The State of the Art Large Language Models for Knowledge Graph Construction from Text: Techniques, Tools, and Challenges

Nandana Mihindukulasooriya
Senior Research Scientist | IBM Research

Agenda

1

Relation Extraction Task

Relation extraction and related tasks in the literature.

2

RE Benchmarks

Summary of relation extraction benchmarks.

3

RE Approches

Methods and approaches for relation extraction focused on LLMs.

KGC

Relation Extraction and Related Tasks

Relation Extraction

Irene Morgan, who was born and raised in Baltimore, lived on Long Island.

Named Entity Recognition

Irene Morgan, who was born and raised in **Baltimore**, lived on **Long Island**.

[PERSON] **[PLACE]** **[PLACE]**

Relation Extraction

Irene Morgan, who was born and raised in **Baltimore**, lived on Long Island.

per:city_of_birth

(Binary) Relation Extraction

Irene Morgan, who was born and raised in **Baltimore**, lived on Long Island.

per:city_of_birth

<**Head Entity**, Relationship, **Tail Entity**>

<**Subject**, Relationship, **Object**>

Relationship is selected from a set of predefined canonical relations.

Open Information Extraction (OpenIE)

KGC 2024 took place in **New York**.

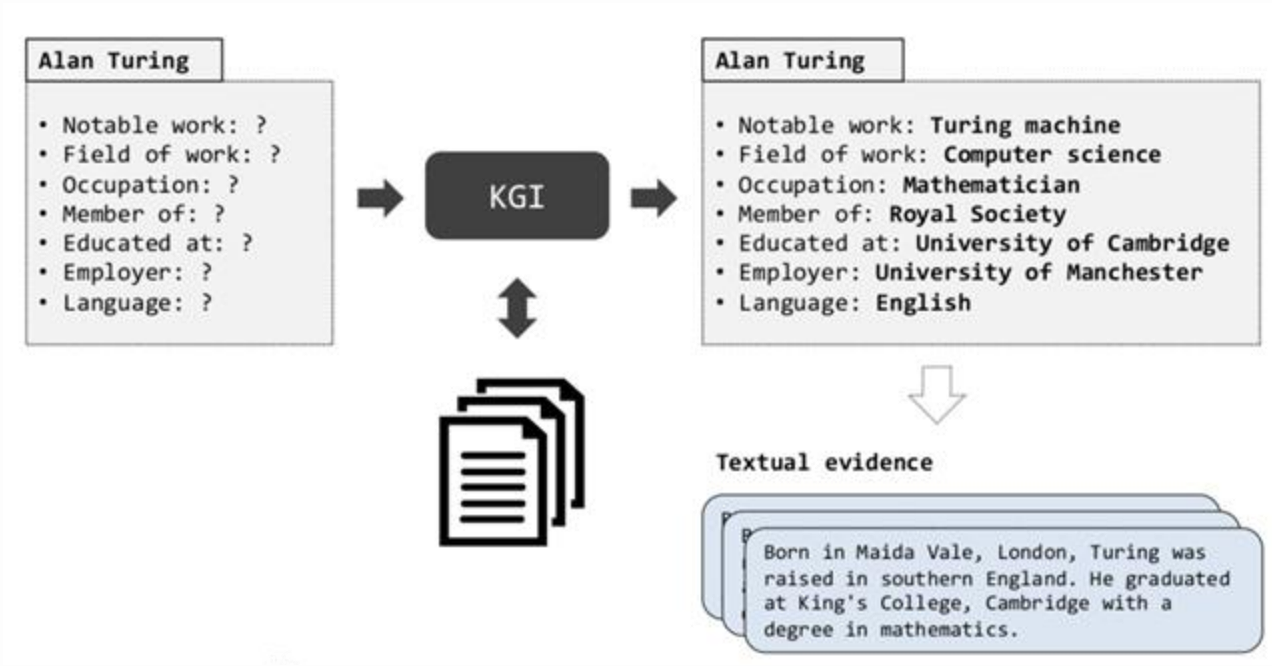
(**KGC 2024**, took place in, **New York**)

- Relations are not predefined, automatically discovered in text.
- A large number of sparse and diverse relations
- Need to further steps of clustering, canonicalization, alignment to map to a set of KG relations.

Open Information Extraction from the Web. Banko et al. IJACAI 2007.

KGC

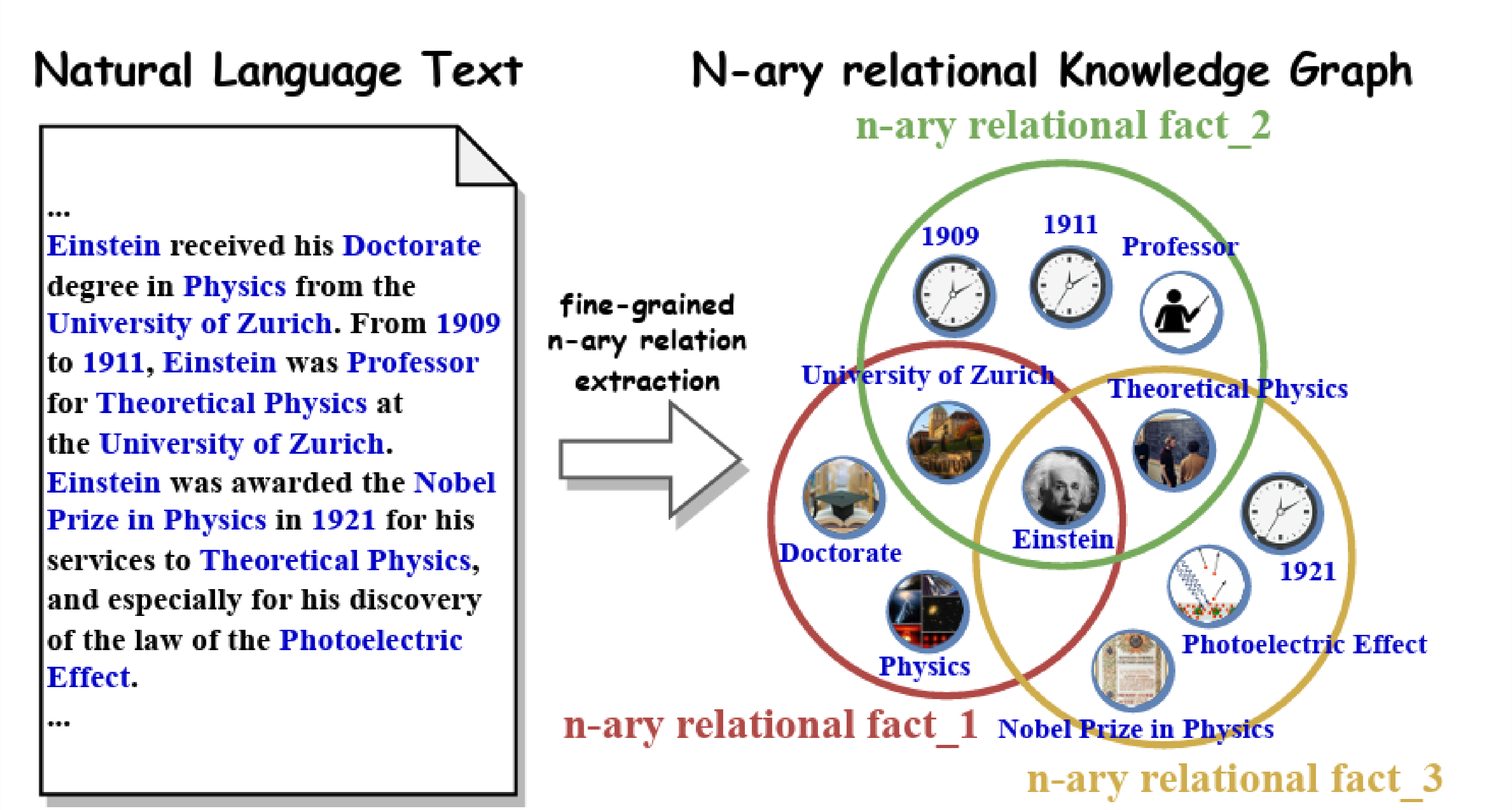
Slot Filling / Knowledge Base Population



- Mapping unstructured data (text) to a structured format (e.g., filling a database or knowledge graph)
- Focus each entity at a time
- Filling predefined slots (attributes) associated with specific entities

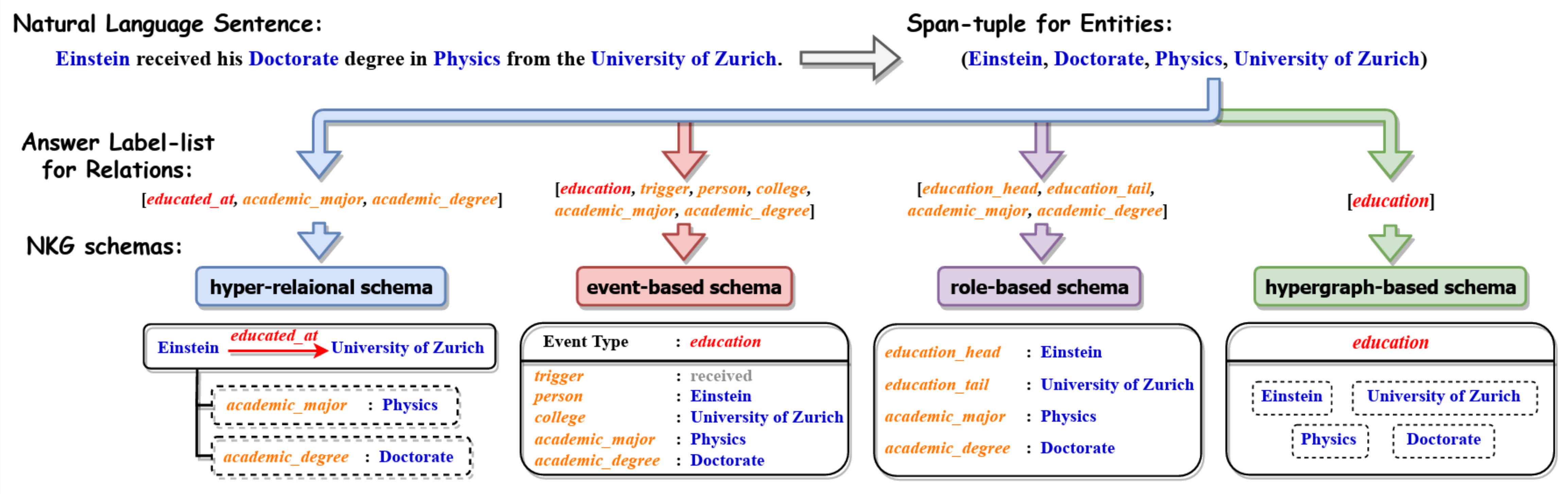
Robust Retrieval Augmented Generation for Zero-shot Slot Filling. Glass et al. EMNLP 2021.

N-ary relation extraction



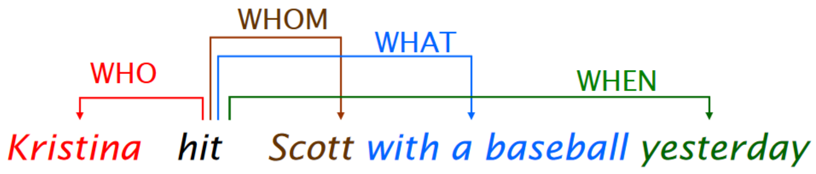
Text2nkg: Fine-grained n-ary relation extraction for n-ary relational knowledge graph construction. Haoran et al. NeurIPS 2024.

N-ary relation extraction



Text2nkg: Fine-grained n-ary relation extraction for n-ary relational knowledge graph construction. Haoran et al. NeurIPS 2024.

Semantic Role Labelling – Thematic Roles



- Who hit Scott with a baseball?
- Whom did Kristina hit with a baseball?
- What did Kristina hit Scott with?
- When did Kristina hit Scott with a baseball?

Table 1.1: A set of widely recognized Semantic Roles

Role	Description	Examples
Agent	Initiator of action, capable of volition	**The batter** smashed the pitch into left field. **The pilot** landed the plane as lightly as a feather.
Patient	Affected by action, undergoes change of state	David trimmed **his beard**. John broke **the window**.
Theme	Entity moving, or being "located"	Paola threw **the Frisbee**. **The picture** hangs above the fireplace.
Experiencer	Perceives action but not in control	**He** tasted the delicate flavor of the baby lettuce. **Chris** noticed the cat slip through the partially open door.
Beneficiary	For whose benefit action is performed	He sliced **me** a large chunk of prime rib, and I could hardly wait to sit down to start in on it. The Smiths rented an apartment **for their son**.
Instrument	Intermediary/means used to perform an action	He shot the wounded buffalo with **a rifle**. The surgeon performed the incision with **a scalpel**.
Location	Place of object or action	There are some real monsters hiding **in the anxiety closet**. The band played **on the stage**.
Source	Starting point	The jet took off **from Nairobi**. We heard the rumor **from a friend**.
Goal	Ending point	The ball rolled **to the other end of the hall**. Laura lectured **to the class**.

Palmer, Martha, Daniel Gildea, and Nianwen Xue. Semantic Role Labeling.
Scott Wen-tau Yih, Kristina Toutanova. Automatic Semantic Role Labeling. ACL 2006.

Semantic Role Labelling – PropBank & AMR

describe.01 - *assign a label or attribute*

DESCRIBE-V NOTES: Comparison to 'call' attributive sense. Member of VNcls ch
DESCRIPTION-N NOTES: Based on nouns00080. Vn class characterize-29.2-1-1.

Aliases:

description (n.)
describe (v.)

Roles:

ARG0-PAG: *describer*
ARG1-PPT: *thing described*
ARG2-PRD: *secondary attribute, described-as*

LOC: location	CAU: cause
EXT: extent	TMP: time
DIS: discourse connectives	PNC: purpose
ADV: general purpose	MNR: manner
NEG: negation marker	DIR: direction
MOD: modal verb	

```
(d / describe-01
  :arg0 (m / man)
  :arg1 (m2 / mission)
  :arg2 (d / disaster))

The man described the mission as a disaster.
The man's description of the mission:
  disaster.
As the man described it, the mission was a
  disaster.
```

Banarescu, Laura, et al. "Abstract Meaning Representation for Sembanking." 2013.

Sentence-level Relation Extraction

- The task of extracting relationships between entities within a single sentence

Irene Morgan, who was born and raised in **Baltimore**, lived on Long Island.

per:city_of_birth

Document-level Relation Extraction

- The task of extracting relationships between entities across an entire document considering the **broader context, coreferences, long-distance dependencies**, etc.

