	55.1(4.6)	63.9(0.8)	65.8(2.0)
	+ Ensemble (S)	59.6(1.7)	61.8(1.2)	62.6(1.0)	64.9(1.5)	71.9(2.2)	74.1(1.7)	56.9(4.7)	64.2(2.1)	66.5(1.7)
	+ Rerank (S)	59.4(1.5)	61.0(1.7)	61.5(1.7)	64.2(2.3)	70.8(2.3)	74.3(2.2)	56.1(0.3)	64.0(1.0)	66.7(1.7)
SLM + LLM	**Vicuna-13B**									
	+ Rerank (L)	60.0(1.8)	61.9(2.1)	62.2(1.4)	65.2(1.4)	70.8(1.6)	73.8(1.7)	56.9(4.0)	63.5(2.7)	66.0(2.6)
	+ Ensemble (S) + Rerank (L)	59.9(0.7)	62.1(0.7)	62.8(1.1)	66.5(0.5)	73.6(1.4)	75.0(1.5)	57.9(5.2)	64.4(1.2)	66.2(2.4)
	InstructGPT									
	+ Rerank (L)	60.6(2.1)	62.7(0.8)	63.3(0.6)	66.8(2.6)	72.3(1.4)	75.4(1.5)	57.8(4.6)	65.3(1.7)	67.3(2.2)
	+ Ensemble (S) + Rerank (L)	**61.3**(1.9)	**63.2**(0.9)	**63.7**(1.8)	**68.9**(1.3)	**74.8**(1.3)	**76.8**(1.2)	59.5(3.7)	65.3(1.9)	67.8(2.1)
	GPT-4									
	+ Rerank (L)	60.8(2.3)	62.6 (2.7)	63.0(1.3)	65.9(2.7)	72.3(0.3)	74.5(1.5)	59.6(2.9)	64.9(2.5)	67.1(2.5)
	+ Ensemble (S) + Rerank (L)	61.1(2.2)	62.8(0.9)	63.6(1.2)	68.6(1.3)	73.9(1.4)	75.9(2.4)	**60.9**(3.9)	**65.6**(1.5)	**67.8**(1.7)

Outline

- Span Detection
- Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring
  - Knowledge Graph Construction
    - General KG
    - Domain-specific KG
    - Theme-specific KG
  - Database Population

LLM-assisted Construction

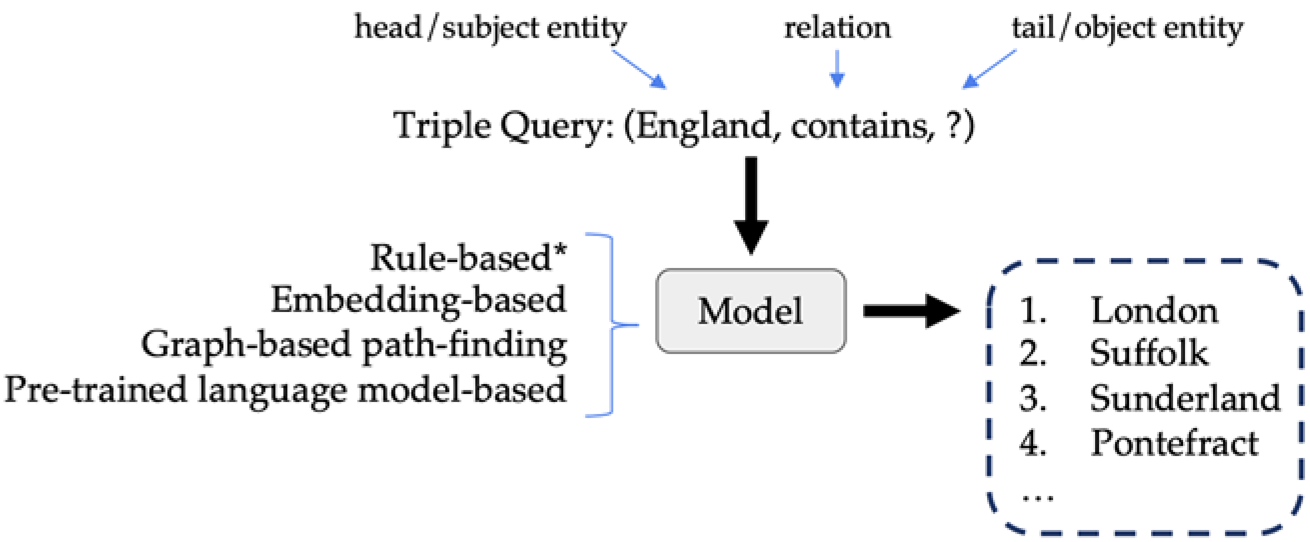
- **Knowledge Graph GPT**
  - **Using prompting methods to construct knowledge graphs from plain text.**

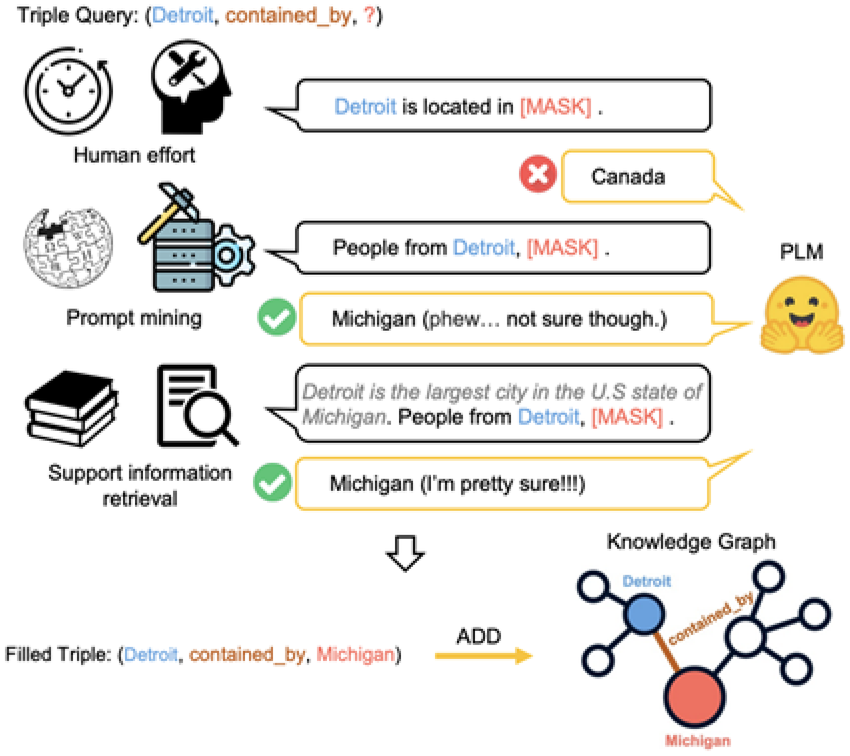
You are given a prompt. Extrapolate as many relationships as you can from the prompt and generate tuples like (source, relation, target). Make sure there are always source, relation and target in the tuple.
Example:
prompt: John knows React, Golang, and Python. John is good at Software Engineering and Leadership
tuple: (John, knows, React); (John, knows, Golang); (John, knows, Python); (John, good_at, Software_Engineering); (John, good_at, Leadership);
prompt: Bob is Alice's father. Alice has one brother John.
tuple: (Bob, father_of, Alice); (John, brother_of, Alice)
prompt: $prompt
tuple:

https://github.com/iAmmarTahir/KnowledgeGraphGPT

Knowledge Graph Completion

- Knowledge graph completion
  - Construction aims to build a foundational representation of knowledge
  - Completion seeks to fill gaps and extend the graph's coverage