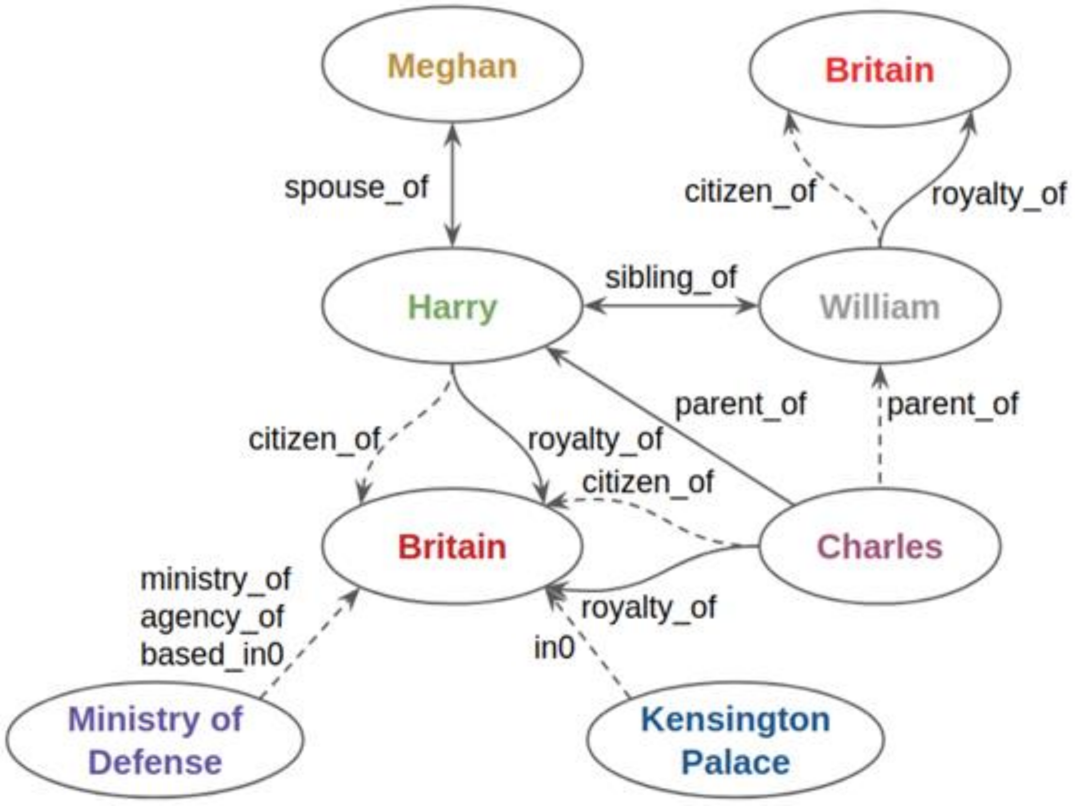
Reasoning Types	%	Examples
Pattern recognition	38.9	[1] *Me Musical Nephews* is a 1942 one-reel animated cartoon directed by Seymour Kneitel and animated by Tom Johnson and George Germanetti. [2] Jack Mercer and Jack Ward wrote the script. ... **Relation:** `publication_date` **Supporting Evidence:** 1
Logical reasoning	26.6	[1] "Nisei" is the ninth episode of the third season of the American science fiction television series The X-Files. ... [3] It was directed by David Nutter, and written by Chris Carter, Frank Spotnitz and Howard Gordon. ... [8] The show centers on FBI special agents *Fox Mulder* (David Duchovny) and Dana Scully (Gillian Anderson) who work on cases linked to the paranormal, called X-Files. ... **Relation:** `creator` **Supporting Evidence:** 1, 3, 8
Coreference reasoning	17.6	[1] *Dwight Tillery* is an American politician of the Democratic Party who is active in local politics of Cincinnati, Ohio. ... [3] He also holds a law degree from the University of Michigan Law School. [4] *Tillery* served as mayor of Cincinnati from 1991 to 1993. **Relation:** `educated_at` **Supporting Evidence:** 1, 3
Common-sense reasoning	16.6	[1] *William Busac* (1020-1076), son of William I, Count of Eu, and his wife Lesceline. ... [4] *William* appealed to King Henry I of France, who gave him in marriage Adelaide, the heiress of the county of Soissons. [5] Adelaide was daughter of Renaud I, Count of Soissons, and Grand Master of the Hotel de France. ... [7] *William* and Adelaide had four children: ... **Relation:** `spouse` **Supporting Evidence:** 4, 7

DocRED: A Large-Scale Document-Level Relation Extraction Dataset. Yao et al. ACL 2019.

Sentences from the Document Corpus

Some additional steps are required

- Entity clustering / canonicalization
- Entity resolution
- Entity linking
  - (adding new entities if necessary)
- Schema matching
- Relation linking



DWIE: An entity-centric dataset for multi-task document-level information extraction. Klim Zaporojets, Johannes Deleu, Chris Develder, Thomas Demeester. Information Processing and Management 2021.

Relation Extraction Benchmarks

Relation Extraction Benchmarks

Corpus Name	General	Specific	Multi-lingual	Relation	Train/Test	Leaderboard
NYT [125]	✔			24	5.6 k/5 k	✔ [3]
WebNLG [51]	✔			171	5019/703	✔
WikiReading* [64]	✔			884	14.85 M/3.73 M	✔
WIKI-TIME [197]	✔			57	97.6 k/40 k	
SciERC [97]		Scientific		7	2,136/551	✔
FOBIE [81]		Scientific		3	1,238/300	✔
DialogRE [211]	✔			37	6 k/1.9 k	✔
FewRel 2.0 [50]		Medical		100+25	56 k/14 k	✔ [4]
ChemProt [115]		Biochemical		14	19.5 k/16.9 k	✔
DDI [63]		Biochemical		5	25.3 k/5.7 k	✔
DocRED* [204]	✔			96	4 k/1 k	✔
CUAD [62]		Legal		25	10.48 k/2.62 k	✔
FinRED [138]		Finance		29	5,699/1,068	✔
SMiLER [135]	✔		✔	36	733 k/15 k	
mLAMA [80]	✔		✔	5	–	
ACE 2023 [41]	✔		✔	24	100 k/50 k	✔
ACE 2024 [41]	✔		✔	24	300 k/50 k	✔

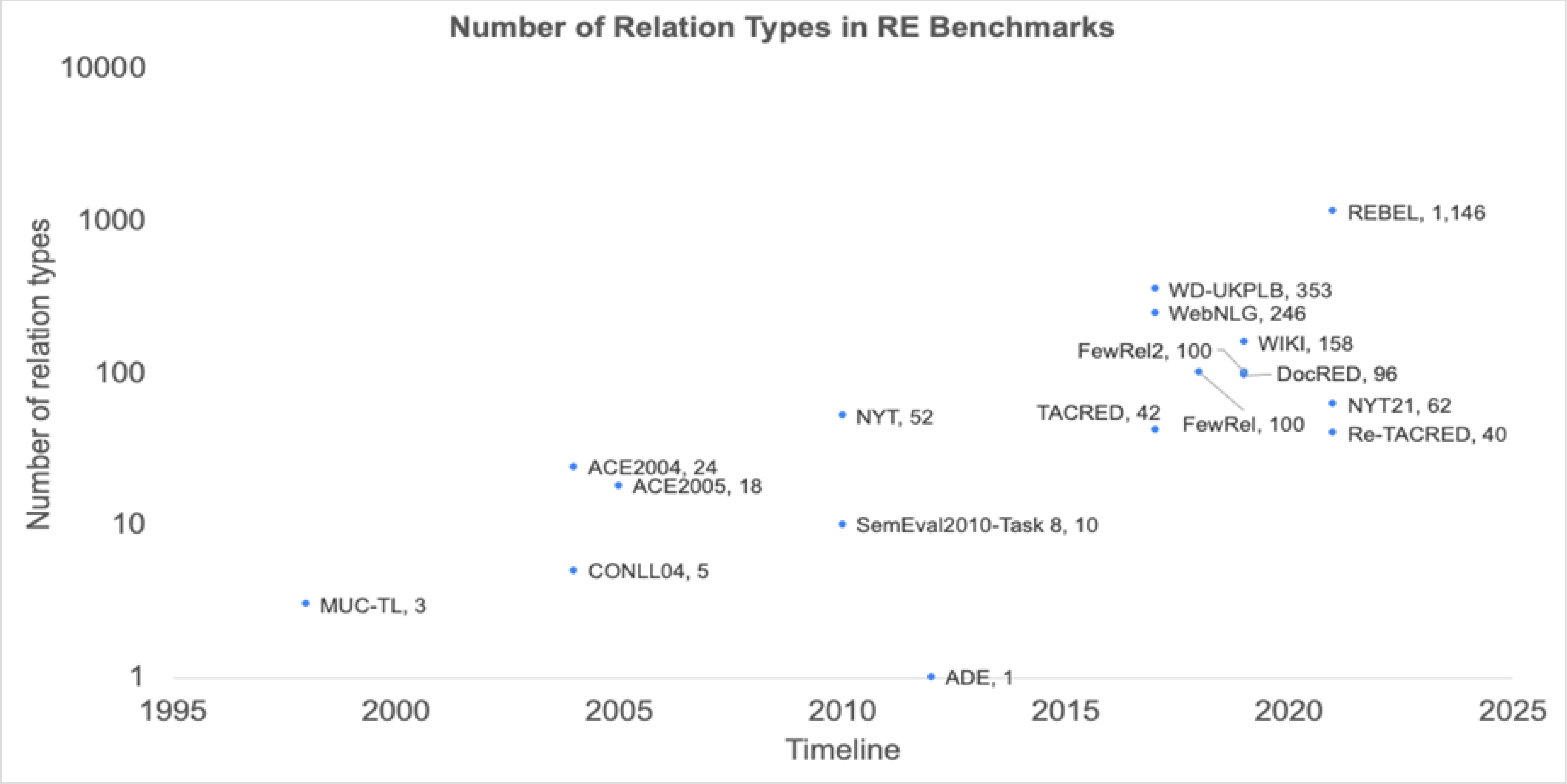
A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers. ACM Computing Surveys (2024).

KGC

Relation Extraction Benchmarks

	Benchmarks
1998	MUC-TR (1998)
2004	CoNLL (2004); ACE (2004)
2005	ACE (2005)
2007	SemEval-07 Task 4 (2007)
2010	NYT (2010); SemEval2010 Task 8 (2010)
2012	ADE (2012)
2013	Google RE (2013)
2017	WebNLG (2017); TACRED (2017); WD-UKPLAB (2017); WebNLG (2017)
2018	FewRel (2018); T-REx (2018)
2019	FewRel2 (2019); DocRED (2019); WIKI (2019)
2021	NYT21 (2021); Re-TACRED (2021); DWIE (2021); REBEL (2021); KELM (2021); TEKGEN (2021)
2022	KERED (2022)
2023	Text2KGBench (2023); SynthIE (2023); WdRef (2025)

Number of relation types in benchmarks



Automatic Content Extraction (ACE)

- Entity Detection and Tracking (EDT)
- Relation Detection and Characterization (RDC)

ACE 2003				ACE 2004		
Type	Subtype	Count		Type	Subtype	Count
AT	based-in	496		PHYS	LOCATED	745
	located	2879			NEAR	87
	residence	395			PART-WHOLE	384
NEAR	relative-location	288		PER-SOC	BUSINESS	179
PART	other	6			FAMILY	130
	part-of	1178			OTHER	56
	subsidiary	366		EMP-ORG	EMPLOY-EXEC	503
ROLE	affiliate-partner	219			EMPLOY-STAFF	554
	citizen-of	450			EMPLOY-undetermined	79
	client	159			MEMBER-OF-GROUP	192
	founder	37			SUBSIDIARY	209
	general-staff	1507			PARTNER	12
	management	1559			OTHER	82
	member	1404		ART	USER/OWNER	200
	other	174			INVENTOR/MANUFACTURER	9
	owner	274			OTHER	3
SOCIAL	associate	119		OTHER-AFF	ETHNIC	39
	grandparent	10			IDEOLOGY	49
	other-personal	108			OTHER	54
	other-professional	415		GPE-AFF	CITIZEN/RESIDENT	273
	other-relative	86			BASED-IN	216
	parent	149			OTHER	40
	sibling	23		DISC	DISC	279
	spouse	89				

Relation Extraction : A Survey. Pawar et al. 2017.

REBEL

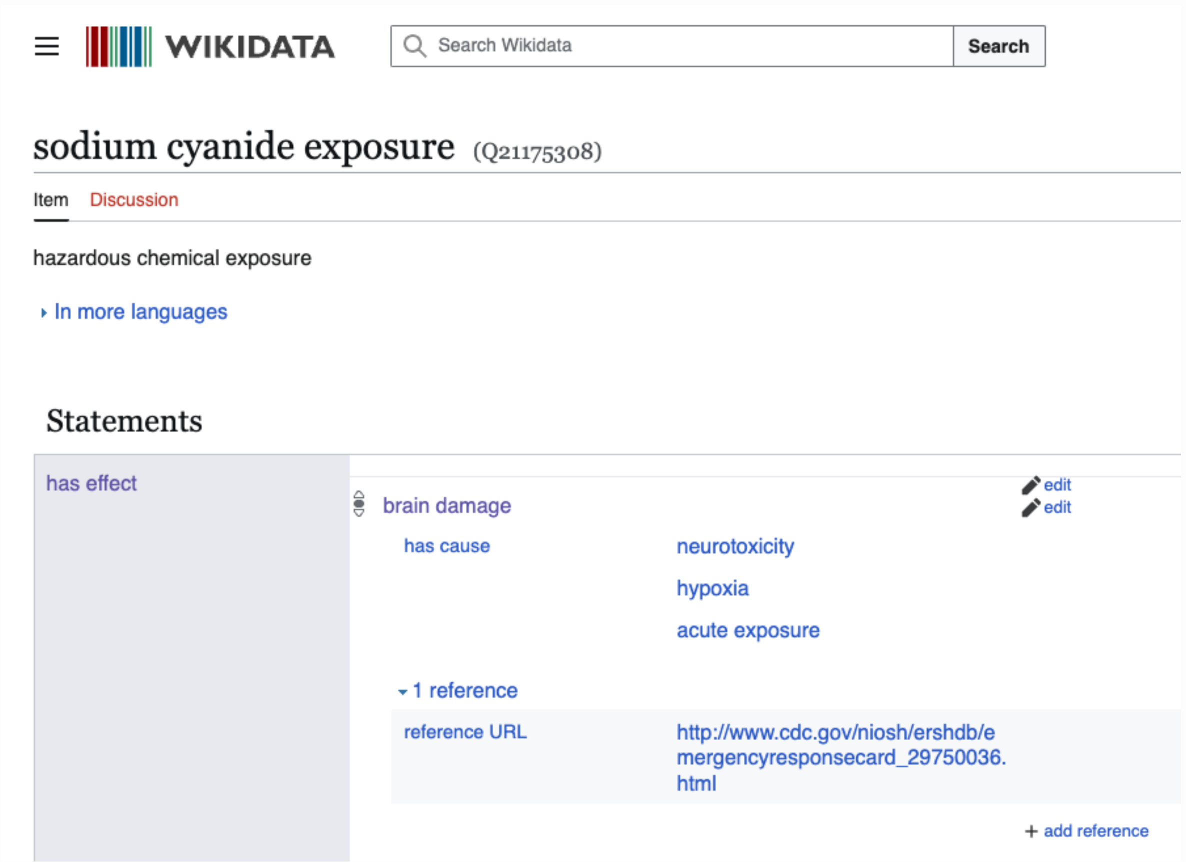
- Wikipedia abstracts as text
- Entity spans are extracted using hyperlinks
- All relations and values from Wikidata for those entities
- A Natural Language Inference (NLI) model used to check for entailment