Pengcheng Jiang, et al., "Text-Augmented Open Knowledge Graph Completion via Pre-Trained Language Models", ACL'23 Findings

TagReal: Knowledge Graph Completion

- TagReal
  - Automatically generates quality query prompts and retrieves support information from large text corpora to probe knowledge from PLM for KG completion.

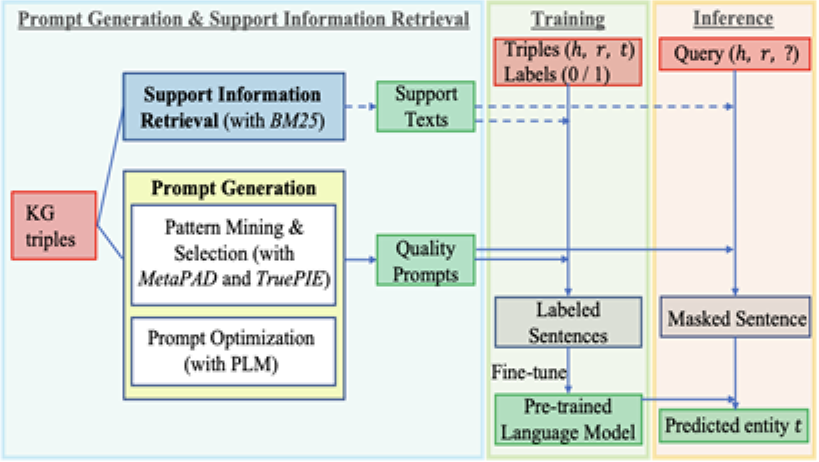
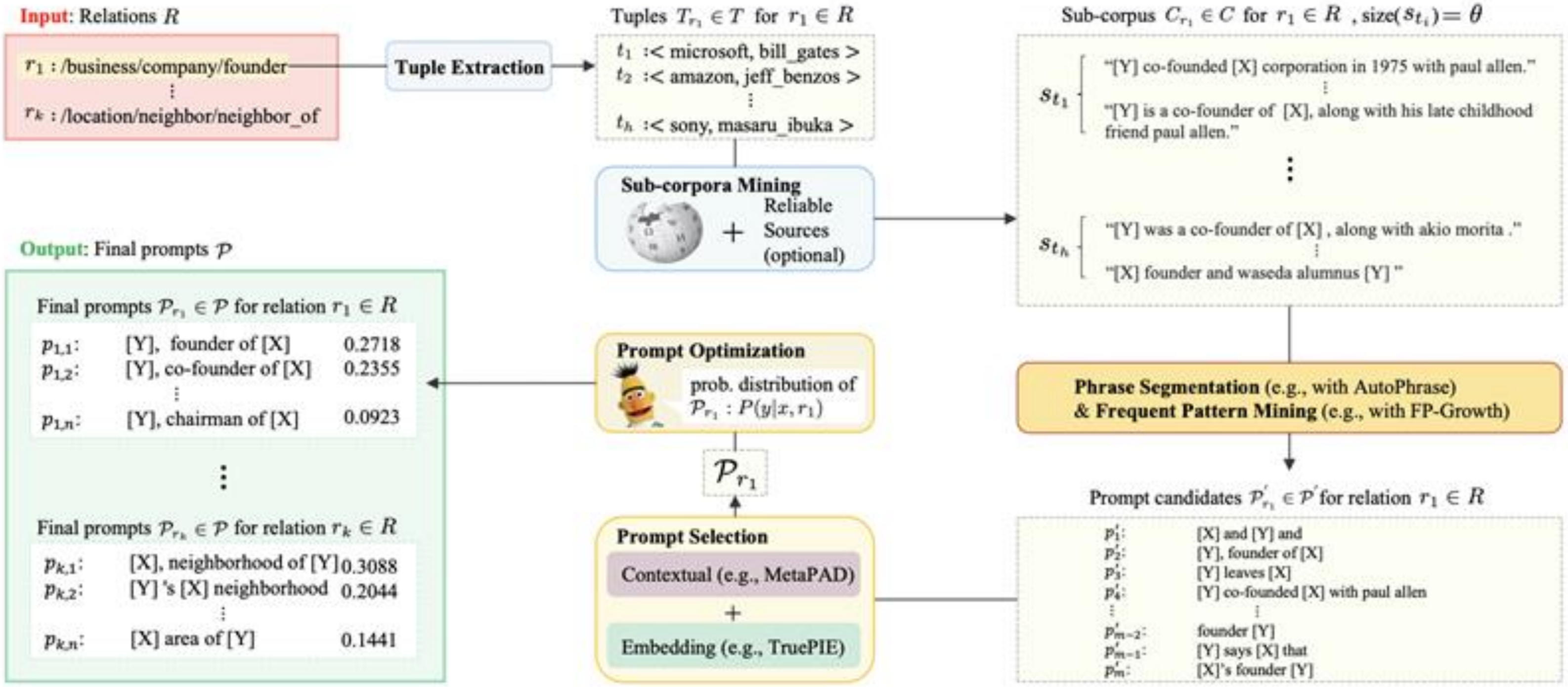


Figure 2: **TAGREAL Framework**. The input and output of each phase are highlighted by red and green, respectively. The dotted arrow indicates the optional process.



TagReal: Knowledge Graph Completion

- Experimental Results

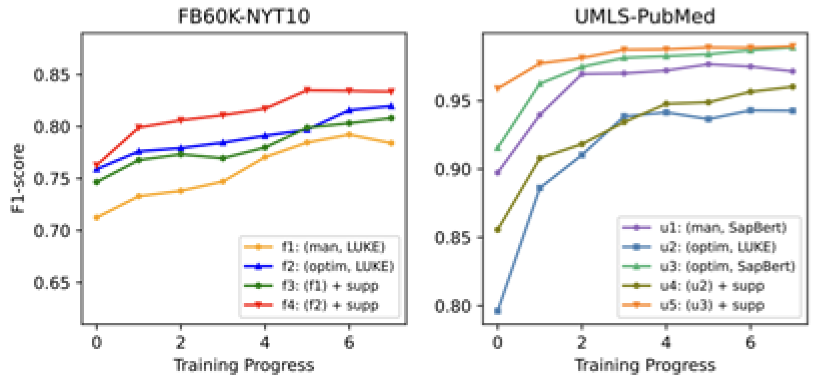
Results on FB60K-NYT10

	Model	20% Hits@5	20% Hits@10	20% MRR	50% Hits@5	50% Hits@10	50% MRR	100% Hits@5	100% Hits@10	100% MRR
KGE-based	TransE (Bordes et al., 2013)	29.13	32.67	15.80	41.54	45.74	25.82	42.53	46.77	29.86
	DisMult (Yang et al., 2014)	3.44	4.31	2.64	15.98	18.85	13.14	37.94	41.62	30.56
	ComplEx (Trouillon et al., 2016a)	4.32	5.48	3.16	15.00	17.73	12.21	35.42	38.85	28.59
	ConvE (Dettmers et al., 2018)	29.49	33.30	24.31	40.10	44.03	32.97	50.18	54.06	40.39
	TuckER (Balažević et al., 2019)	29.50	32.48	24.44	41.73	45.58	33.84	51.09	54.80	40.47
	RotatE (Sun et al., 2019)	15.91	18.32	12.65	35.48	39.42	28.92	**51.73**	55.27	**42.64**
Text&KGE-based	RC-Net (Xu et al., 2014)	13.48	15.37	13.26	14.87	16.54	14.63	14.69	16.34	14.41
	TransE+Line (Fu et al., 2019)	12.17	15.16	4.88	21.70	25.75	8.81	26.76	31.65	10.97
	JointNRE (Han et al., 2018)	16.93	20.74	11.39	26.96	31.54	21.24	42.02	47.33	32.68
RL-based	MINERVA (Das et al., 2017)	11.64	14.16	8.93	25.16	31.54	22.24	43.80	44.70	34.62
	CPL (Fu et al., 2019)	15.19	18.00	10.87	26.81	31.70	23.80	43.25	49.50	33.52
PLM-based	PKGC (Lv et al., 2022)	35.77	43.82	28.62	41.93	46.70	31.81	41.98	52.56	32.11
	TagReal (our method)	**45.59**	**51.34**	**35.41**	**48.98**	**55.64**	**38.03**	50.85	**60.64**	38.86

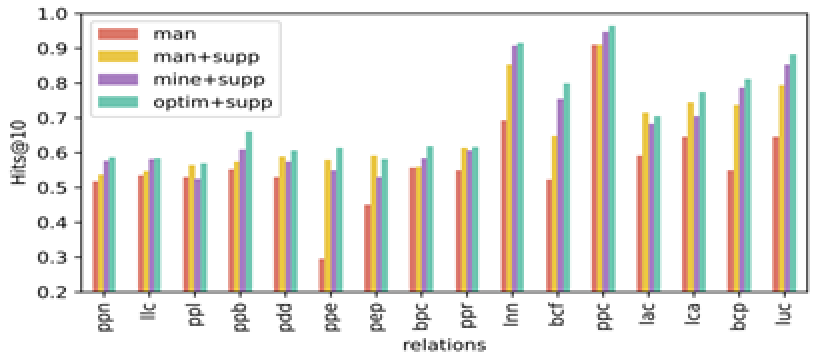
Table 1: **Performance comparison of KG completion on FB60K-NYT10 dataset**. Results are averaged values of ten independent runs of head/tail entity predictions. The highest score is highlighted in **bold.**

- TAGREAL has superb performance even with limited training data, outperforming existing embedding-based, graph-based, and PLM-based methods.

Performance Variation of F1-score



Relation-wise Hits@10 on FB60K-NYT10



Outline

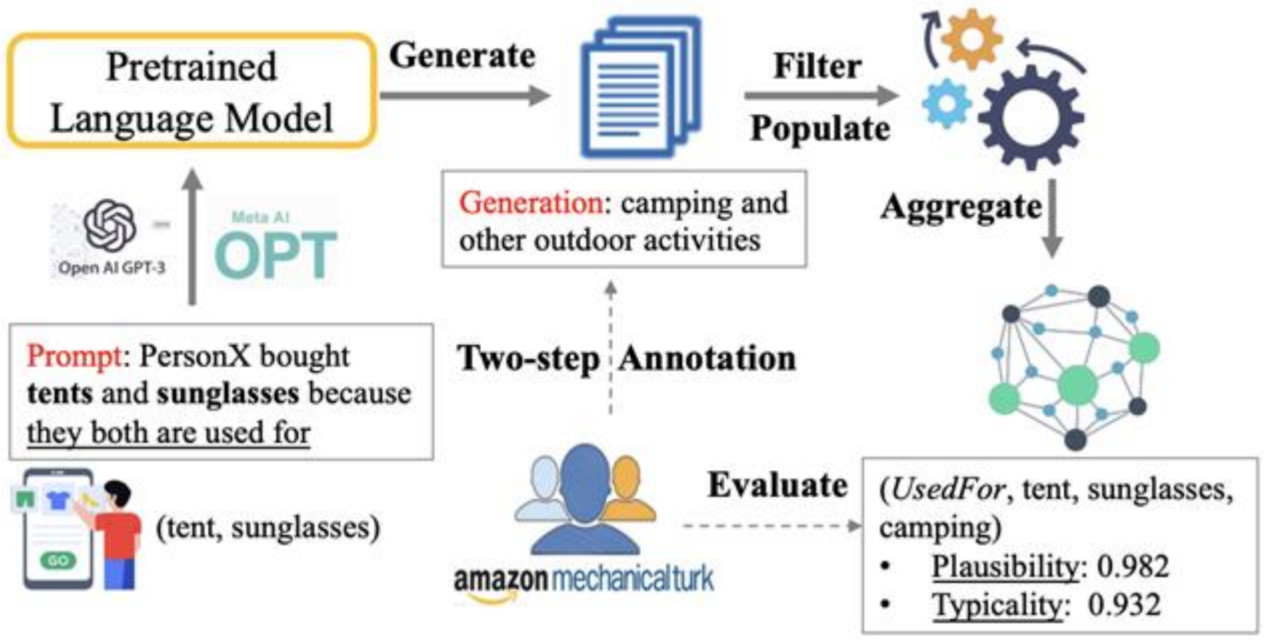
- Span Detection
- Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring
  - Knowledge Graph Construction
    - General KG
    - Domain-specific KG
    - Theme-specific KG
  - Database Population

Domain-specific KG

- Define the domain scope (biomedical, music, chemistry)
- Design the ontology by domain expertise
  - classes, properties, relations, and hierarchies
- Identify data source:
  - relevant databases, publications, and datasets (both structured and unstructured data sources)
- Entity recognition and linking
- Relationship extraction
- Knowledge integration
  - merge information from various sources
  - resolve conflicts and inconsistencies
  - Validation and refinement