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

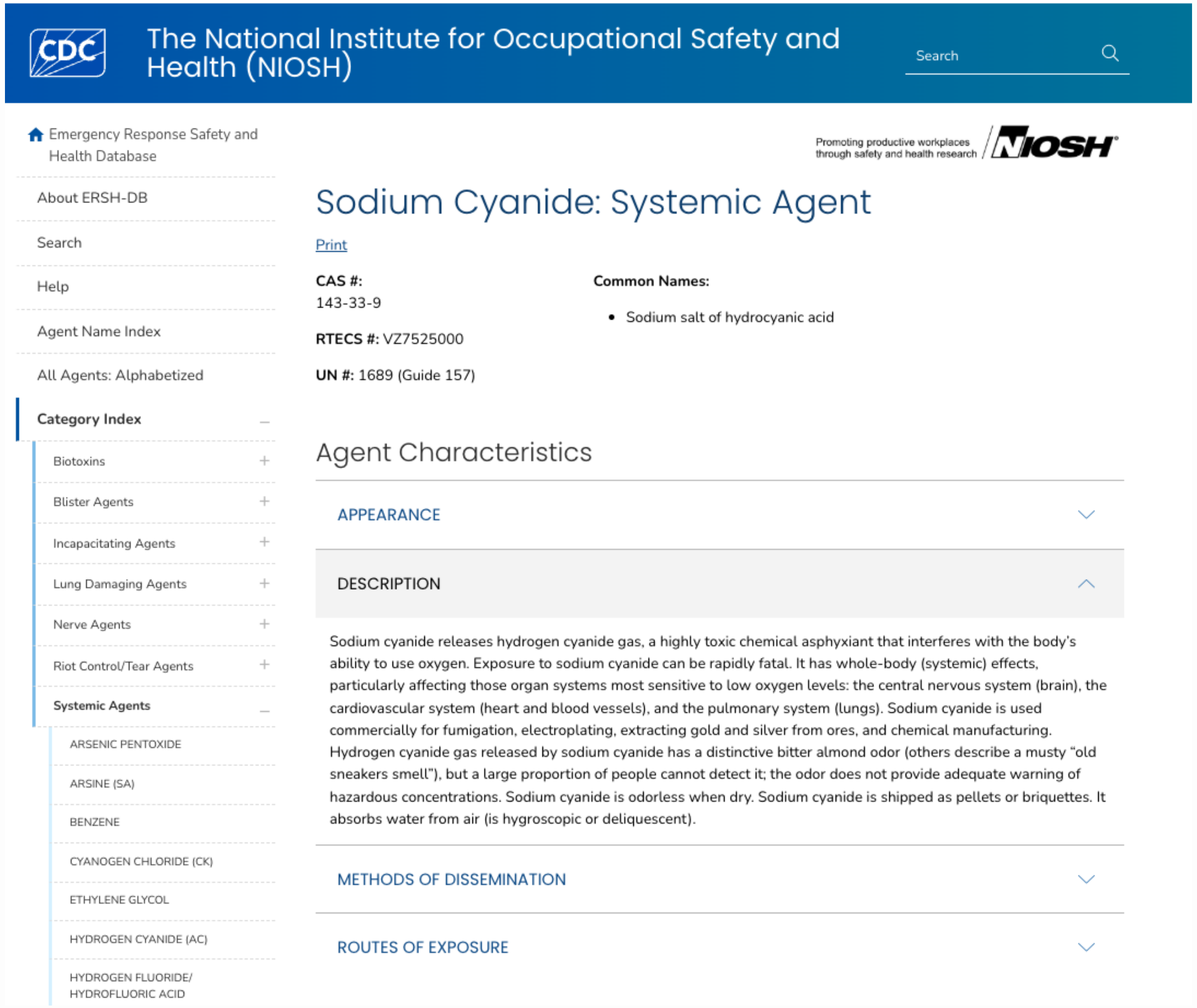
(This Must Be the Place, performer, Talking Heads)
(Talking Heads, genre, new wave)
(This Must Be the Place, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

REBEL: Relation extraction by end-to-end language generation. Cabot and Navigli. EMNLP 2021.

Wikidata Ref Bench



https://www.wikidata.org/wiki/Q21175308



https://www.cdc.gov/niosh/ershdb/emergencyresponsecard_29750036.html

Mihindukulasooriya and Hertling. Towards Triple-to-Text Alignments for Relation Extraction from Wikidata References using LLMs. ESWC 2025.

Wikidata Ref Bench

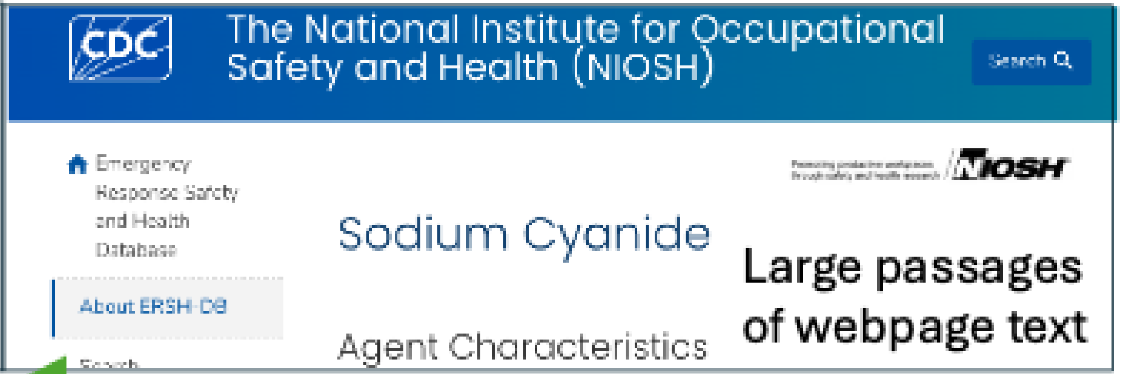
Triple – Reference Pair

Reference:
http://www.cdc.gov/niosh/ershdb/emergencyresponsecard_29750036.html

subject: sodium cyanide exposure (Q21175308)
relation: has effect (P1542)
object: brain damage (Q720026)

WIKIDATA

Web Page Content



Text entailment validation

Validation:
Fluent Sentences(s): TRUE
Subject mentioned in text: TRUE
Relation mentioned in text: TRUE
Object mentioned in text: TRUE
Fact entailed by text: TRUE

LLM Output

Aligned text span extraction

Selected Text: Survivors of severe sodium cyanide exposures may suffer brain damage due to a direct effect of the poison on nerve cells, or to a lack of oxygen, or possibly due to insufficient blood circulation.

LLM Output

Mihindukulasooriya and Hertling. Towards Triple-to-Text Alignments for Relation Extraction from Wikidata References using LLMs. ESWC 2025.

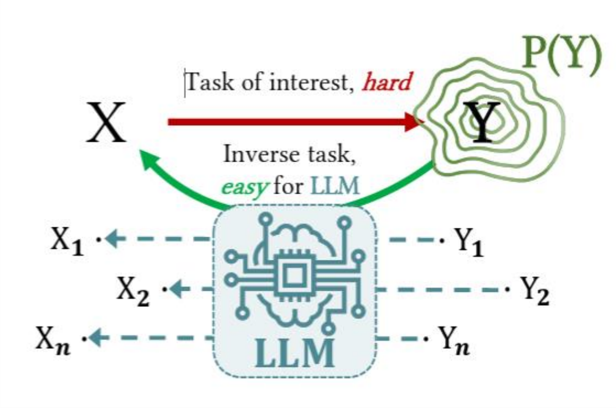
Wikidata Ref Bench – diversity of sources

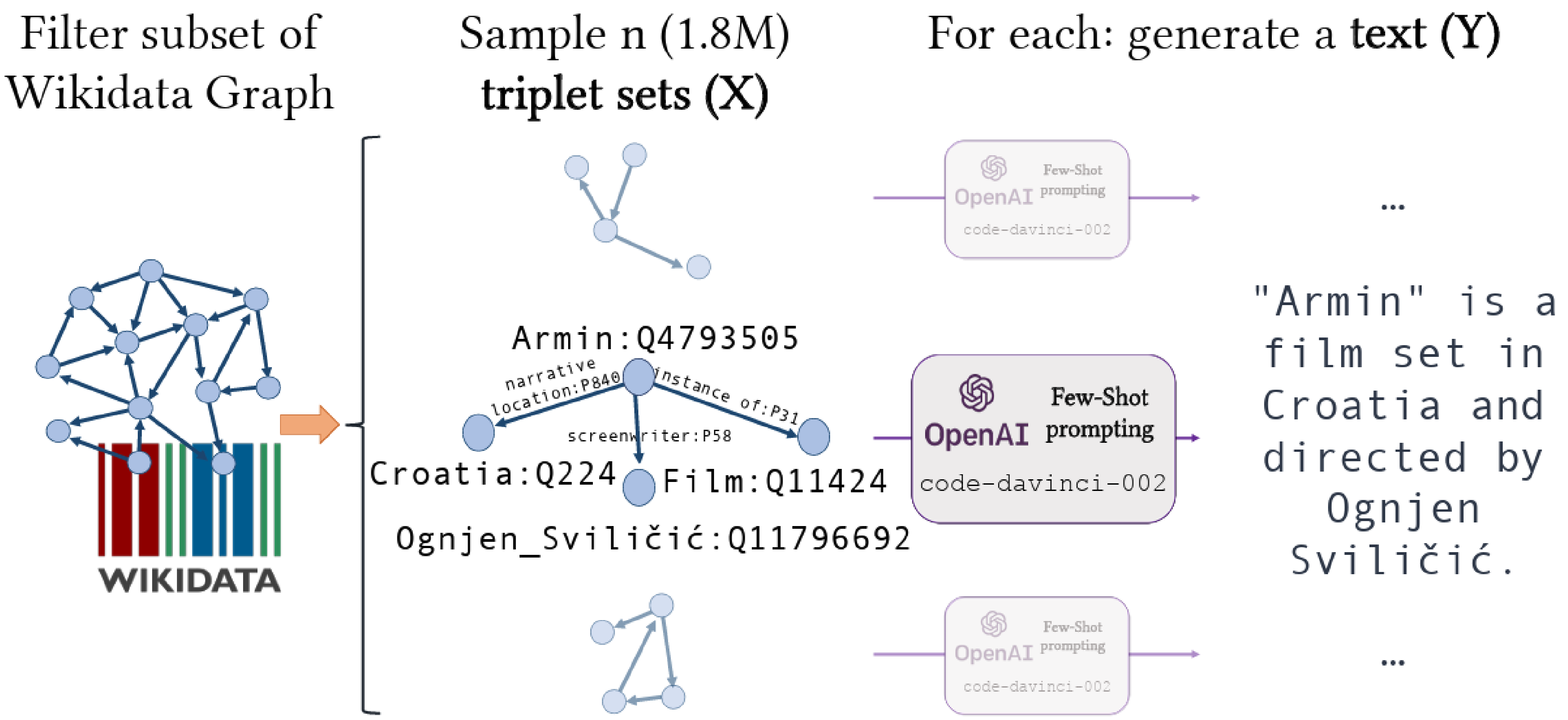
Subject	relation	Object	Source	Sentence
open-angle glaucoma	drug or therapy used for treatment	dipivefrin	bioportal.bioontology.org	Dipivefrin is used as initial therapy for the control of intraocular pressure in chronic open-angle glaucoma.
XPO, Inc.	chief executive officer	Mario Harik	www.cio.com	As XPO's first CIO, Mario Harik played a key role in making the logistics company an industry leader and innovation powerhouse. Harik, now having become CEO and a member of the board last August, leading the the company in its next phase.
MIMIC	funding scheme	AHRC Research Gran	gtr.ukri.org	"Musically Intelligent Machines Interacting Creatively" (MIMIC) is a three year AHRC-funded project, run by teams at Goldsmiths College, Durham University and the University of Sussex.
Erlend Berg	academic thesis	The financial lives of the poor	etheses.lse.ac.uk	Berg, Erlend (2009) The financial lives of the poor. PhD the-sis, London School of Economics and Political Science.

Mihindukulasooriya and Hertling. Towards Triple-to-Text Alignments for Relation Extraction from Wikidata References using LLMs. ESWC 2025.

KGC

SynthIE





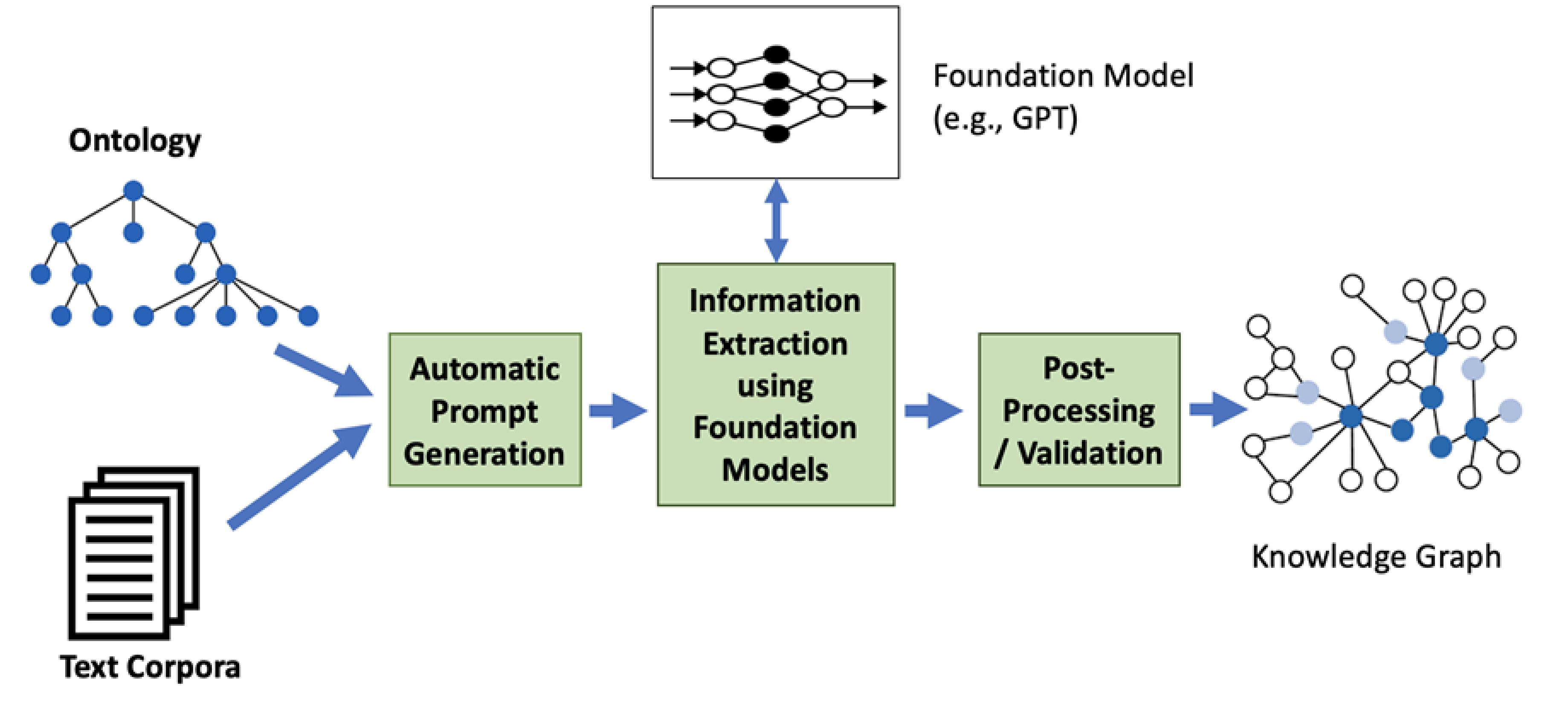
Exploiting Asymmetry for Synthetic Training Data Generation: SynthIE and the Case of Information Extraction.
Josifoski et al. EMNLP 2023

Text2KGBench

- Inputs
  - An ontology
  - A set of sentences
- Outputs
  - Triples aligned with each sentence(s) of text conforming to the ontology

Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text.
Mihindukulasooriya et al. ISWC 2023.