Domain-specific KG

- FolkScope (E-commerce)
  - leverage the generation power of LLMs and human in-the-loop annotation to semi-automatically construct KG



- STEP1: use LLMs first generate intention assertions via e-commerce specific prompts to explain shopping behaviors
- STEP2: annotate plausibility and typicality labels of sampled intentions as training data in order to populate human judgments to all automatic generations.
- STEP3: propose pattern mining and conceptualization to form more condensed and abstract knowledge

Changlong Yu, et al., "FolkScope: Intention Knowledge Graph Construction for E-commerce Commonsense Discovery", ACL'23 Findings

Domain-specific KG

- Experimental Results: FolkScope (E-commerce)

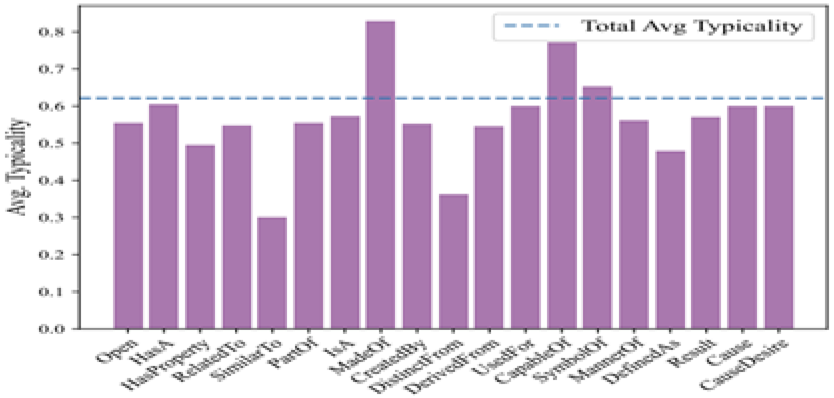
Relation-wise Accuracy

Relation	Acc. Rate	# Edges	# Tails	Avg. Length
Open	87.54%	703,059	151,748	7.86
HasA	94.08%	710,331	68,516	5.53
HasProperty	79.13%	317,938	133,877	5.00
RelatedTo	91.89%	571,918	130,551	3.08
SimilarTo	86.35%	685,737	18,603	3.53
PartOf	79.60%	674,928	114,983	4.36
IsA	89.05%	591,037	98,262	3.82
MadeOf	90.05%	528,289	70,246	5.06
CreatedBy	95.15%	267,459	74,920	3.93
DistinctFrom	91.74%	861,929	80,295	4.66
DerivedFrom	85.54%	444,131	61,696	4.90
UsedFor	91.79%	630,462	45,206	2.58
CapableOf	87.73%	681,480	101,170	5.23
SymbolOf	78.04%	809,196	52,075	3.46
MannerOf	89.44%	371,892	122,829	4.38
DefinedAs	85.59%	288,411	151,986	6.31
Result	44.79%	568,523	166,018	8.80
Cause	80.50%	696,392	185,042	7.06
CauseDesire	67.23%	833,524	155,422	5.61
Total	83.40%	11,236,636	1,874,782	5.02

Matched KG subsets

Knowledge Graph	Clothing		Electronics	
	# Edges	Coverage	# Edges	Coverage
Matched Knowledge Graph	432,119	79.83%	117,836	82.40%
+ Plau. >0.5	323,263	79.83%	78,908	82.40%
+ Plau. >0.5 and Typi. >0.5	141,422	79.67%	40,978	80.20%
+ Plau. >0.9	269,210	79.83%	58,013	82.39%
+ Plau. >0.9 and Typi. >0.9	103,262	79.36%	27,288	76.94%

Relation-wise Human evaluation on typicality

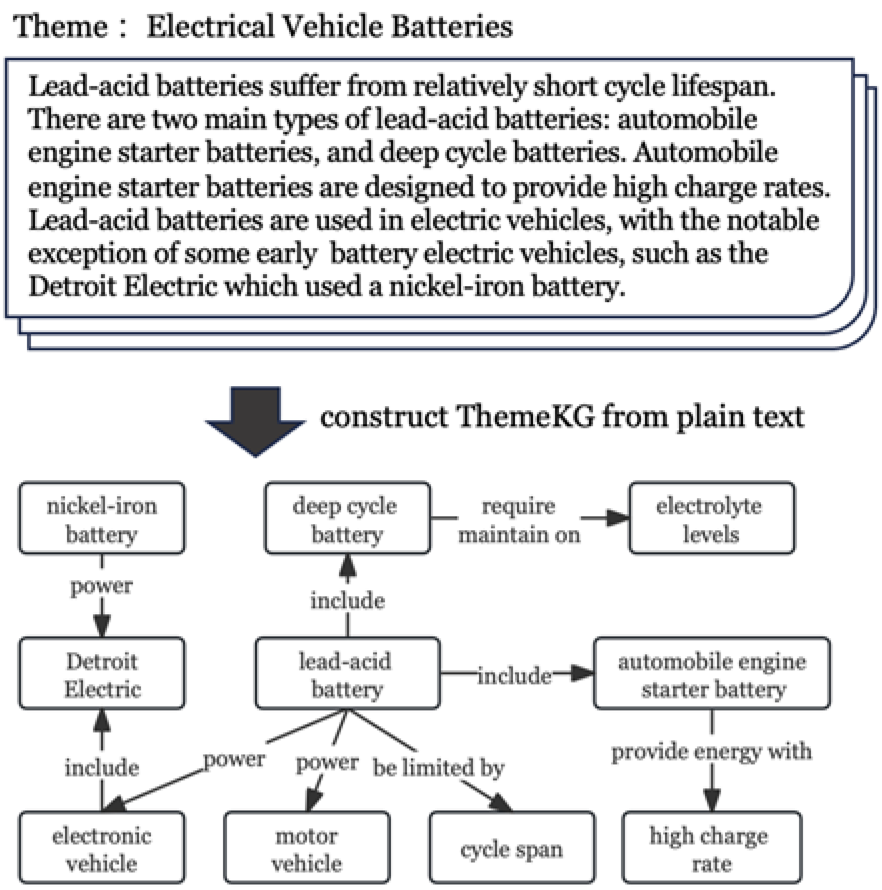


Outline

- Span Detection
- Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring
  - Knowledge Graph Construction
    - General KG
    - Domain-specific KG
    - Theme-specific KG
  - Database Population

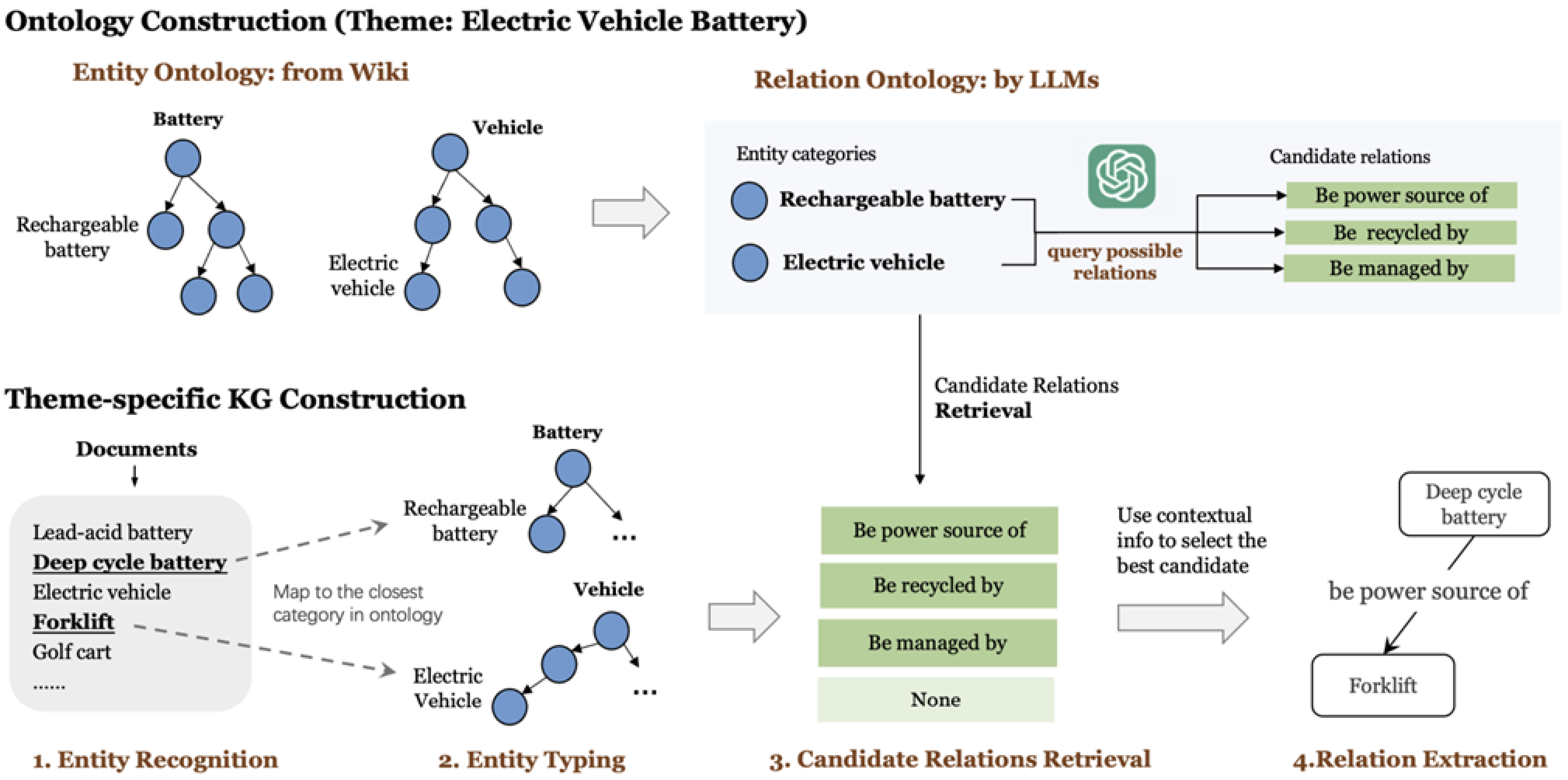
Theme-specific KG Construction

- **Task Setting**
  - Given a set of documents of a specific theme, extract the entities and corresponding relations to construct a knowledge graph
- **Challenges**
  - Open vocabulary
  - No annotation
  - Small set of documents.
  - Directly use LLMs easily leads to unclear/wrong relations (*have, have due to, lit up in colors of*) and inaccurate entities (*two main types*)



Linyi Ding, Sizhe Zhou, Jinfeng Xiao, Jiawei Han, "Automated Construction of Theme-specific Knowledge Graphs", Arxiv'24

Theme-specific KG Construction

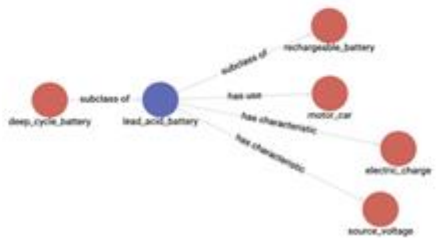


Theme-specific KG Construction

Main Results

Dataset	Method	Entity Metric			Triple Metric			Theme Metric
		Recall	Precision	F1-score	Recall	Precision	F1-score	Coherence
EVB	OpenIE [37]	0.62	0.36	0.46	0.13	0.24	0.17	0.46
	REBEL [24]	0.22	**0.80**	0.35	0.11	**0.80**	0.19	0.80
	IMoJIE [31]	0.44	0.49	0.46	0.26	0.45	0.33	0.78
	KG-GPT [47]	0.72	0.69	0.70	0.67	0.64	0.65	0.95
	GPT-4 [1]	0.68	0.71	0.69	0.64	0.65	0.64	**0.97**
	TKGCon (w/o ontology)	/	/	/	0.67	0.57	0.62	0.92
	TKGCon	**0.92**	**0.80**	**0.86**	**0.78**	0.73	**0.75**	**0.97**
HAI	OpenIE [37]	0.52	0.28	0.36	0.17	0.22	0.19	0.35
	REBEL [24]	0.16	0.87	0.27	0.15	0.75	0.25	0.75
	IMoJIE [31]	0.33	0.39	0.36	0.25	0.31	0.28	0.83
	KG-GPT [47]	0.84	0.79	0.81	0.72	0.69	0.70	0.91
	GPT-4 [1]	0.82	0.80	0.83	0.70	0.72	0.71	**0.93**
	TKGCon (w/o ontology)	/	/	/	0.75	0.62	0.68	0.88
	TKGCon	**0.90**	**0.88**	**0.89**	**0.81**	**0.75**	**0.78**	0.92

ThemeKG



WikiData

Case Study on Applications of TKGCon

Question	Which countries support Hamas or condemn Israel in the Hamas attack on Israel in Oct 2023?
Vanilla GPT4	I'm sorry, but as of my knowledge cutoff date in march 2023, i do not have information on specific events that occurred in october 2023.
RAG+GPT4	In the **Middle East** and **North Africa**, most countries either condemned Israel or offered full-throated support to Hamas. **North Korea** is also mentioned as condemning Israel.
TKG+GPT4	1. **Iran**, 2. **Persian Gulf countries**, 3. **North Korea**, 4. most **Middle East countries**, 5. most **North Africa countries**