Text2KGBench

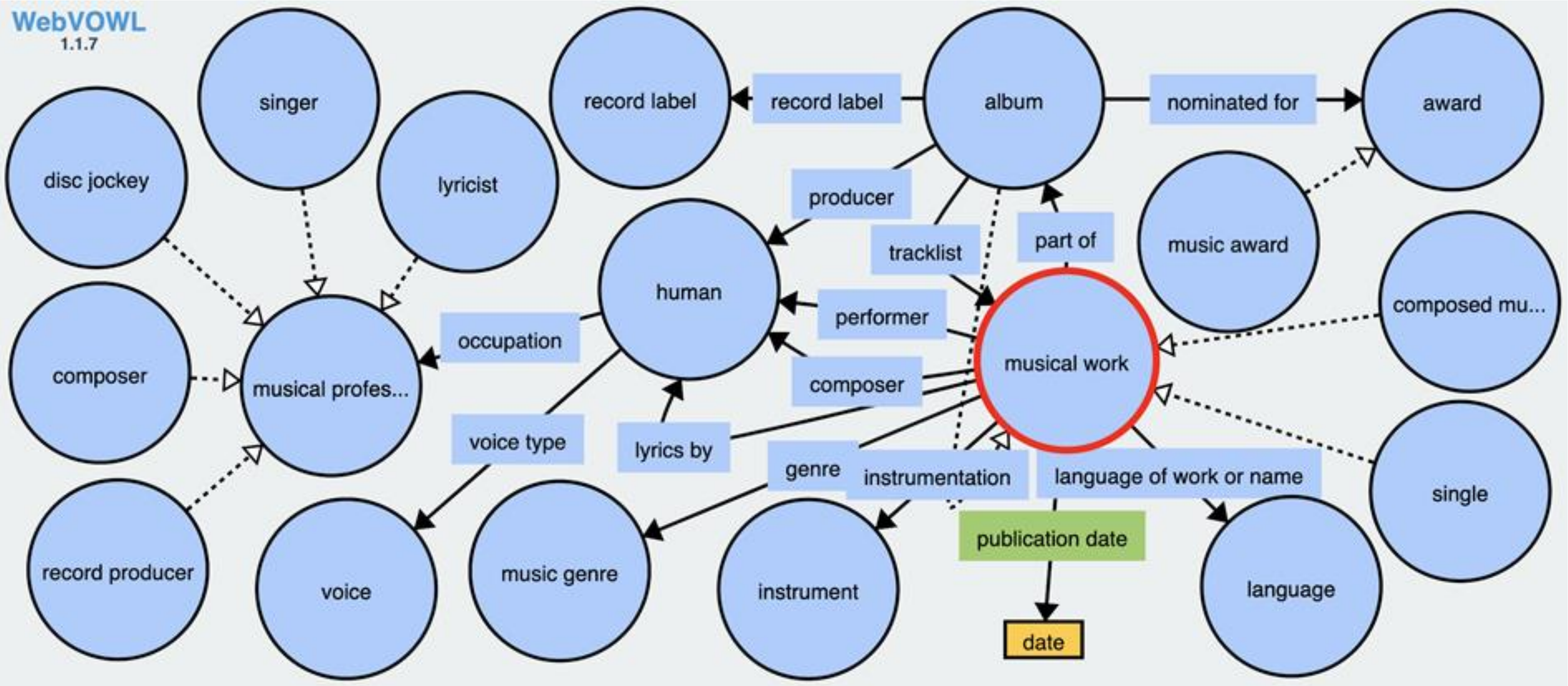


Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text.
Mihindukulasooriya et al. ISWC 2023.

Text2KGBench

- **A smaller domain ontologies from Wikidata**

- Music
- Movie
- Sport
- Book
- Military
- Computer
- Space
- Politics
- Nature
- Culture



Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text. Mihindukulasooriya et al. ISWC 2023.

Text2KGBench

- **Aligned sentence selection**
  - For each ontology, a subset of triple-sentence pairs chosen based on the relations / classes of the ontology
  - TekGen corpus provides Wikipedia sentences aligned to Wikidata triples (16M aligned triple-sentence pairs with 663 Wikidata relations)

The Lion King is an animated musical drama film directed by Roger Allers and Rob Minkoff, produced by Don Hahn.

director (“Lion King”, “Roger Allers”)
director (“Lion King”, “Rob Minkoff”)
producer (“Lion King”, “Don Hah”)

Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text. Mihindukulasooriya et al. ISWC 2023.

Text2KGBench

- **Unseen sentence generation**
  - Language models have seen these sentence during pre-training
  - There is a possibility of data leakage in training data
  - ***Will the results differ significantly on totally unseen sentences?***

John Doe starred in the film **The Fake Movie** that was released in **2025**.

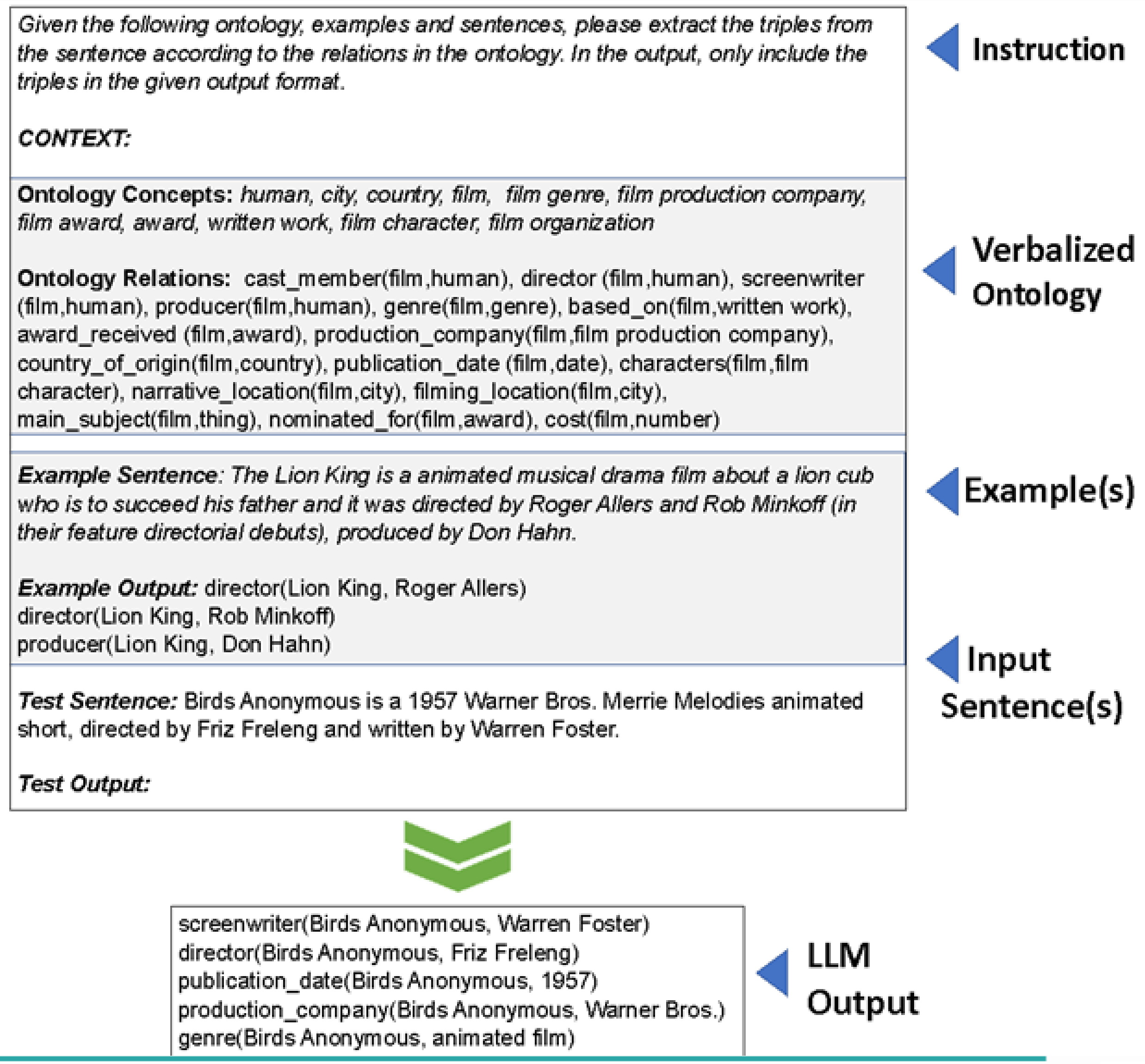
cast member (“The Fake Movie”, “John Doe”)
publication date (“The Fake Movie”, 2025)

Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text.
Mihindukulasooriya et al. ISWC 2023.

KGC

Text2KGBench

- The baseline uses in-context learning
- Best examples are selected using sentence similarity (SBERT) from training data
- Other approaches to use training data to improve
  - For fine-tuning/instruction-tuning the models
  - For prompt-tuning the models
- Other improvements
  - Ontology verbalization
  - Self-reflection / critique
  - Chain-of-thought
  - Constrained beam search



KGC

Text2KGBench – Evaluation Metrics

- Fact extraction accuracy – Precision (P), Recall (R), and F1
  - System generated triples compared to ground truth triples
  - Ground truth may not be exhaustive
  - to avoid false negative and out-of-context triples, a locally closed approach considering only relations in ground truth
- Ontology conformance
  - Does the LLM adhere to the ontology provided in the prompt
  - Currently only checks if the relations are canonical ones from the ontology
  - Can be further extended to ensure domain/range and other constraints
- Hallucinations
  - We expect the triples to be faithful to the input sentence provided
  - Subject entity / object should be present in the sentence
  - Stemming is used to address morphological variations (further work is needed to make this robust)

Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text.
Mihindukulasooriya et al. ISWC 2023.

KGC

Text2KGBench – Error Analysis

Sentence	Extracted Triple	Error / Related metric
Aparajito won 11 international **awards**, including the **Golden Lion and Critics Award** at the Venice Film Festival, be-coming the first ever film to win both.	**award_received** (Aparajito, Venice Film Festival)	An incorrect fact extracted. The model mistook the film festival for an award
The Gallopin Gaucho was a second attempt at success by co-directors **Walt Disney** and Ub Iwerks.	**directed** (The Gallopin Gaucho, Walt Disney)	Ontology conformance error. The canonical relation is the director.
American Born Chinese is a graphic novel by Gene Luen Yang.	**narrative_location** (American Born Chinese, San Francisco)	Object hallucination. Neither the object nor the relation is mentioned in the text.
Schreck was a founding member of the **Sturmabteilung**.	**member_of_political_party** (Hermann Goring, Sturmabteilung)	Subject hallucination. Hermann Goring is not mentioned in the text.

Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text.
Mihindukulasooriya et al. ISWC 2023.

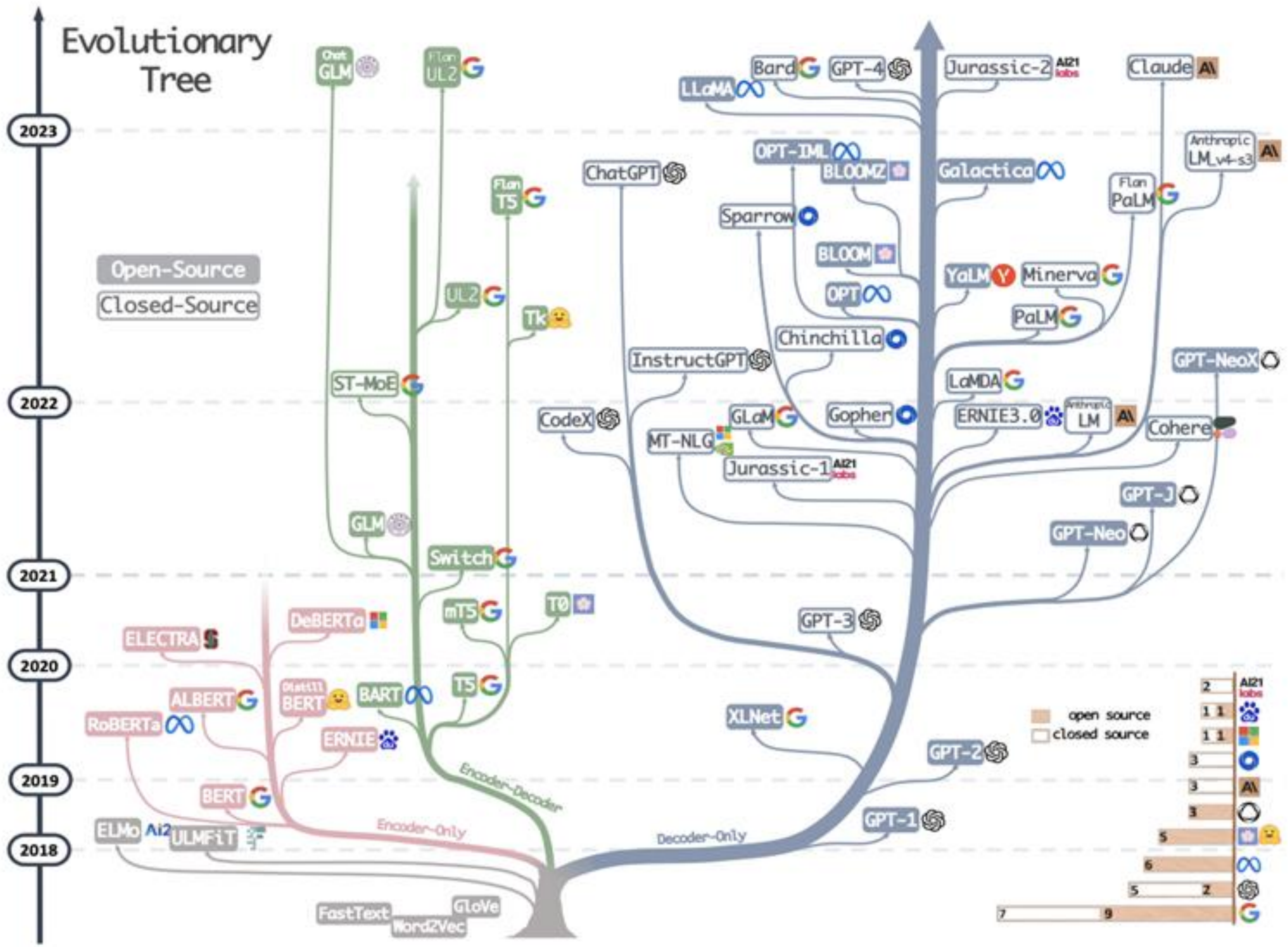
Relation Extraction Approaches

Traditional relation extraction approaches

- Rule-based approaches
  - Lexical analysis and phrase patterns
  - Syntax analysis, dependency trees, and semantic parsing
- Weak supervision / distant supervision approaches
  - Noise reduction approaches
  - Embedding based approaches
  - Auxiliary information
- Supervised learning approaches
  - Feature-based methods
  - Kernel-based methods
  - LSTMs, CNNs, GNNs

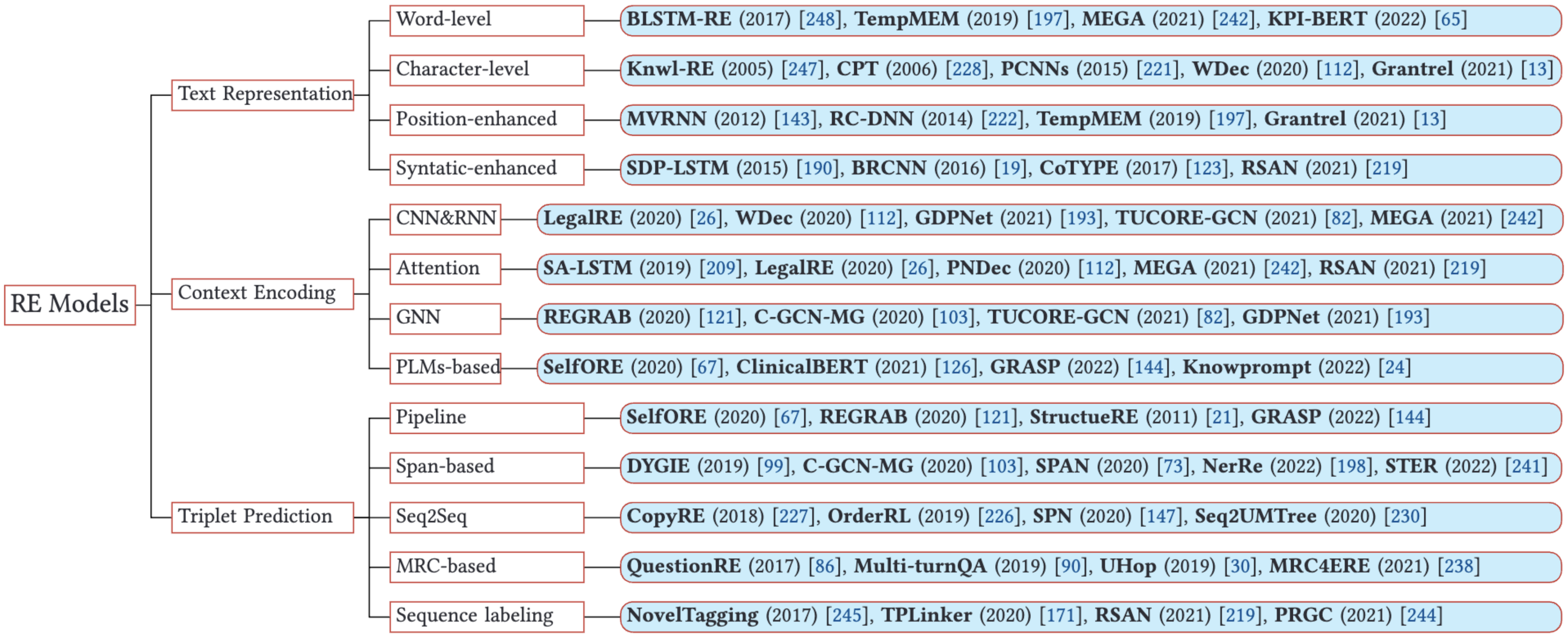
A Review of Relation Extraction. Bach, Nguyen, and Sameer Badaskar. Literature Review for Language and Statistics (2007)
Relation Extraction using Distant Supervision: A Survey. Smirnova, A. and Cudré-Mauroux. ACM Computing Surveys (2018).