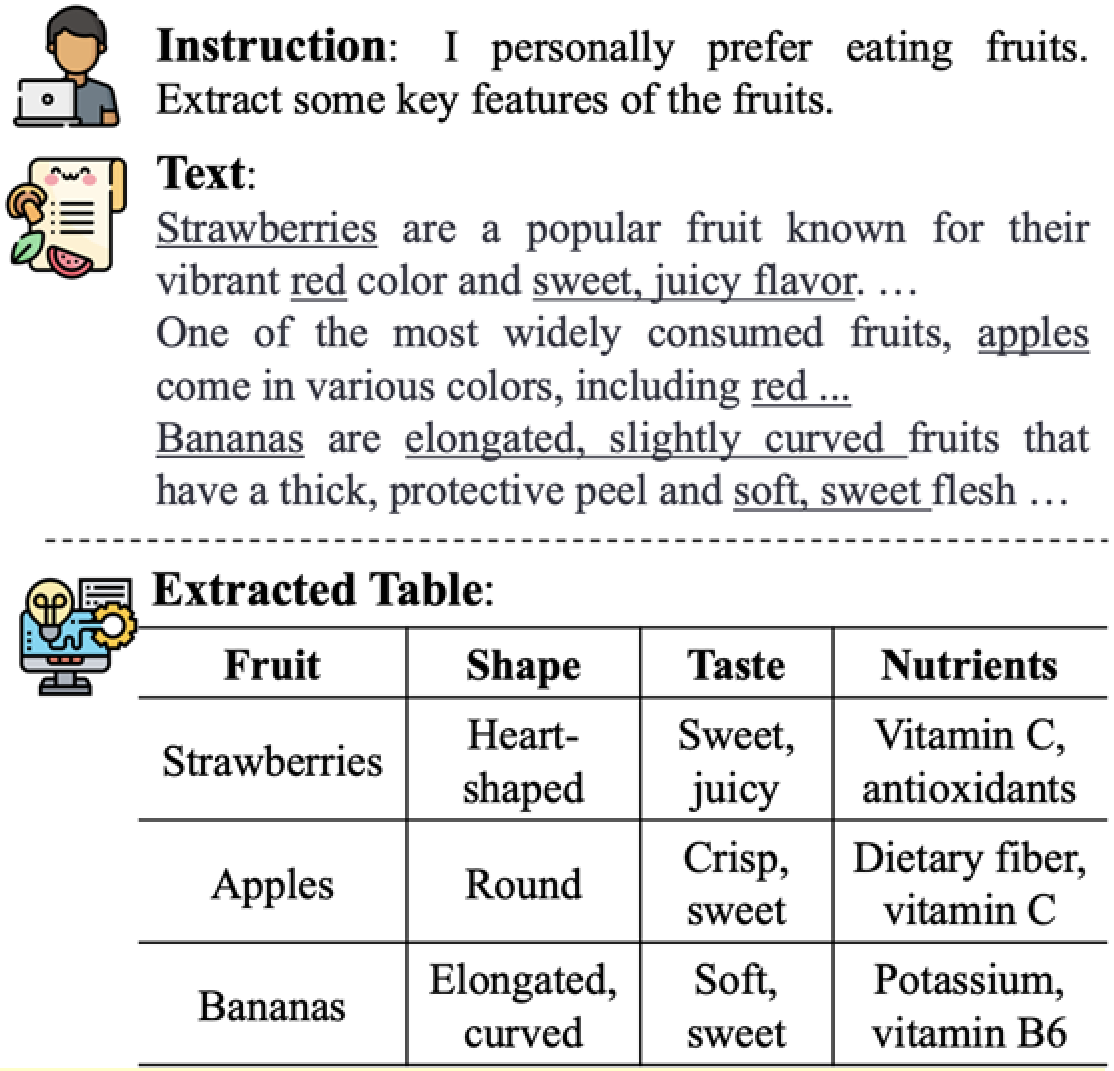
Comparison of knowledge in ThemeKG and Wikidata

Outline

- Span Detection
- Entity Typing
- Relation Extraction
- Comprehensive Knowledge Structuring
  - Knowledge Graph Construction
  - Database Population

On-demand Information Extraction

- Average users may have IE needs that don’t align well with existing well-defined tasks.
- Task: On-demand IE
  - Given a user instruction and the related text, extract the sought-after information into a structured table format

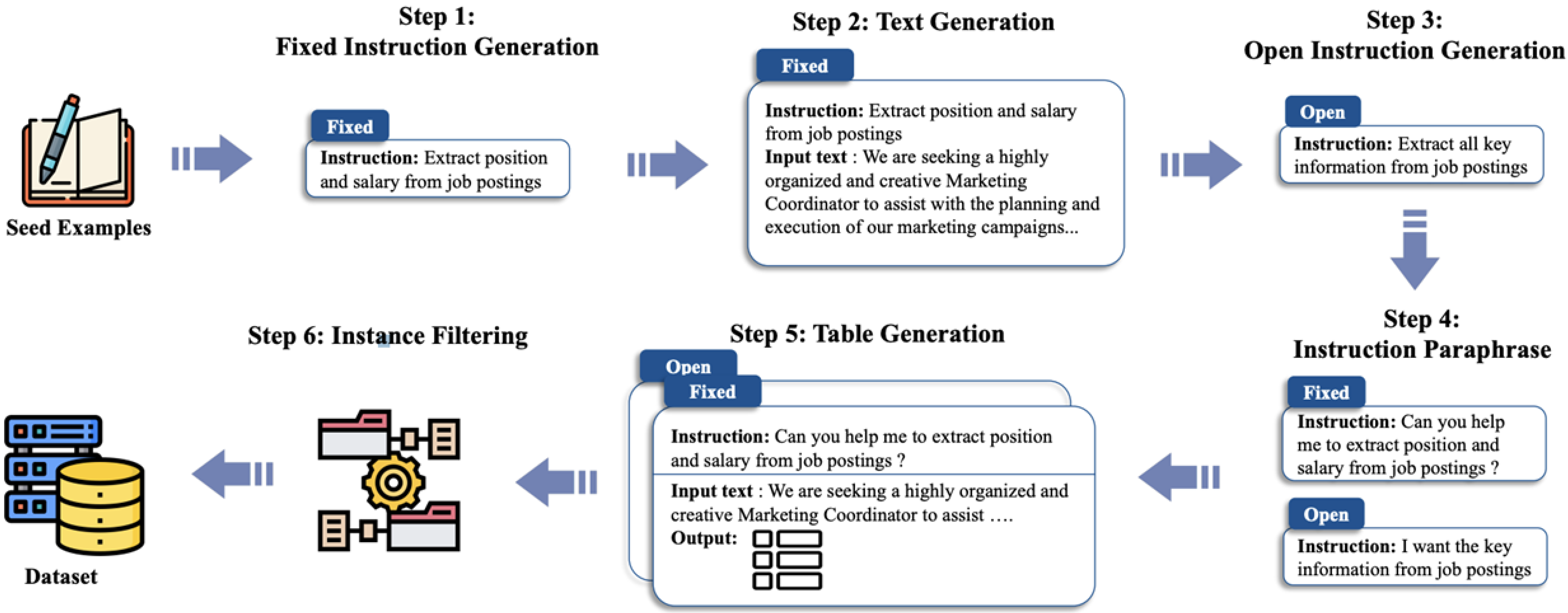


Fruit	Shape	Taste	Nutrients
Strawberries	Heart-shaped	Sweet, juicy	Vitamin C, antioxidants
Apples	Round	Crisp, sweet	Dietary fiber, vitamin C
Bananas	Elongated, curved	Soft, sweet	Potassium, vitamin B6

Jiao, Yizhu, et al., “Instruct and Extract: Instruction Tuning for On-Demand Information Extraction.” EMNLP’23.

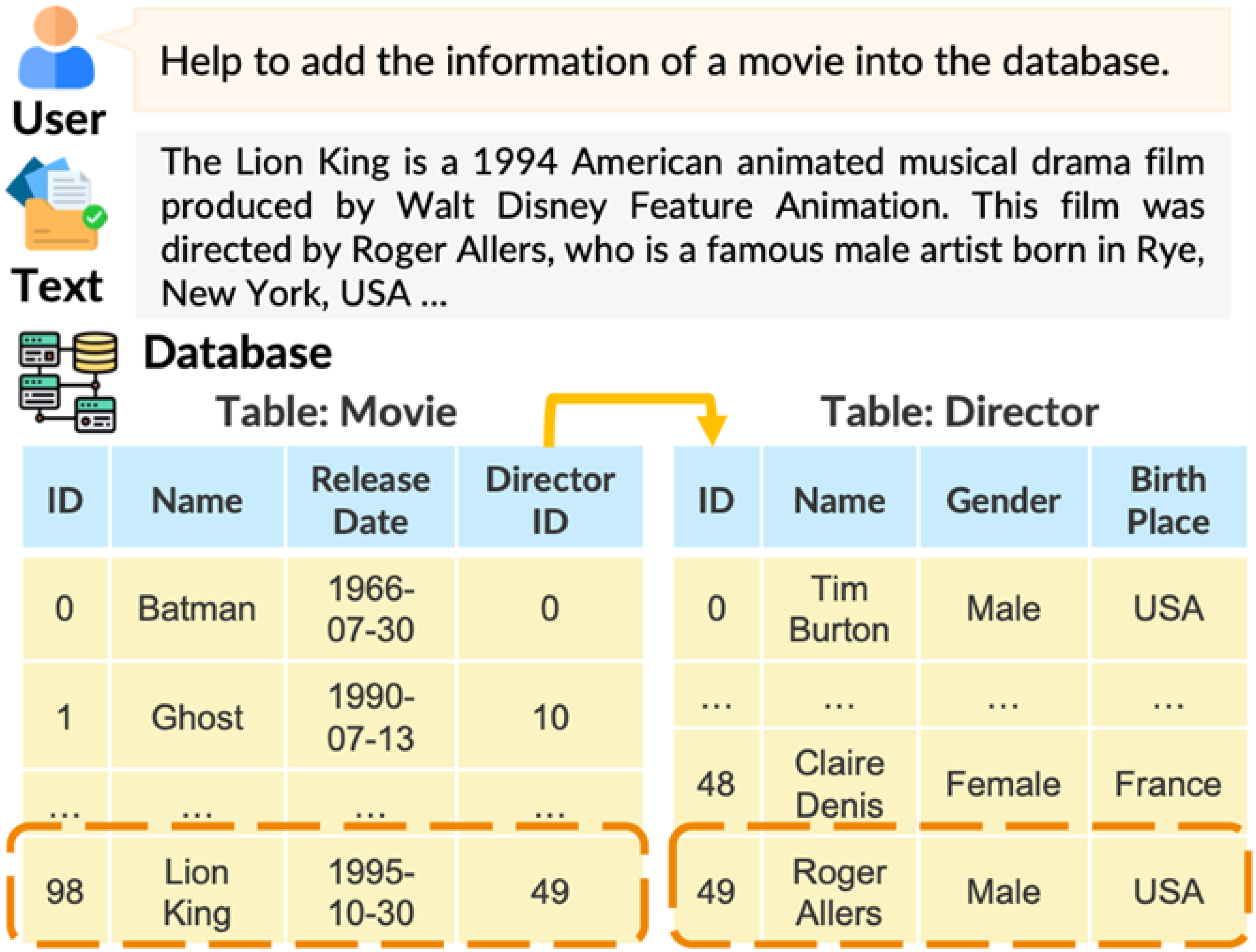
Method – Instruction Tuning for LLMs

❑ Synthetic data generation + Instruction tuning LLaMa-7b



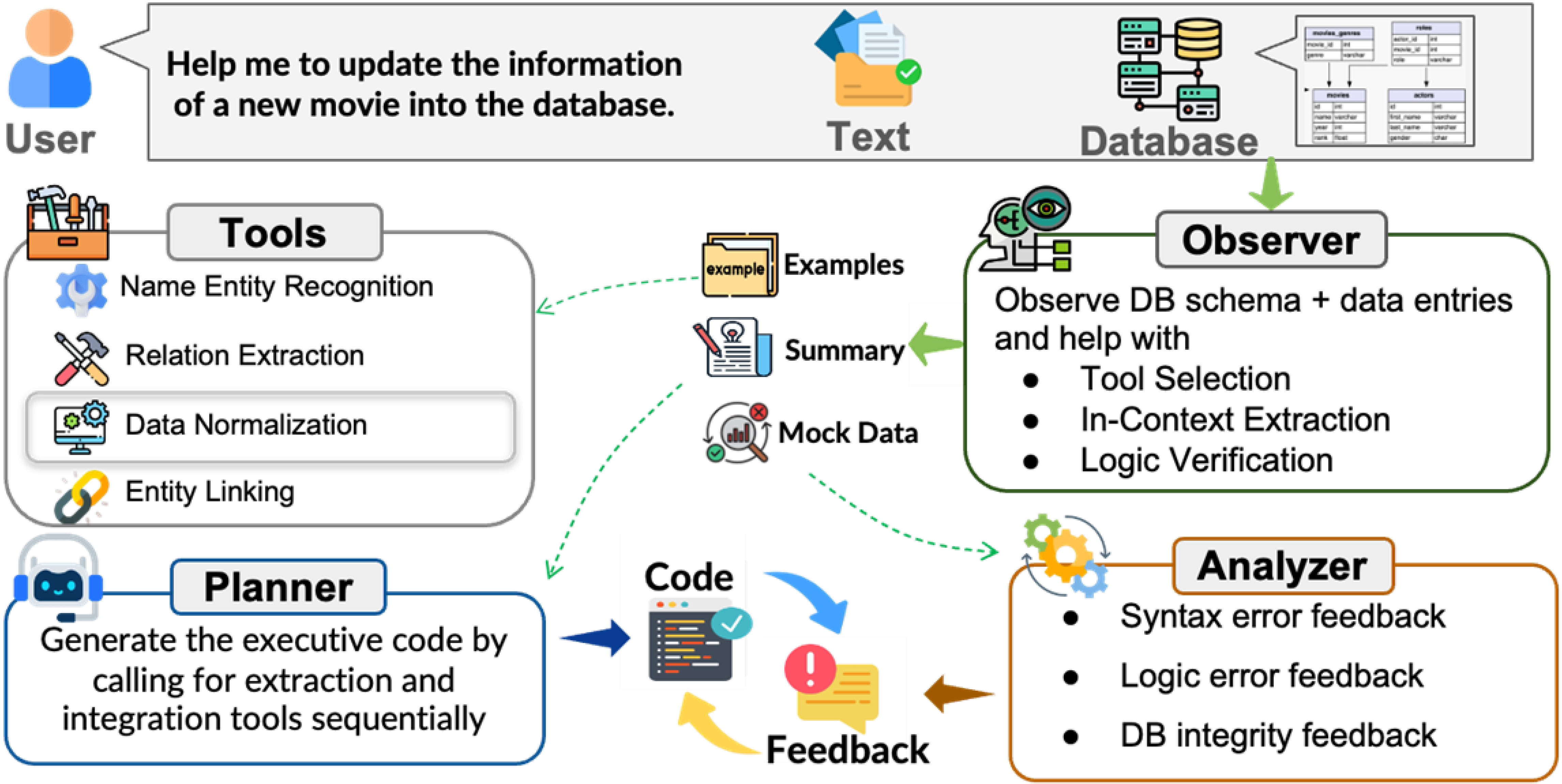
Integration-Aware Information Extraction