Large Language Models-based Approaches



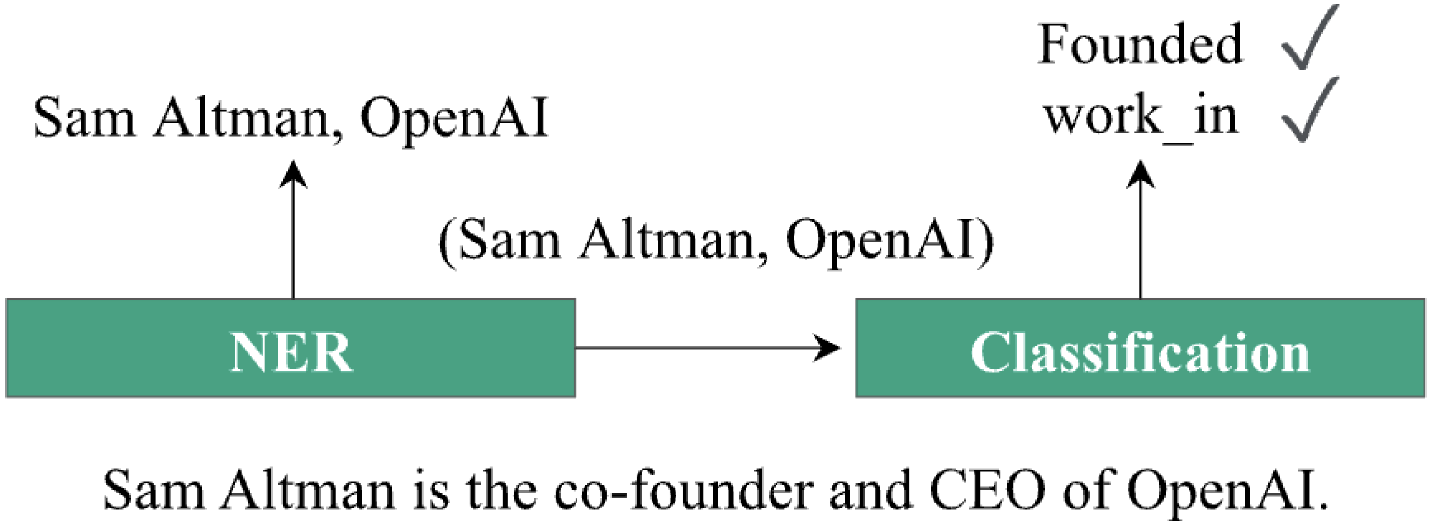
Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. Yang et al.
ACM Transactions on Knowledge Discovery from Data (2024)

Relation Extraction Models



A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers. ACM Computing Surveys (2024).

Pipeline based RE



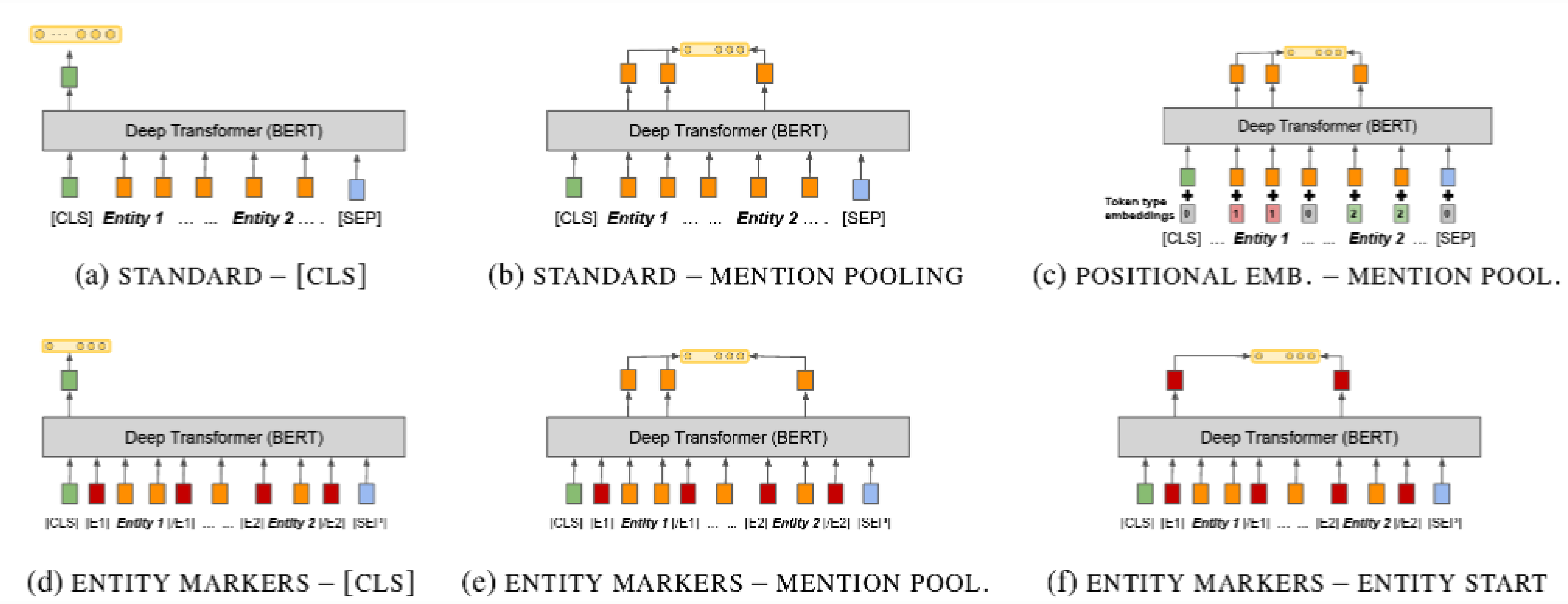
(a) The pipeline approach

Entity Overlaps

NEO: No Entity Overlap
SEO: Single Entity Overlap
EPO: Entity Pair Overlap

	Text	Triplets
NEO	The [United States] president [Donald Trump] will visit [Beijing], [China].	(Donald Trump, President_of, United States) (China, Contains, Beijing)
SEO	The [United States] president [Donald Trump] was born in [New York City].	(Donald Trump, President_of, United States) (Donald Trump, Born_in, New York City)
EPO	Martin went to [Tokyo] last week, which is the capital of [Japan].	(Japan, Contains, Tokyo) (Japan, Capital, Tokyo)

Text representation



(a) STANDARD – [CLS]

(b) STANDARD – MENTION POOLING

(c) POSITIONAL EMB. – MENTION POOL.

(d) ENTITY MARKERS – [CLS]

(e) ENTITY MARKERS – MENTION POOL.

(f) ENTITY MARKERS – ENTITY START

Matching the Blanks: Distributional Similarity for Relation Learning. Baldini Soares et al. ACL 2019.

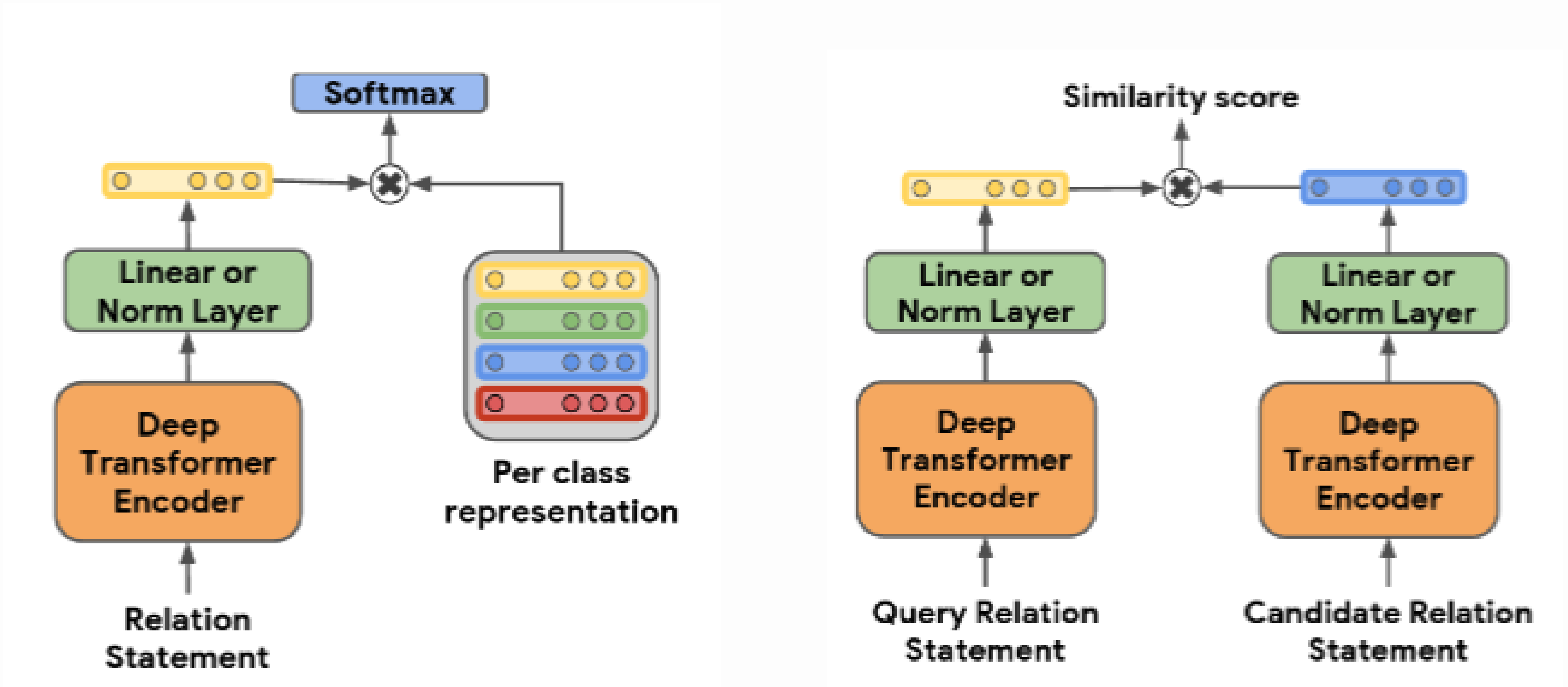
KGC

Text representation

		SemEval 2010 Task 8		KBP37		TACRED		FewRel 5-way-1-shot
# training annotated examples		8,000 (6,500 for dev)		15,916		68,120		44,800
# relation types		19		37		42		100
		Dev F1	Test F1	Dev F1	Test F1	Dev F1	Test F1	Dev Acc.
Wang et al. (2016)*		–	88.0	–	–	–	–	–
Zhang and Wang (2015)*		–	79.6	–	58.8	–	–	–
Bilan and Roth (2018)*		–	84.8	–	–	–	68.2	–
Han et al. (2018)		–	–	–	–	–	–	71.6
Input type	**Output type**							
STANDARD	[CLS]	71.6	–	41.3	–	23.4	–	85.2
STANDARD	MENTION POOL.	78.8	–	48.3	–	66.7	–	87.5
POSITIONAL EMB.	MENTION POOL.	79.1	–	32.5	–	63.9	–	87.5
ENTITY MARKERS	[CLS]	81.2	–	68.7	–	65.7	–	85.2
ENTITY MARKERS	MENTION POOL.	80.4	–	68.2	–	69.5	–	87.6
ENTITY MARKERS	ENTITY START	**82.1**	**89.2**	**70**	**68.3**	**70.1**	**70.1**	**88.9**

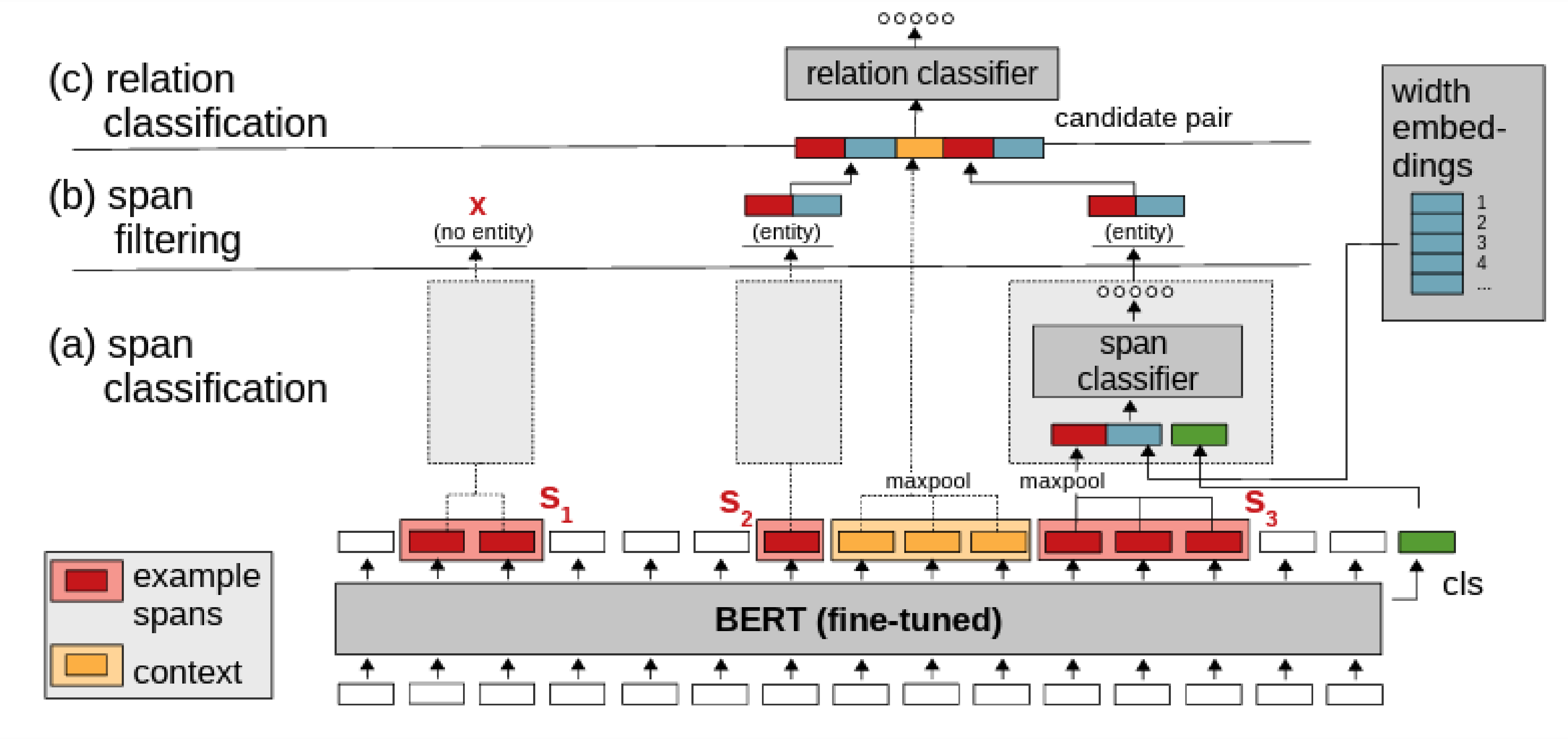
Matching the Blanks: Distributional Similarity for Relation Learning. Baldini Soares et al. ACL 2019.

Relation classification / ranking



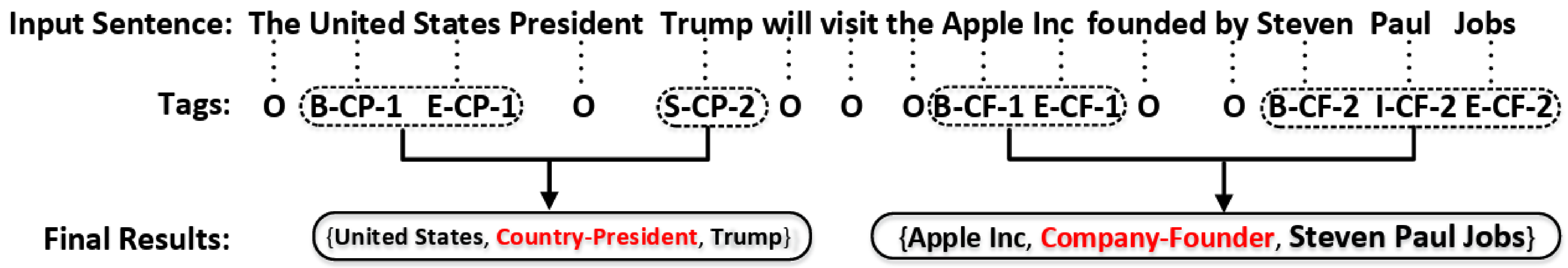
Matching the Blanks: Distributional Similarity for Relation Learning. Baldini Soares et al. ACL 2019.

Span-based approaches



Span-based joint entity and relation extraction with transformer pre-training.
Markus Eberts and Adrian Ulges. 2019. ECAI'20.

Sequence labeling approaches

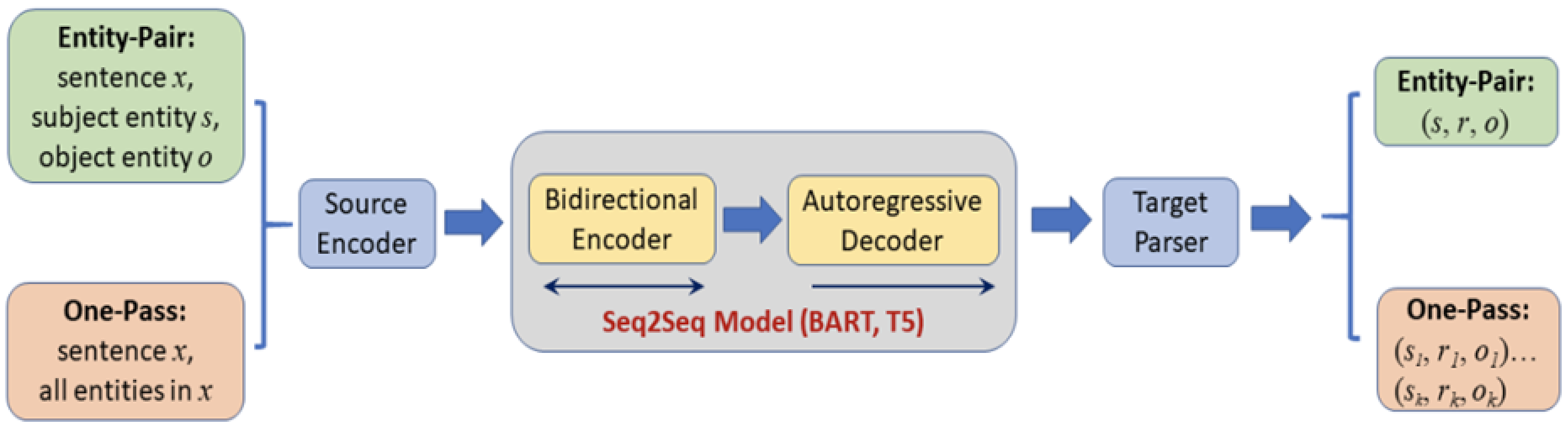


Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme.
Suncong Zheng et al. ACL 2017.

Relation extraction with Seq2Seq models (Encoder-Decoder)

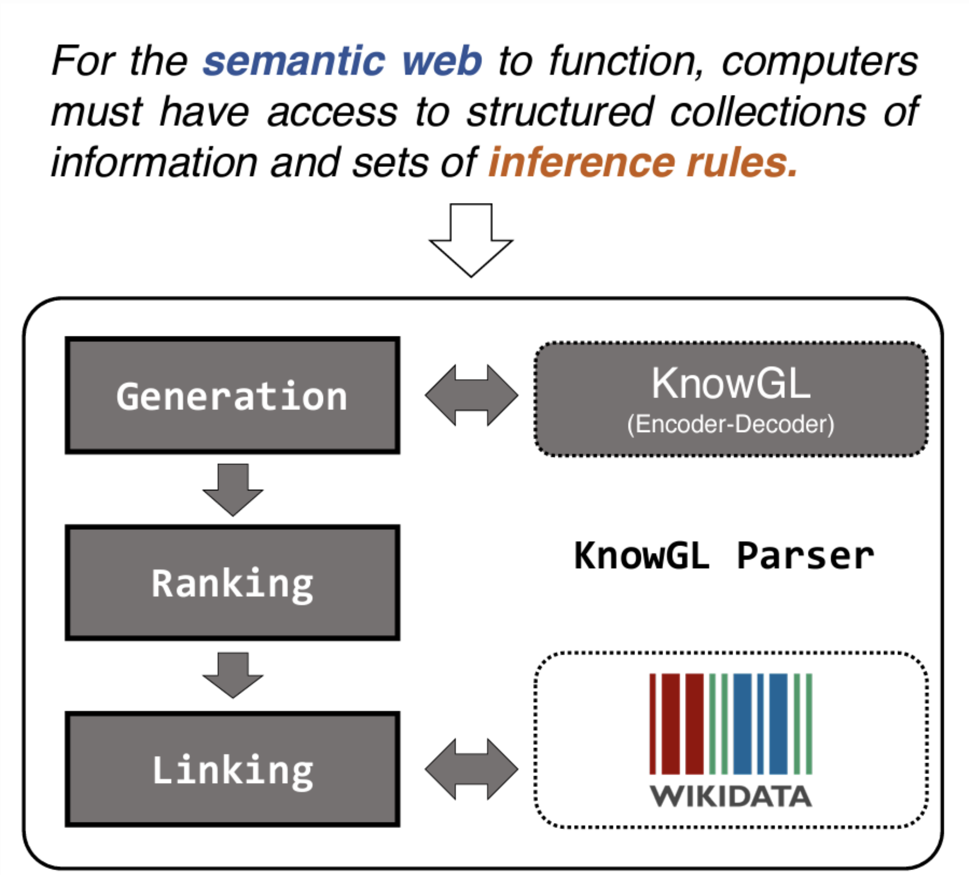
Irene Morgan, who was born and raised in Baltimore, lived on Long Island.

(Irene Morgan, birthPlace, Baltimore)
(Irene Morgan, residence, Long Island)



REBEL: Relation Extraction By End-to-end Language generation. Pere-Lluís Huguet Cabot, Roberto Navigli. EMNLP 2021.
A Generative Model for Relation Extraction and Classification. Jian Ni, Gaetano Rossiello, Alfio Gliozzo, Radu Florian. Arxiv 2022.
Generative Relation Linking for Question Answering over Knowledge Bases. Gaetano Rossiello, Nandana Mihindukulasooriya, Ibrahim Abdelaziz, Mihaela Bornea, Alfio Gliozzo, Tahira Naseem, Pavan Kapanipathi. ISWC 2021.

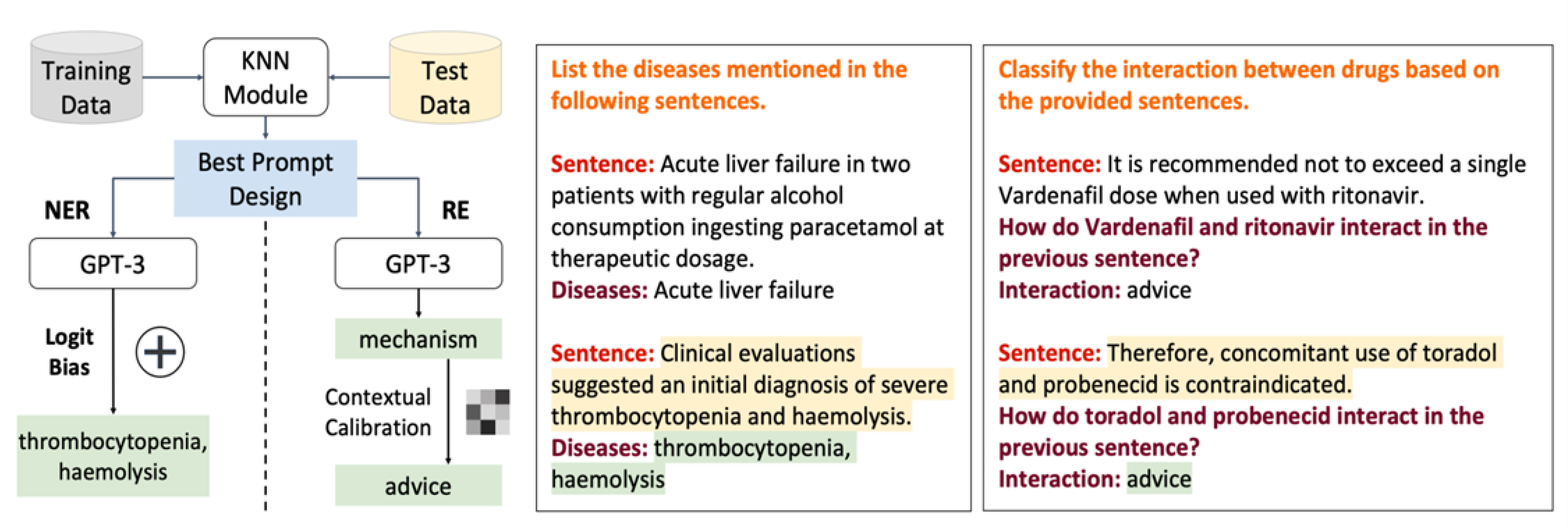
Relation extraction with Seq2Seq models (Encoder-Decoder)




```
[{
  "subject": {
    "mention": "semantic web",
    "entity_label": "Semantic Web",
    "type_label": "academic discipline",
    "entity_link": "Q54837",
    "type_link": "Q11862829"
  },
  "relation": {
    "label": "uses",
    "link": "Property:P2283"
  },
  "object": {
    "mention": "inference rules",
    "entity_label": "inference",
    "type_label": "process",
    "entity_link": "Q408386",
    "type_link": "Q619671"
  },
  "score": -0.98
}]
```

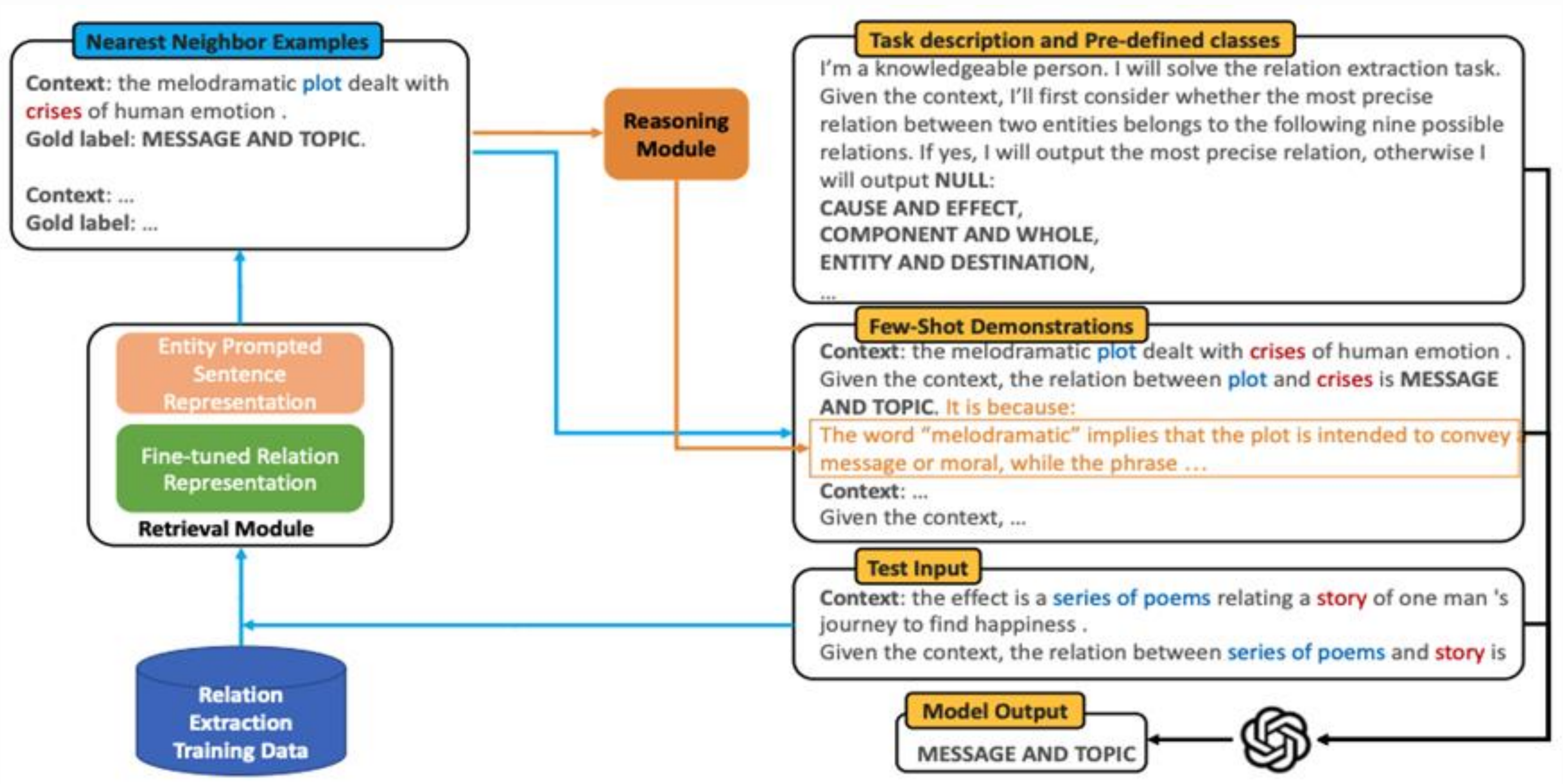

KnowGL: Knowledge Generation and Linking from Text. Gaetano Rossiello, Faisal Chowdhury, Nandana Mihindukulasooriya, Owen Cornec and Alfio Gliozzo. AAAI 2023.

Relation Extraction with In-Context Learning (Decoder-Only)



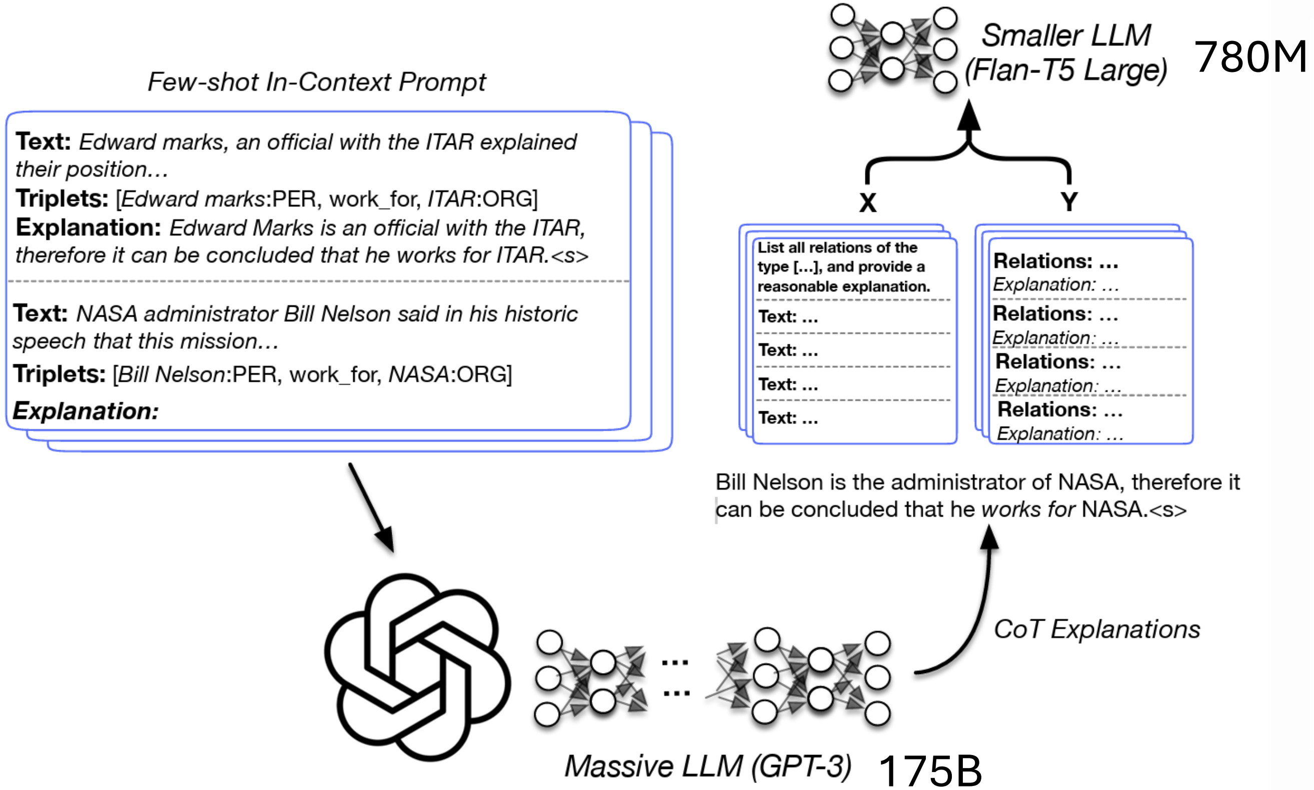
Thinking about GPT-3 In-Context Learning for Biomedical IE? Think Again. Jimenez Gutierrez et al. EMNLP Findings 2022.

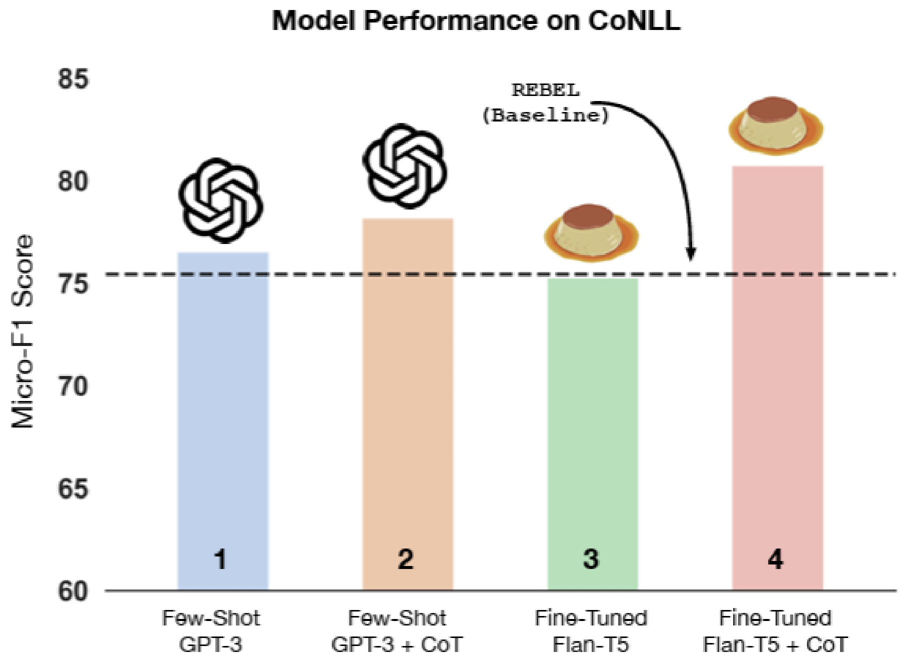
Relation Extraction with In-Context Learning (Decoder-Only)



GPT-RE: In-context Learning for Relation Extraction using Large Language Models. Zhen Wan, Fei Cheng, Zhuoyuan Mao, Qianying Liu, Haiyue Song, Jiwei Li, Sadao Kurohashi. EMNLP 2023.

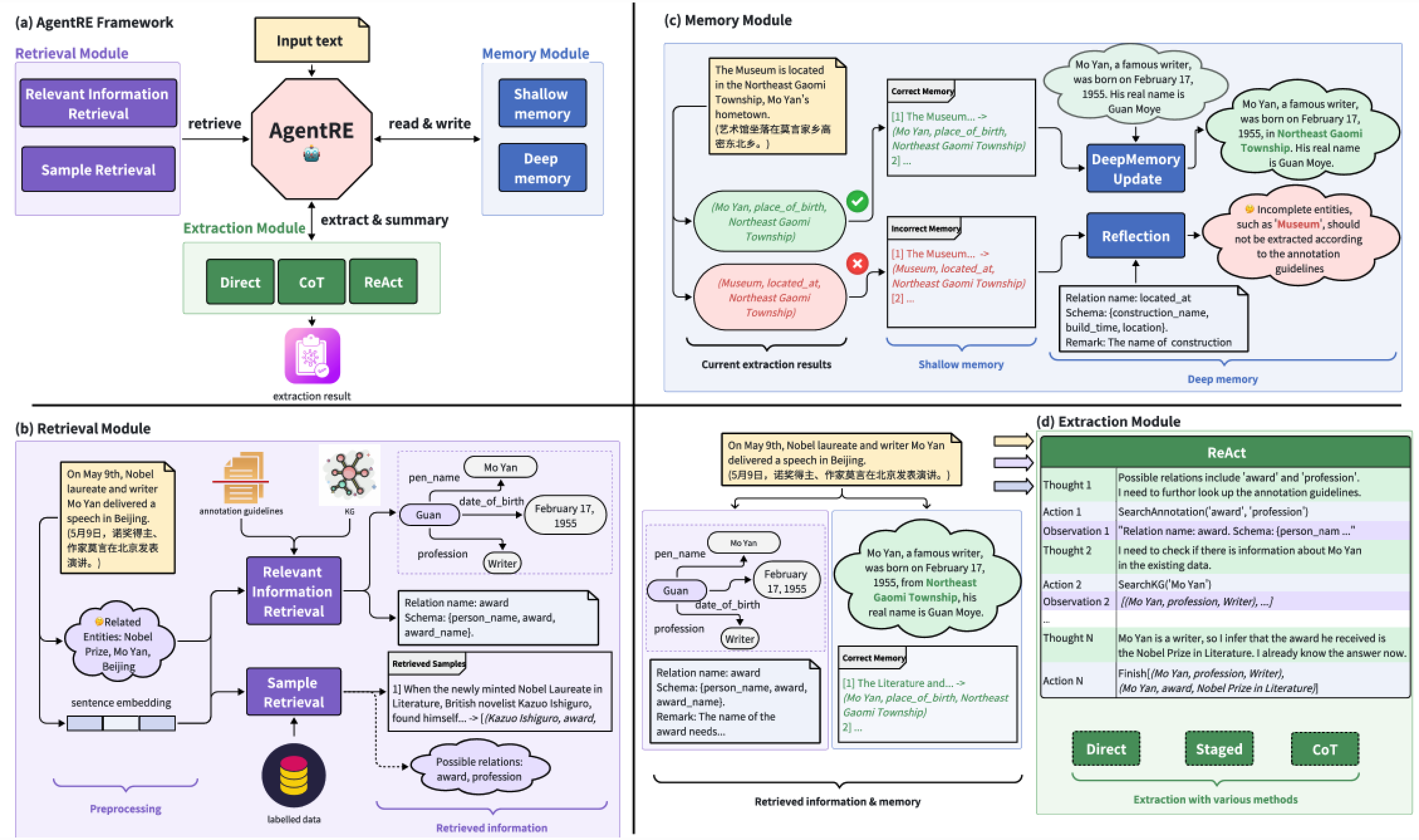
Relation Extraction with smaller models





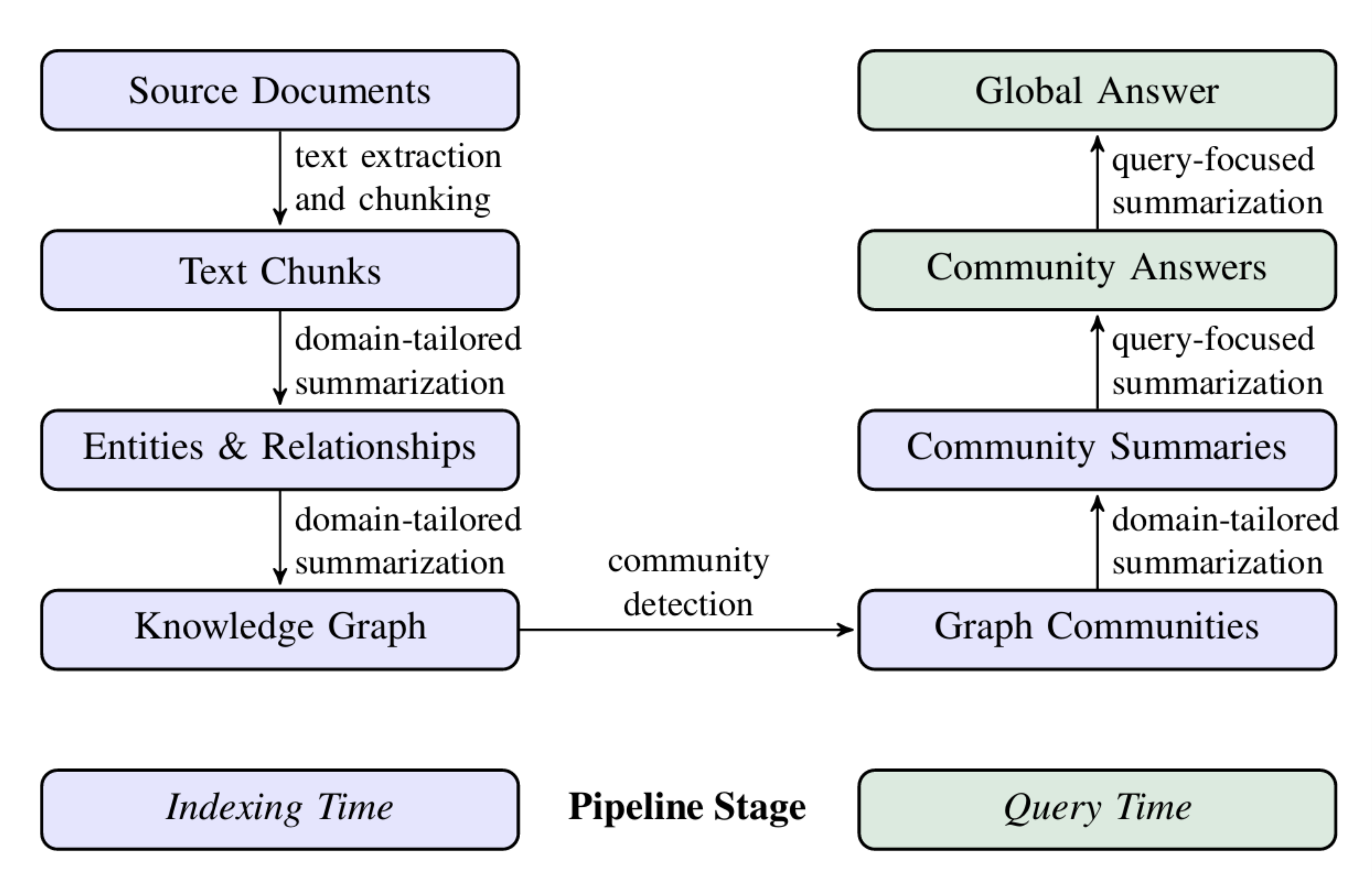
Revisiting Relation Extraction in the era of Large Language Models. Somin Wadhwa et al. ACL 2023.

Agent RE



AgentRE: An Agent-Based Framework for Navigating Complex Information Landscapes in Relation Extraction.
Yuchen Shi et al. CIKM 2024.

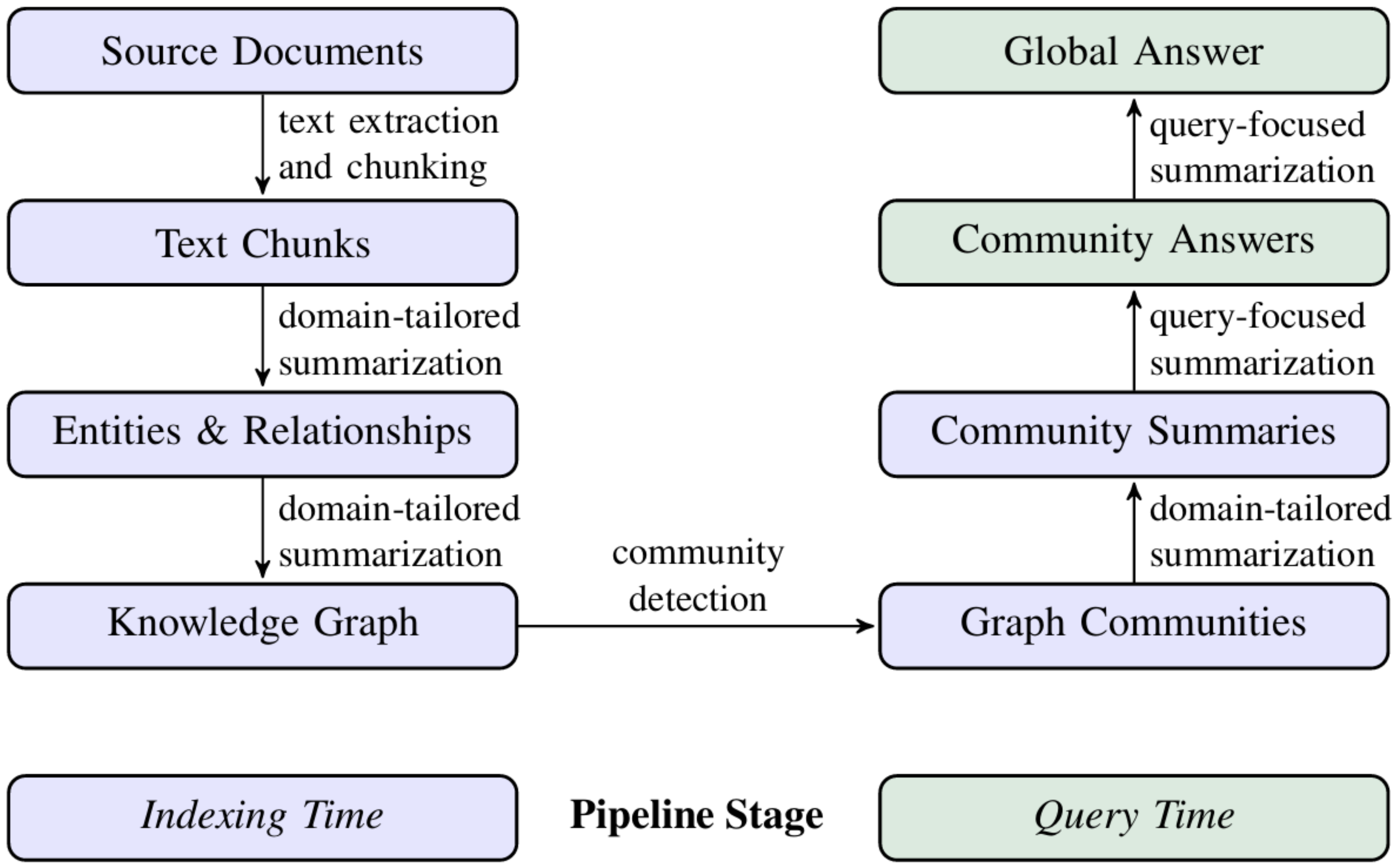
GrpahRAG



From Local to Global: A GraphRAG Approach to Query-Focused Summarization. Darren Edge et al. 2025.

KGC

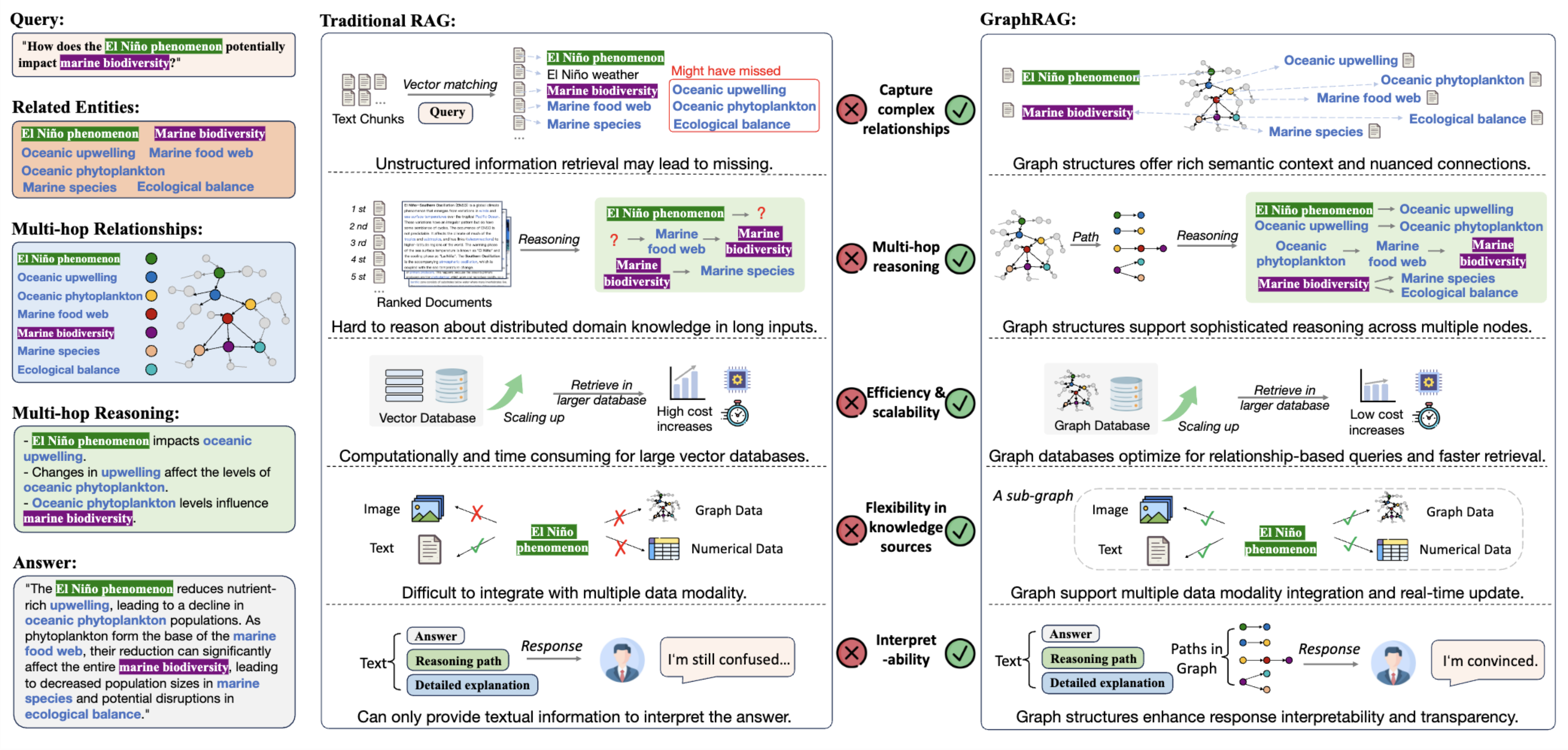
GrpahRAG



From Local to Global: A GraphRAG Approach to Query-Focused Summarization. Darren Edge et al. 2025.

KGC

RAG Vs GraphRAG



A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models.
Qinggang Zhang et al. 2025.

KGC

GraphRAG

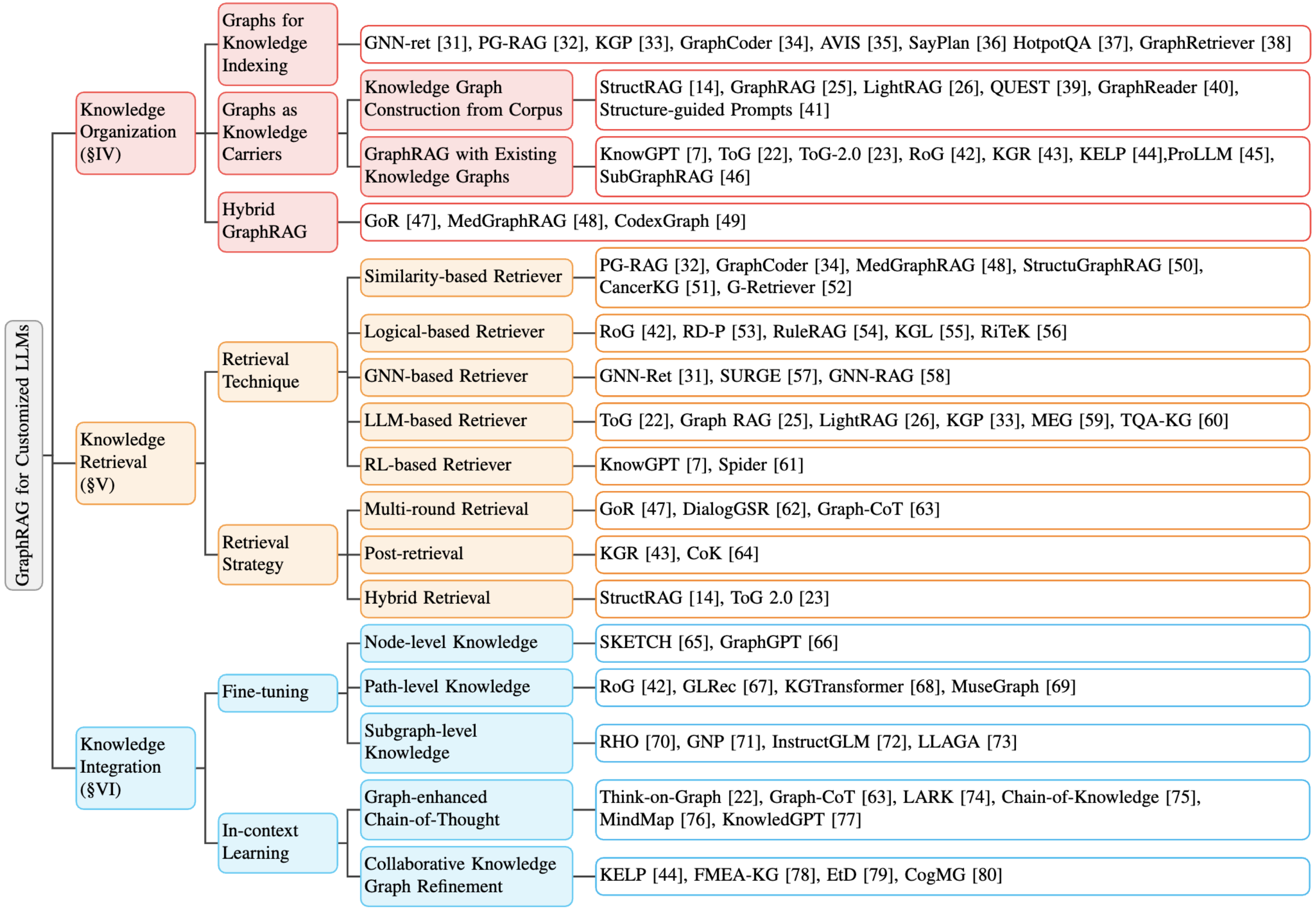
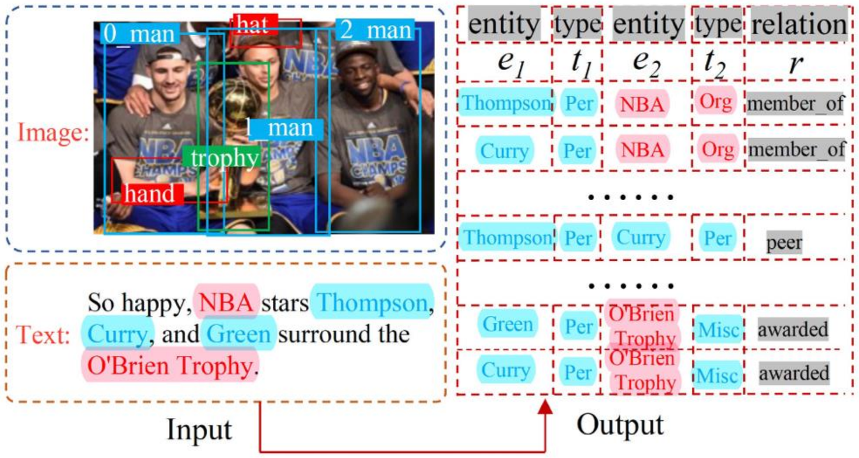


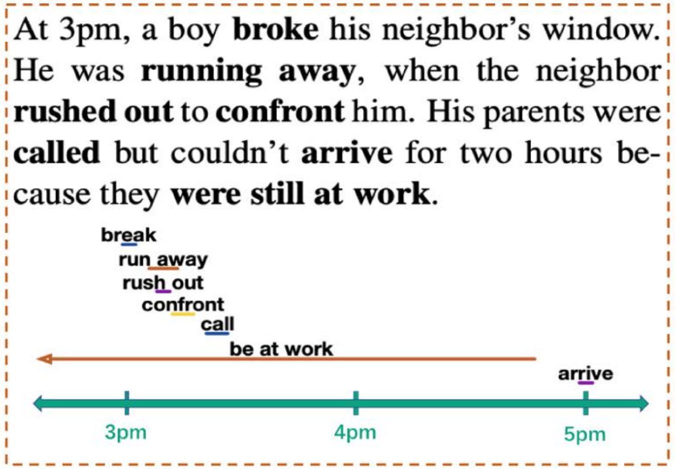
Fig. 3: The taxonomy for existing GraphRAG methods in the survey.

A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models.
Qinggang Zhang et al. 2025.

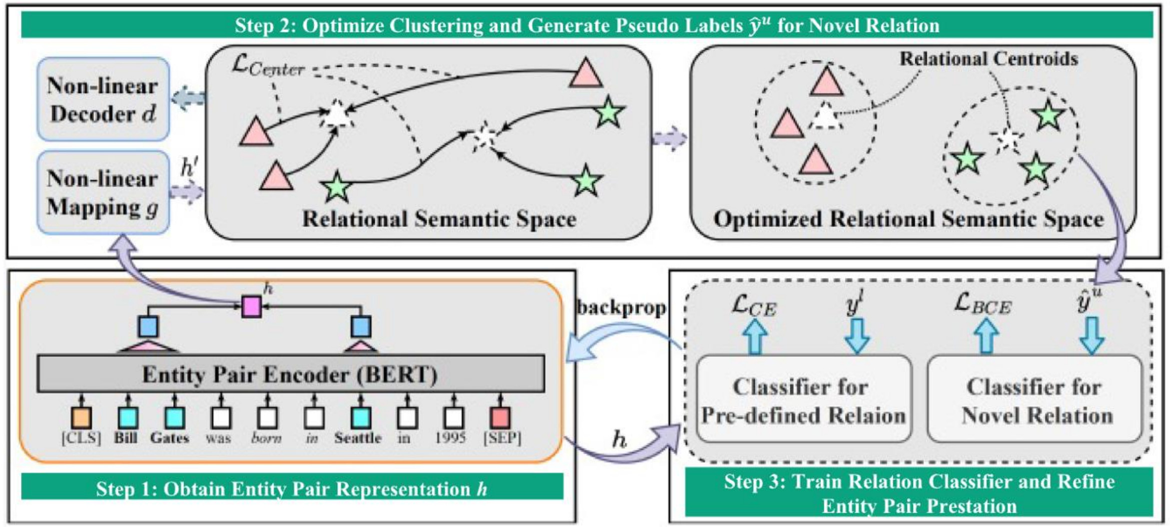
RE Challenges



(a) Multimodal relation extraction. It requires extracting different characteristic information from vision.



(b) Temporal relation extraction. The narrative and corresponding timeline are shown.



(c) Open relation extraction. There are three steps iteratively performed: (1) encode both the labeled and unlabeled instances into entity pair representations. (2) transfer the pair of entity representations into the relation-oriented representations by gathering towards their relational centroids. (3) optimize the entity pair representations and classifier by minimizing a joint objective function to reduce the clustering bias on predefined classes.

A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers.
Zhao et al. ACM Computing Surveys. 2024.

The State of the Art Large Language Models for Knowledge Graph Construction from Text: Techniques, Tools, and Challenges

Nandana Mihindukulasooriya
Senior Research Scientist | IBM Research

KGC | The Knowledge Graph Conference

#KGC2025

Join the Conversation

@TheKnowledgeGraphConference

Insert Handle Here

@TheKnowledgeGraphConference

www.knowledgegraph.tech