- ❑ Task: Text2Database
- ❑ Input
  - ❑ Database with its schema
  - ❑ User Instruction
  - ❑ A set of text documents
- ❑ Output
  - ❑ Populated database



Jiao, Yizhu, et al., "Text2DB: Integration-Aware Information Extraction with Large Language Model Agents." ACL'24 Findings.

Method – LLM based Agent



References

- Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, Huajun Chen, "KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction", WWW'22
- Yew Ken Chia, Lidong Bing, Soujanya Poria, Luo Si, "RelationPrompt: Leveraging Prompts to Generate Synthetic Data for Zero-Shot Relation Triplet Extraction", ACL'22 Findings
- Linyi Ding, Sizhe Zhou, Jinfeng Xiao, Jiawei Han, "Automated Construction of Theme-specific Knowledge Graphs", arXiv'24
- Xiaotao Gu , Zihan Wang , Zhenyu Bi , Yu Meng, Liyuan Liu, Jiawei Han, Jingbo Shang. "UCPhrase: Unsupervised Context-aware Quality Phrase Tagging." (KDD'21)
- Xiaotao Gu, Yikang Shen, Jiaming Shen, Jingbo Shang, Jiawei Han, "Phrase-aware Unsupervised Constituency Parsing" (ACL'22)
- Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, Maosong Sun, "PTR: Prompt Tuning with Rules for Text Classification", AI Open
- Pere-Lluís Huguet Cabot, Roberto Navigli, "REBEL: Relation Extraction By End-to-end Language generation", ACL'21
- Pengcheng Jiang , Shivam Agarwal , Bowen Jin , Xuan Wang, Jimeng Sun, Jiawei Han, "Text-Augmented Open Knowledge Graph Completion via Pre-Trained Language Models", ACL'23 Findings
- Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji, Jiawei Han. "Open-vocabulary argument role prediction for event extraction." EMNLP'22 Findings.
- Yizhu Jiao, Ming Zhong, Sha Li, Ruining Zhao, Siru Ouyang, Heng Ji, and Jiawei Han. "Instruct and Extract: Instruction Tuning for On-Demand Information Extraction." EMNLP 2023.
- Yizhu Jiao, Sha Li, Sizhe Zhou, Heng Ji, and Jiawei Han. " Text2DB: Integration-Aware Information Extraction with Large Language Model Agents ." ACL 2024 findings.
- Seoyeon Kim, Kwangwook Seo, Hyungjoo Chae, Jinyoung Yeo, Dongha Lee, "VerifiNER: Verification-augmented NER via Knowledge-grounded Reasoning with Large Language Models", ACL'24

References

- Yubo Ma, Yixin Cao, YongChing Hong, Aixin Sun, “Large Language Model Is Not a Good Few-shot Information Extractor, but a Good Reranker for Hard Samples!”, EMNLP’23 Findings
- Siru Ouyang, Jiaxin Huang, Pranav Pillai, Yunyi Zhang, Yu Zhang, Jiawei Han, “Ontology Enrichment for Effective Fine-grained Entity Typing”, KDD’24
- Tanay Komarlu, Minhao Jiang, Xuan Wang, Jiawei Han, “OntoType: Ontology-Guided and Pre-Trained Language Model Assisted Fine-Grained Entity Typing”, KDD’24
- Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, Jiawei Han, “EIDER: Evidence-enhanced Document-level Relation Extraction”, ACL’22 Findings
- Changlong Yu, Weiqi Wang, Xin Liu, Jiaxin Bai, Yangqiu Song, Zheng Li, Yifan Gao, Tianyu Cao, and Bing Yin. “FolkScope: Intention Knowledge Graph Construction for E-commerce Commonsense Discovery”, ACL’23 Findings
- Kai Zhang, Bernal Jiménez Gutiérrez, Yu Su, “Aligning Instruction Tasks Unlocks Large Language Models as Zero-Shot Relation Extractors”, ACL'23 Findings
- Yu Zhang, Yu Meng, Xuan Wang, Sheng Wang, Jiawei Han. "Seed-guided topic discovery with out-of-vocabulary seeds." NAACL’22.
- Yu Zhang, Yunyi Zhang, Yucheng Jiang, Martin Michalski, Yu Deng, Lucian Popa, ChengXiang Zhai, Jiawei Han. "Entity Set Co-Expansion in StackOverflow." IEEE Big Data 2022
- Ming Zhong, Siru Ouyang, Minhao Jiang, Vivian Hu, Yizhu Jiao, Xuan Wang, Jiawei Han, “ReactIE: Enhancing Chemical Reaction Extraction with Weak Supervision”, ACL’23 Findings
- Sizhe Zhou, Suyu Ge, Jiawei Han, “Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns”, ECMLPKDD’23
- Sizhe Zhou, Yu Meng, Bowen Jin, Jiawei Han, “Grasping the Essentials: Tailoring Large Language Models for Zero-Shot Relation Extraction”, arXiv’24
- Wenxuan Zhou, Muhao Chen, “An Improved Baseline for Sentence-level Relation Extraction”, AACL’22

Q&A

Tutorial Website: