Trends in Joint Information Extraction

Hye-Yoon Baek (백혜윤)

Graph & Language Intelligence Laboratory
Department of Computer Science and Engineering
Konkuk University

2025.07.10

CONTENTS

Background

What is Information Extraction?

- Information Extraction
  - ✓ to extract structured information from unstructured natural language text
  - ✓ can extract various types of structures
    - Such as entities, relations, events , …

Background

Information Extraction

- Named Entity Recognition (NER)
  - ✓ 사람, 조직, 장소 등 미리 정의된 엔티티 범주에 해당하는 단어를 문서에서 인식하여 추출 및 분류

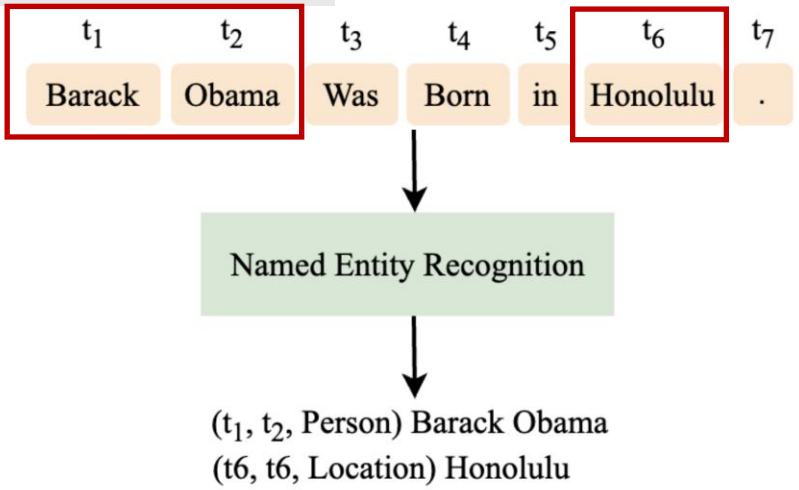


Fig. 1

- Relation Extraction (RE)
  - ✓ 두 엔티티 사이의 의미적 관계를 포착
  - ✓ 비정형 텍스트로부터 (subject, relation, object) 형태의 구조화된 트리플을 추출

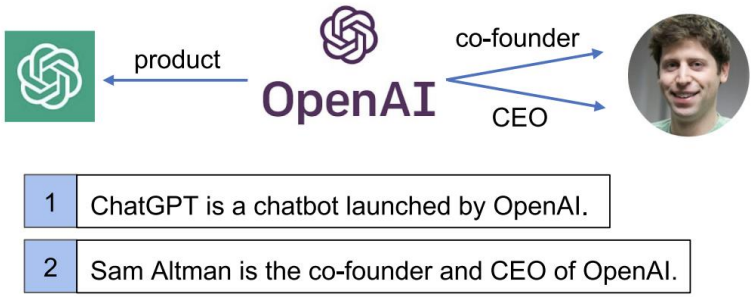


Fig. 2

(OpenAI, product, ChatGPT)
(OpenAI, co-founder, Sam Altman)
(OpenAI, CEO, Sam Altman)

Fig. 1. [2024][arXiv][-] Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study
Fig. 2. [2024][ACS][-] A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers

Background

OpenIE vs. ClosedIE

- Open Information Extraction

  : 사전에 정의된 스키마나 온톨로지 없이, 텍스트에서 직접 정보를 추출

  ✓ 문장에서 subject, object 역할의 argument 후보를 탐색

  ✓ 그 사이에 등장하는 동사, 전치사구 등을 predicate(relation)으로 설정

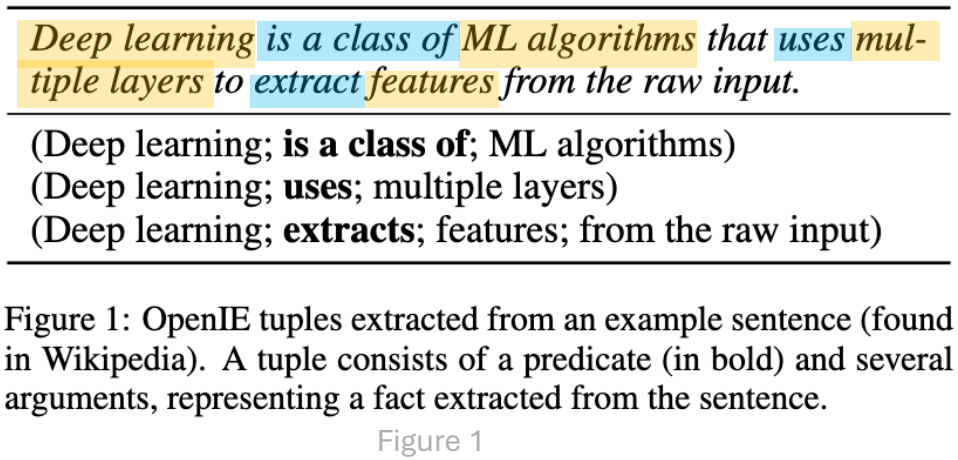


Figure 1: OpenIE tuples extracted from an example sentence (found in Wikipedia). A tuple consists of a predicate (in bold) and several arguments, representing a fact extracted from the sentence.

Figure 1

- Closed Information Extraction

  : 사전에 정의된 relation types를 기준으로 정보를 추출

  Ex) located_in, used_for …

Figure1. [2022][IJCAI][-] A Survey on Neural Open Information Extraction: Current Status and Future Directions

Background

OpenIE vs. ClosedIE

- Open Information Extraction

  : 사전에 정의된 스키마나 온톨로지 없이, 텍스트에서 직접 정보를 추출

  ✓ 문장에서 subject, object 역할의 argument 후보를 탐색

  ✓ 그 사이에 등장하는 동사, 전치사구 등을 predicate(relation)으로 설정

Deep learning is a class of ML algorithms that uses multiple layers to extract features from the raw input.

(Deep learning; **is a class of**; ML algorithms)
(Deep learning; **uses**; multiple layers)
(Deep learning; **extracts**; features; from the raw input)

Figure 1: OpenIE tuples extracted from an example sentence (found in Wikipedia). A tuple consists of a predicate (in bold) and several arguments, representing a fact extracted from the sentence.

Figure 1

- Closed Information Extraction

  : 사전에 정의된 relation types를 기준으로 정보를 추출

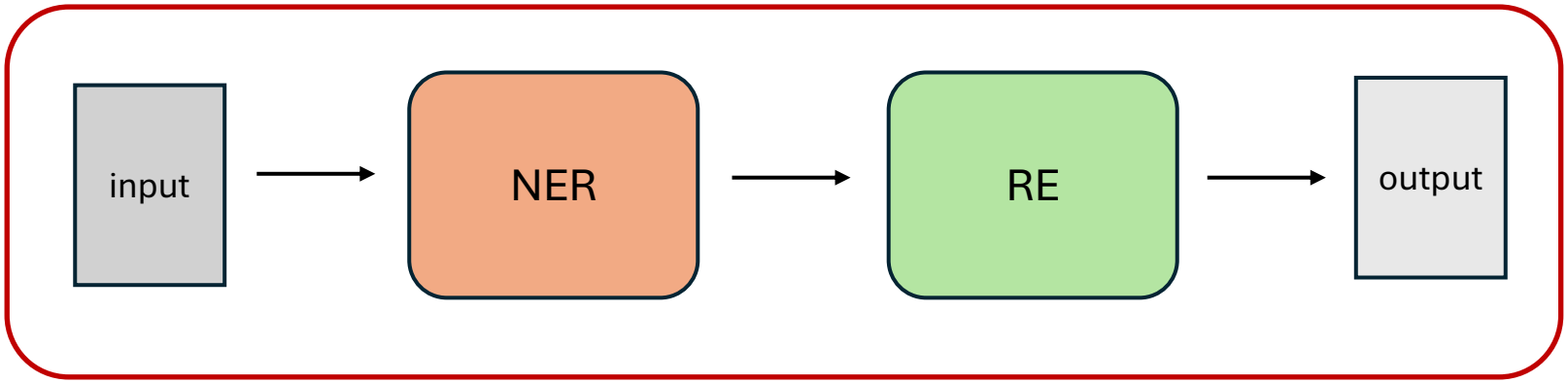
  Ex) located_in, used_for …

Figure1. [2022][IJCAI][-] A Survey on Neural Open Information Extraction: Current Status and Future Directions

Background

Pipeline Approach vs. Joint Approach

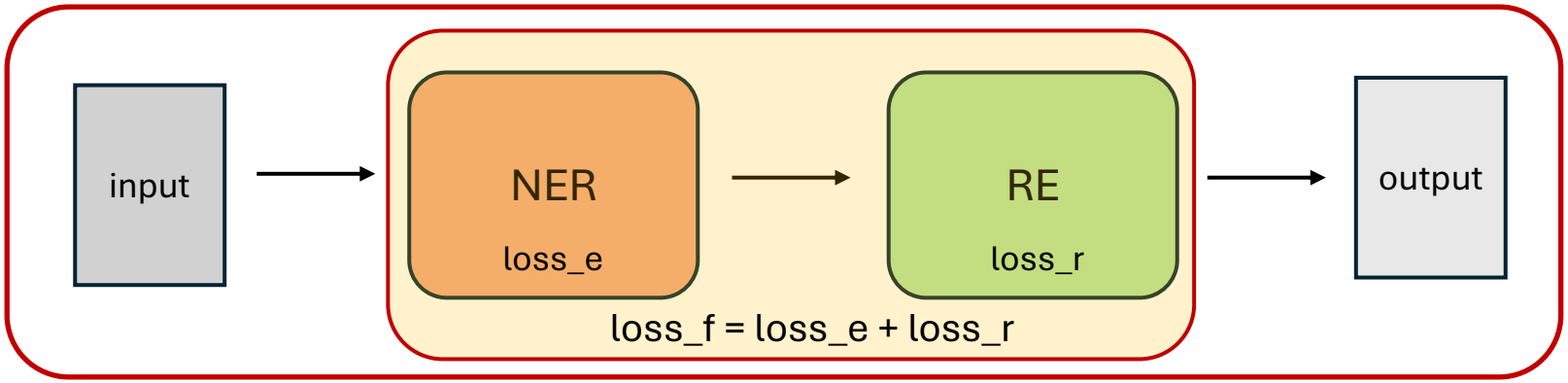
- Pipeline approach
  - ✓ NER, RE가 독립적인 모델에서 처리되고 학습됨
  - ✓ NER 모델의 아웃풋이 RE 모델에 입력됨
- Limitations
  - ✓ error propagation
  - ✓ 두 subtask 간 상호작용 불가



Background

Pipeline Approach vs. Joint Approach

- Joint Approach (Joint Information Extraction)
  - ✓ Extracting entity mentions and semantic relations between entities from unstructured text with a single model
  - ✓ 목표: jointly optimize [entity&relation extraction] in a single model

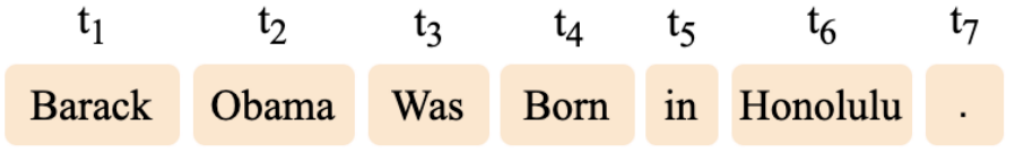


CONTENTS

JointIE 주요 모델

Span-based Approach

- Span: 연속된 n 개의 토큰 구간



Barak
Barak Obama
Barak Obama was
...
in Honolulu
Honolulu

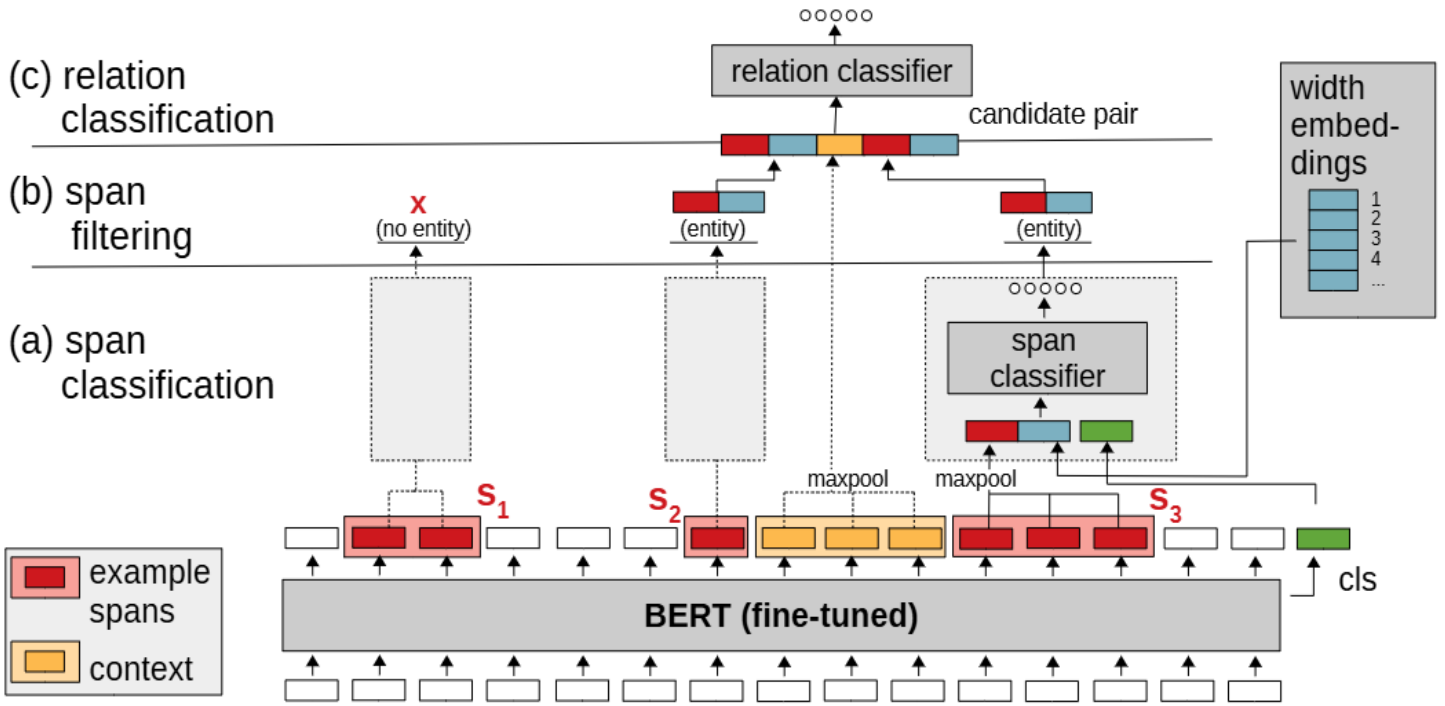
- Span-based 방식은 구간 단위로 처리함으로써 멀티 토큰의 엔티티를 탐지하기에 유리함

JointIE 주요 모델

Span-based Approach

- 문장 내 모든 연속된 토큰 구간을 후보 스팬으로 간주함
- 각 스팬이 실제 엔티티인지, 두 스팬 간 의미있는 관계가 존재하는지 판단
- 대표 모델:

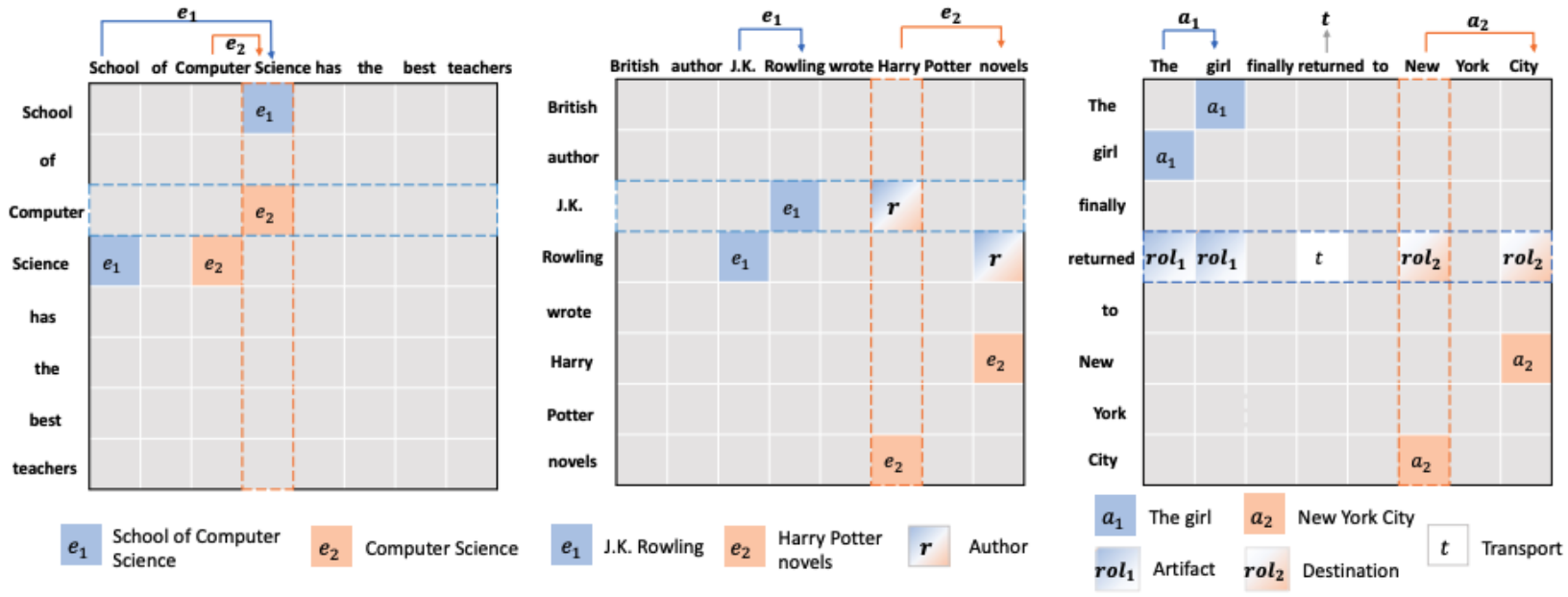
  [2020][ECAI][SpERT] Span-based Joint Entity and Relation Extraction with Transformer pre-training



JointIE 주요 모델

Token-pair-based Approach

- 문장 내 모든 토큰 쌍에 대해 이 쌍이 어떤 구조적 의미를 가지는지를 판단
  - ✓ 한 엔티티 스팬의 시작-끝 토큰 쌍 (start-end)
  - ✓ 관계를 이루는 두 엔티티의 start-start/end-end 토큰 쌍
- 2차원 매트릭스 상에서 모든 토큰 쌍의 관계를 직관적으로 다룰 수 있음



(a) Named Entity Recognition (b) Relation Extraction (c) Event Extraction

[2023][ACL][UTC-IE] A Unified Token-pair Classification Architecture for Information Extraction

JointIE 주요 모델

Token-pair-based Approach

- 대표 모델:

  [2023][ACL][UTC-IE] A Unified Token-pair Classification Architecture for Information Extraction

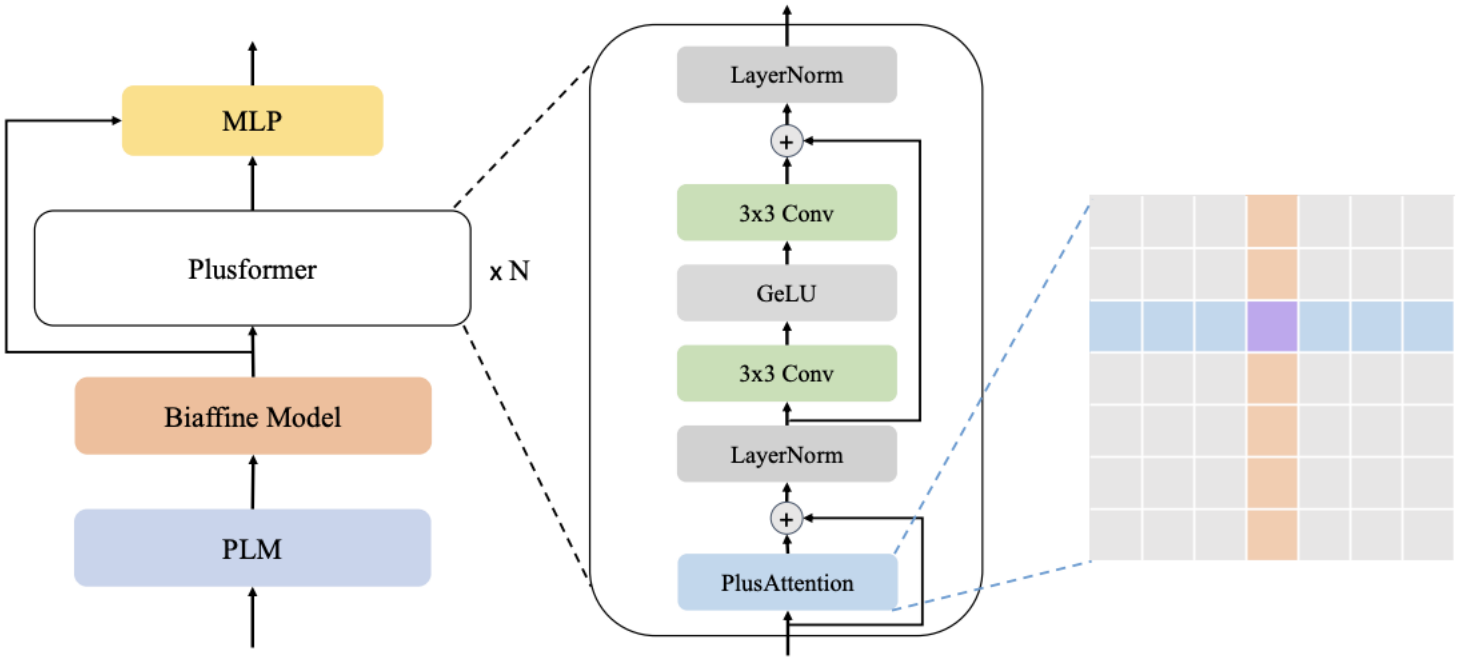


Figure 2: An overview of the UTC-IE Model.

JointIE 주요 모델

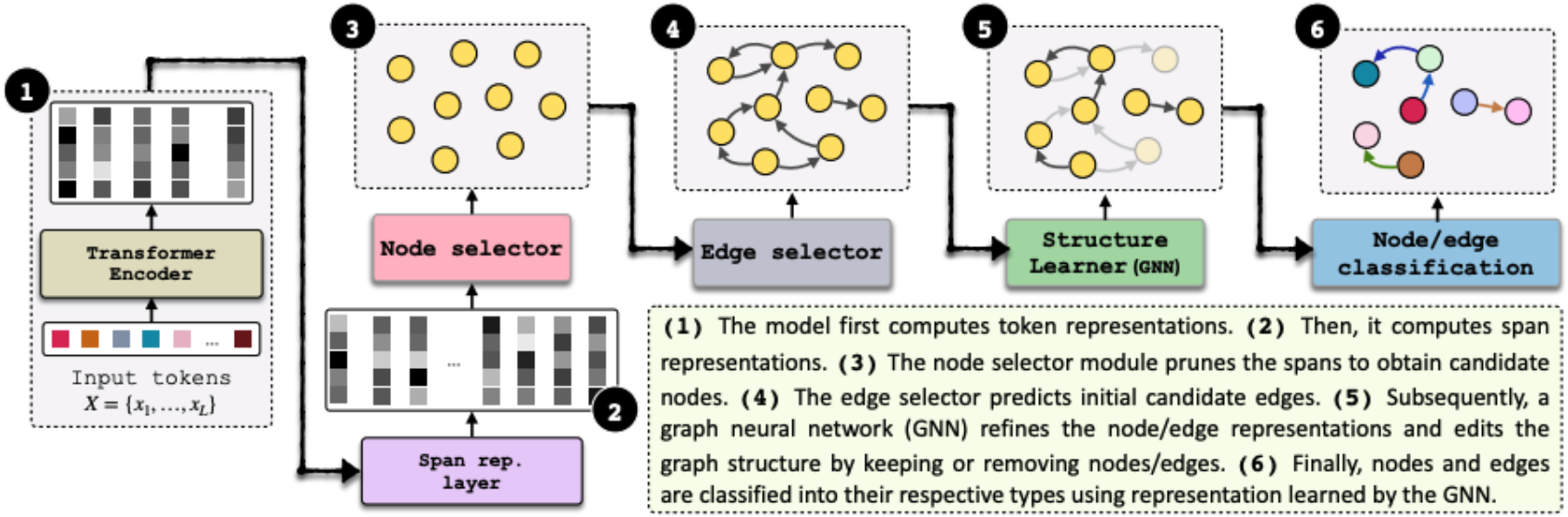
Graph-based Approach _ Text-to-Graph

- 대표 모델:

  [2024][arXiv][GraphER] A Structure-aware Text-to-Graph Model for Entity and Relation Extraction

  ✓ IE task를 Graph Structure Learning으로 정형화

  ✓ Dynamic Structure Refinement: 추출 과정에서 그래프 구조를 동적으로 개선하고 최적화함



$$\mathcal{L}_{total} = \mathcal{L}_{\mathcal{V}} + \mathcal{L}_{\mathcal{E}} + \mathcal{L}_{edit} + \mathcal{L}_{cls}$$

JointIE 주요 모델

Graph-based Approach _ Text-to-Graph

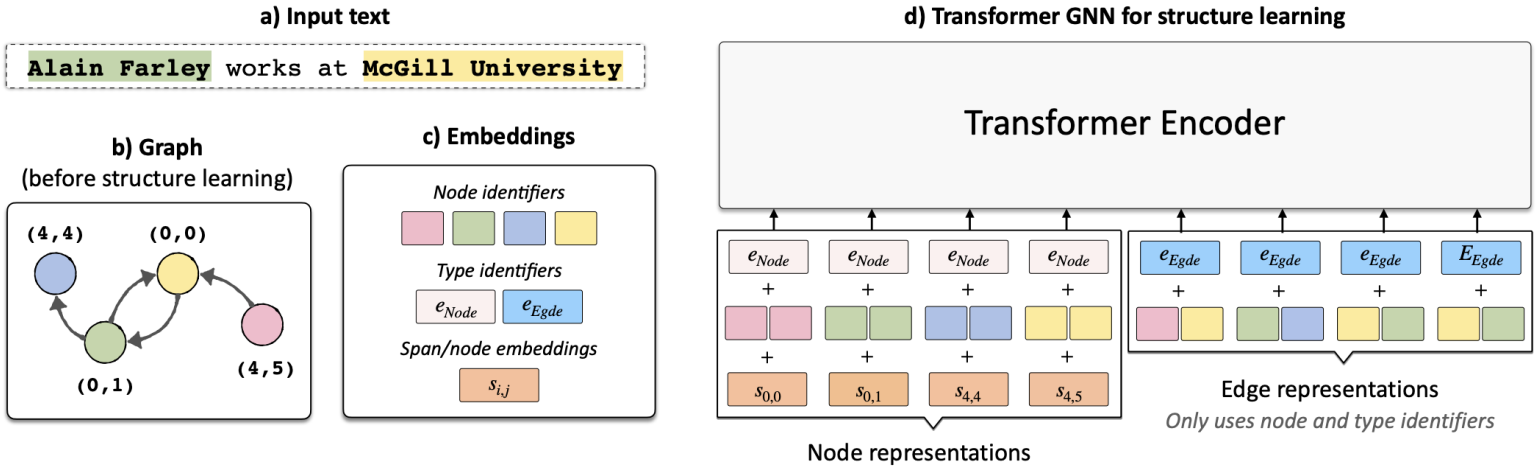
- 대표 모델:

  [2024][arXiv][GraphER] A Structure-aware Text-to-Graph Model for Entity and Relation Extraction

  ✓ IE task를 Graph Structure Learning으로 정형화

  ✓ Dynamic Structure Refinement: 추출 과정에서 그래프 구조를 동적으로 개선하고 최적화함

  ✓ GNN 기반의 Message Passing이 아닌 TokenGT를 도입하여 글로벌한 문맥을 고려할 수 있도록 함



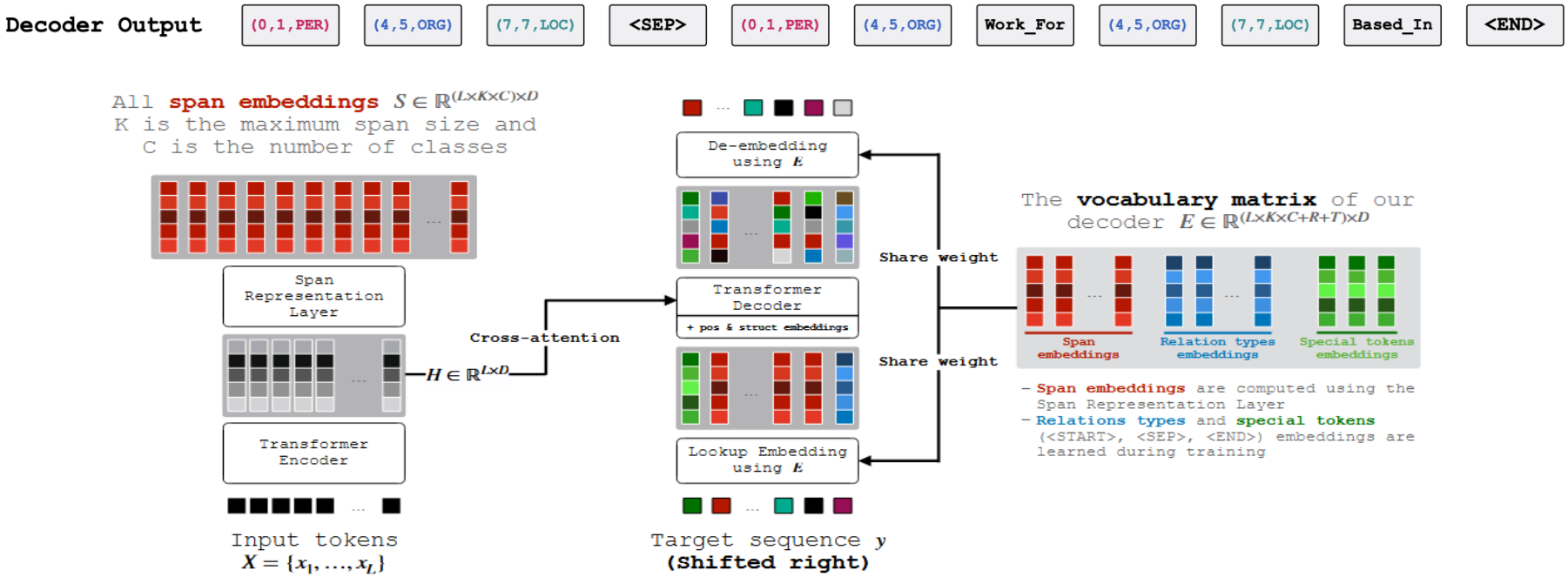
JointIE 주요 모델

Generative Approach

- Entity&relation 예측 태스크를 text generation 형태로 처리하는 방법 등장
- 대표 모델:

[2024][AAAI][ATG] An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction

- ✓ Pointing mechanism 사용 -> 생성된 출력이 원본 텍스트에 근거를 두어 문맥적 일관성을 높임
- ✓ Constrained decoding 도입 -> 생성되는 시퀀스가 올바른 그래프 구조를 유지하도록 보장함



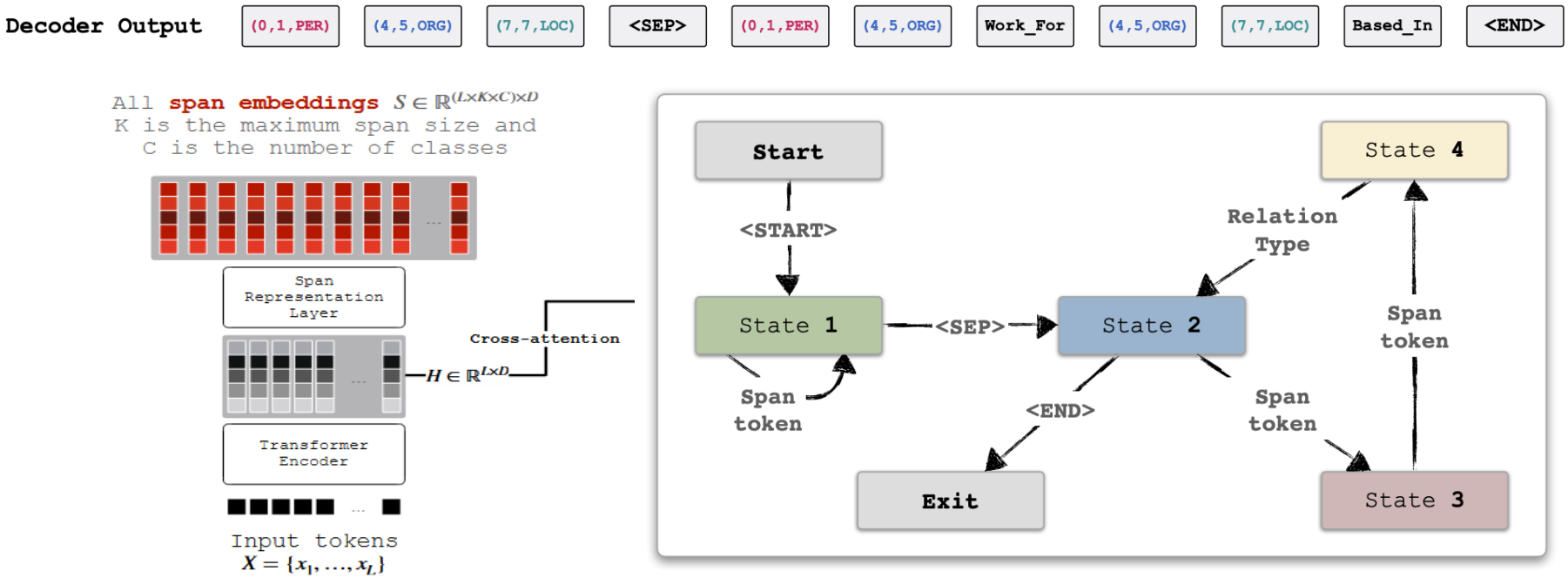
JointIE 주요 모델

Generative Approach

- Entity&relation 예측 태스크를 text generation 형태로 처리하는 방법 등장
- 대표 모델:

[2024][AAAI][ATG] An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction

- ✓ Pointing mechanism 사용 -> 생성된 출력이 원본 텍스트에 근거를 두어 문맥적 일관성을 높임
- ✓ Constrained decoding 도입 -> 생성되는 시퀀스가 올바른 그래프 구조를 유지하도록 보장함



CONTENTS

데이터셋 비교

JointIE _ ACE05, SciERC, CoNLL04 ...


```
{
  "clusters": [[[62, 64], [90, 91], [96, 98], [112, 114]], [[6, 6], [170, 170]], [[81, 82], [126, 127]], [[129, 131], [142, 142]]],
  "sentences": [
    ["In", "this", "paper", ",", "a", "novel", "method", "to", "learn", "the", "intrinsic", "object", "structure", "for", "robust", "visual", "tracking", "is", "proposed", "."],
    ["The", "basic", "assumption", "is", "that", "the", "parameterized", "object", "state", "lies", "on", "a", "low", "dimensional", "manifold", "and", "can", "be", "learned", "from", "training", "data", "."],
    ["Based", "on", "this", "assumption", ",", "firstly", "we", "derived", "the", "dimensionality", "reduction", "and", "density", "estimation", "algorithm", "for", "unsupervised", "learning", "of", "object", "intrinsic
    ["Secondly", "the", "dynamical", "model", "is", "derived", "and", "trained", "based", "on", "this", "intrinsic", "representation", "."],
    ["Thirdly", "the", "learned", "intrinsic", "object", "structure", "is", "integrated", "into", "a", "particle-filter", "style", "tracker", "."],
    ["We", "will", "show", "that", "this", "intrinsic", "object", "representation", "has", "some", "interesting", "properties", "and", "based", "on", "which", "the", "newly", "derived", "dynamical", "model", "makes", "p
    ["Experiments", "show", "that", "the", "learned", "tracker", "performs", "much", "better", "than", "existing", "trackers", "on", "the", "tracking", "of", "complex", "non-rigid", "motions", "such", "as", "fish", "twi
    ["The", "proposed", "method", "also", "has", "the", "potential", "to", "solve", "other", "type", "of", "tracking", "problems", "."]
  ],
  "ner": [
    [[6, 6, "Method"], [10, 12, "OtherScientificTerm"], [14, 16, "Task"]],
    [[28, 28, "OtherScientificTerm"], [32, 34, "OtherScientificTerm"]],
    [[52, 57, "Method"], [59, 64, "Task"], [62, 64, "Method"], [68, 72, "OtherScientificTerm"]],
    [[81, 82, "Method"], [90, 91, "Method"]], [[96, 98, "OtherScientificTerm"], [103, 105, "Method"]],
    [[112, 114, "Method"], [126, 127, "Method"], [129, 131, "Method"]],
    [[142, 142, "Generic"], [148, 148, "Generic"], [151, 155, "Task"], [153, 155, "OtherScientificTerm"], [158, 159, "OtherScientificTerm"], [161, 161, "Othe
    [[170, 170, "Generic"], [180, 181, "Task"]]
    ],
  "relations": [
    [[6, 6, 10, 12, "USED-FOR"], [10, 12, 14, 16, "USED-FOR"]],
    [[32, 34, 26, 28, "FEATURE-OF"]],
    [[52, 57, 59, 64, "USED-FOR"]],
    [[90, 91, 81, 82, "USED-FOR"]],
    [[96, 98, 103, 105, "PART-OF"]],
    [[126, 127, 129, 131, "USED-FOR"]],
    [[142, 142, 148, 148, "COMPARE"], [142, 142, 151, 155, "USED-FOR"], [148, 148, 151, 155, "USED-FOR"], [158, 159, 153, 155, "HYPONYM-OF"], [161, 161, 158,
    [[170, 170, 180, 181, "USED-FOR"]]
    ],
  "doc_key": "CVPR_2003_18_abs"
}
```


Dataset	$\|\mathcal{E}\|$	$\|\mathcal{R}\|$
ACE05	7	6
CoNLL 04	4	5
SciERC	6	7


```
{
  "sentences": ["Annie", "Oakley", ",", "also", "known", "as", "Little", "Miss", "Sure", "Shot", ",", "was", "born", "Phoebe", "Ann", "Moses", "in", "Willowdell", ",", "Darke", "County", ",", "in", "1860", "."],
  "ner": [[0, 1, "Peop"], [6, 9, "Peop"], [13, 15, "Peop"], [17, 20, "Loc"]],
  "relations": [[0, 1, 17, 20, "Live_In"], [6, 9, 17, 20, "Live_In"], [13, 15, 17, 20, "Live_In"]],
  "clusters": [],
  "doc_key": "conll04_5284"
}
```

데이터셋 비교

(Relational) Triple Extraction _ WebNLG, NYT

…


```
{
    "text": "Peter Stöger is manager of 1 . FC Köln which has 50000 members and participated in the 2014 season .",
    "id": "train_0",
    "relation_list": [
        {"subject": "1 . FC Köln", "object": "Peter Stöger", "subj_char_span": [27, 38], "obj_char_span": [0, 12], "predicate": "manager", "subj_tok_span": [7, 13], "obj_tok_span": [0, 4]},
        {"subject": "1 . FC Köln", "object": "2014", "subj_char_span": [27, 38], "obj_char_span": [87, 91], "predicate": "season", "subj_tok_span": [7, 13], "obj_tok_span": [22, 23]},
        {"subject": "1 . FC Köln", "object": "50000", "subj_char_span": [27, 38], "obj_char_span": [49, 54], "predicate": "numberOfMembers", "subj_tok_span": [7, 13], "obj_tok_span": [15, 17]}
        ],
    "entity_list": [
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]},
        {"text": "Peter Stöger", "type": "DEFAULT", "char_span": [0, 12], "tok_span": [0, 4]},
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]},
        {"text": "2014", "type": "DEFAULT", "char_span": [87, 91], "tok_span": [22, 23]},
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]},
        {"text": "50000", "type": "DEFAULT", "char_span": [49, 54], "tok_span": [15, 17]}
        ]
},
{
    "text": "Peter Stöger is manager of 1 . FC Köln who were in the 201415 Bundesliga season and have 50000 members .",
    "id": "train_1",
    "relation_list": [
        {"subject": "1 . FC Köln", "object": "Peter Stöger", "subj_char_span": [27, 38], "obj_char_span": [0, 12], "predicate": "manager", "subj_tok_span": [7, 13], "obj_tok_span": [0, 4]},
        {"subject": "1 . FC Köln", "object": "201415 Bundesliga", "subj_char_span": [27, 38], "obj_char_span": [55, 72], "predicate": "season", "subj_tok_span": [7, 13], "obj_tok_span": [17, 20]},
        {"subject": "1 . FC Köln", "object": "50000", "subj_char_span": [27, 38], "obj_char_span": [89, 94], "predicate": "numberOfMembers", "subj_tok_span": [7, 13], "obj_tok_span": [23, 25]}
        ],
    "entity_list": [
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]}, {"text": "Peter Stöger", "type": "DEFAULT", "char_span": [0, 12], "tok_span": [0, 4]},
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]}, {"text": "201415 Bundesliga", "type": "DEFAULT", "char_span": [55, 72], "tok_span": [17, 20]},
        {"text": "1 . FC Köln", "type": "DEFAULT", "char_span": [27, 38], "tok_span": [7, 13]}, {"text": "50000", "type": "DEFAULT", "char_span": [89, 94], "tok_span": [23, 25]}
        ]
},
```


Corpus Name	Relation
NYT [125]	24
WebNLG [51]	171

Background

(Relational) Triple Extraction

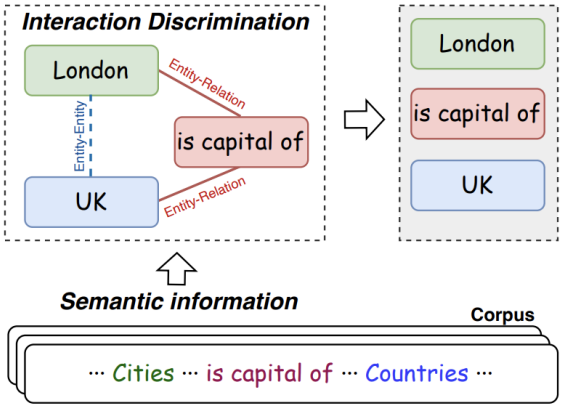
- Raw text로부터 <subject, relation, object> 형태의 트리플을 추출
- 엔티티와 그들 사이의 의미적 관계를 동시에 식별하며,
- 지식 그래프 구축을 위한 핵심 단계로 간주됨

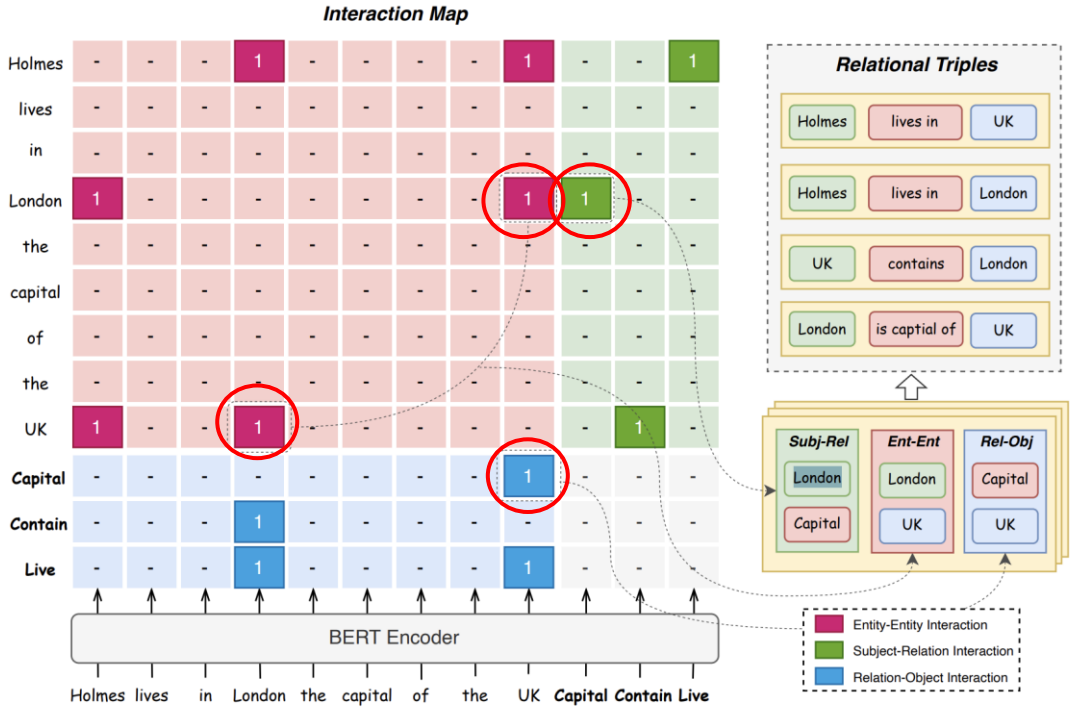
[2022][EMNLP][UniRel] Unified Representation and Interaction for Joint Relational Triple Extraction

주요 모델

[2022][EMNLP][UniRel]
Unified Representation and Interaction for Joint Relational Triple Extraction

- entity-entity, entity-relation 사이의 interaction을 통합적으로 처리
- Interaction Map 메커니즘 활용





CONTENTS

Ontology

[2023][ISWC][Text2KGBench]
A Benchmark for Ontology-Driven Knowledge Graph Generation From Text

- 온톨로지를 모델에 명시적으로 제공한 상태에서, 정확하고 신뢰할 수 있는 트리플을 추출할 수 있는지를 평가함
- 온톨로지 기반 지식그래프를 생성하는 LLM의 능력을 평가하기 위한 벤치마크를 제안함
- 모델 아웃풋: 온톨로지에 맞는 fact triple 목록
  - ✓ 이때 모델이 준수해야 할 원칙
    - Ontology conformance
    - Faithfulness: 입력된 문장에서 언급된 사실만 사용 (원문에 충실)
    - No hallucination: 허구의 엔티티/관계를 생성하지 않음

Text2KGBench


```
{
        "title": "Movie Ontology",
        "id": "ont_1_movie",
        "concepts": [
                {"qid": "Q5", "label": "human"},
                {"qid": "Q515", "label": "city"},
                {"qid": "Q6256", "label": "country"},
                {"qid": "Q11424", "label": "film"},
                {"qid": "Q201658", "label": "film genre"},
                {"qid": "Q483394", "label": "genre"},
                {"qid": "Q1762059", "label": "film production company"},
                {"qid": "Q4220917", "label": "film award"},
                {"qid": "Q618779", "label": "award"},
                {"qid": "Q47461344", "label": "written work"},
                {"qid": "Q15773347", "label": "film character"},
                {"qid": "Q104649845", "label": "film organization"}
        ],
        "relations": [
                {"pid": "P57", "label": "director", "domain": "Q11424", "range": "Q5"},
                {"pid": "P58", "label": "screenwriter", "domain": "Q11424", "range": "Q5"},
                {"pid": "P136", "label": "genre", "domain": "Q11424", "range": "Q483394"},
                {"pid": "P144", "label": "based on", "domain": "Q11424", "range": "Q47461344"},
                {"pid": "P161", "label": "cast member", "domain": "Q11424", "range": "Q5"},
                {"pid": "P166", "label": "award received", "domain": "Q11424", "range": "Q618779"},
                {"pid": "P272", "label": "production company", "domain": "Q11424", "range": "Q1762059"},
                {"pid": "P495", "label": "country of origin", "domain": "Q11424", "range": "Q6256"},
                {"pid": "P577", "label": "publication date", "domain": "Q11424", "range": ""},
                {"pid": "P674", "label": "characters", "domain": "Q11424", "range": "Q15773347"},
                {"pid": "P840", "label": "narrative location", "domain": "Q11424", "range": "Q515"},
                {"pid": "P915", "label": "filming location", "domain": "Q11424", "range": "Q515"},
                {"pid": "P921", "label": "main subject", "domain": "Q11424", "range": ""},
                {"pid": "P1411", "label": "nominated for", "domain": "Q11424", "range": "Q618779"},
                {"pid": "P2130", "label": "cost", "domain": "Q11424", "range": ""}
        ]
}
```


data/wikidata_tekgen/ontologies/1_movie_ontology.json

Text2KGBench


```
{
    "title": "Movie Ontology",
    "id": "ont_1_movie",
    "concepts": [
        {"qid": "Q5", "label": "human"},
        {"qid": "Q515", "label": "city"},
        {"qid": "Q6256", "label": "country"},
        {"qid": "Q11424", "label": "film"},
        {"qid": "Q201658", "label": "film genre"},
        {"qid": "Q483394", "label": "genre"},
        {"qid": "Q1762059", "label": "film production
        {"qid": "Q4220917", "label": "film award"},
        {"qid": "Q618779", "label": "award"},
        {"qid": "Q47461344", "label": "written work"},
        {"qid": "Q15773347", "label": "film character"
        {"qid": "Q104649845", "label": "film organizat
    ],
    "relations": [
        {"pid": "P57", "label": "director", "domain": '
        {"pid": "P58", "label": "screenwriter", "domai
        {"pid": "P136", "label": "genre", "domain": "Q
        {"pid": "P144", "label": "based on", "domain":
        {"pid": "P161", "label": "cast member", "domai
        {"pid": "P166", "label": "award received", "do
        {"pid": "P272", "label": "production company",
        {"pid": "P495", "label": "country of origin", '
        {"pid": "P577", "label": "publication date", "
        {"pid": "P674", "label": "characters", "domain'
        {"pid": "P840", "label": "narrative location",
        {"pid": "P915", "label": "filming location", "
        {"pid": "P921", "label": "main subject", "doma
        {"pid": "P1411", "label": "nominated for", "do
        {"pid": "P2130", "label": "cost", "domain": "Q
    ]
}
```


Given the following ontology, examples and sentences, please extract the triples from the sentence according to the relations in the ontology. In the output, only include the triples in the given output format.

CONTEXT:

Ontology Concepts: *human, city, country, film, film genre, film production company, film award, award, written work, film character, film organization*

Ontology Relations: cast_member(film,human), director (film,human), screenwriter (film,human), producer(film,human), genre(film,genre), based_on(film,written work), award_received (film,award), production_company(film,film production company), country_of_origin(film,country), publication_date (film,date), characters(film,film character), narrative_location(film,city), filming_location(film,city), main_subject(film,thing), nominated_for(film,award), cost(film,number)

Example Sentence: *The Lion King is a animated musical drama film about a lion cub who is to succeed his father and it was directed by Roger Allers and Rob Minkoff (in their feature directorial debuts), produced by Don Hahn.*

Example Output: director(Lion King, Roger Allers)
director(Lion King, Rob Minkoff)
producer(Lion King, Don Hahn)

Test Sentence: Birds Anonymous is a 1957 Warner Bros. Merrie Melodies animated short, directed by Friz Freleng and written by Warren Foster.

Test Output:

screenwriter(Birds Anonymous, Warren Foster)
director(Birds Anonymous, Friz Freleng)
publication_date(Birds Anonymous, 1957)
production_company(Birds Anonymous, Warner Bros.)
genre(Birds Anonymous, animated film)

Instruction

Verbalized Ontology

Example(s)

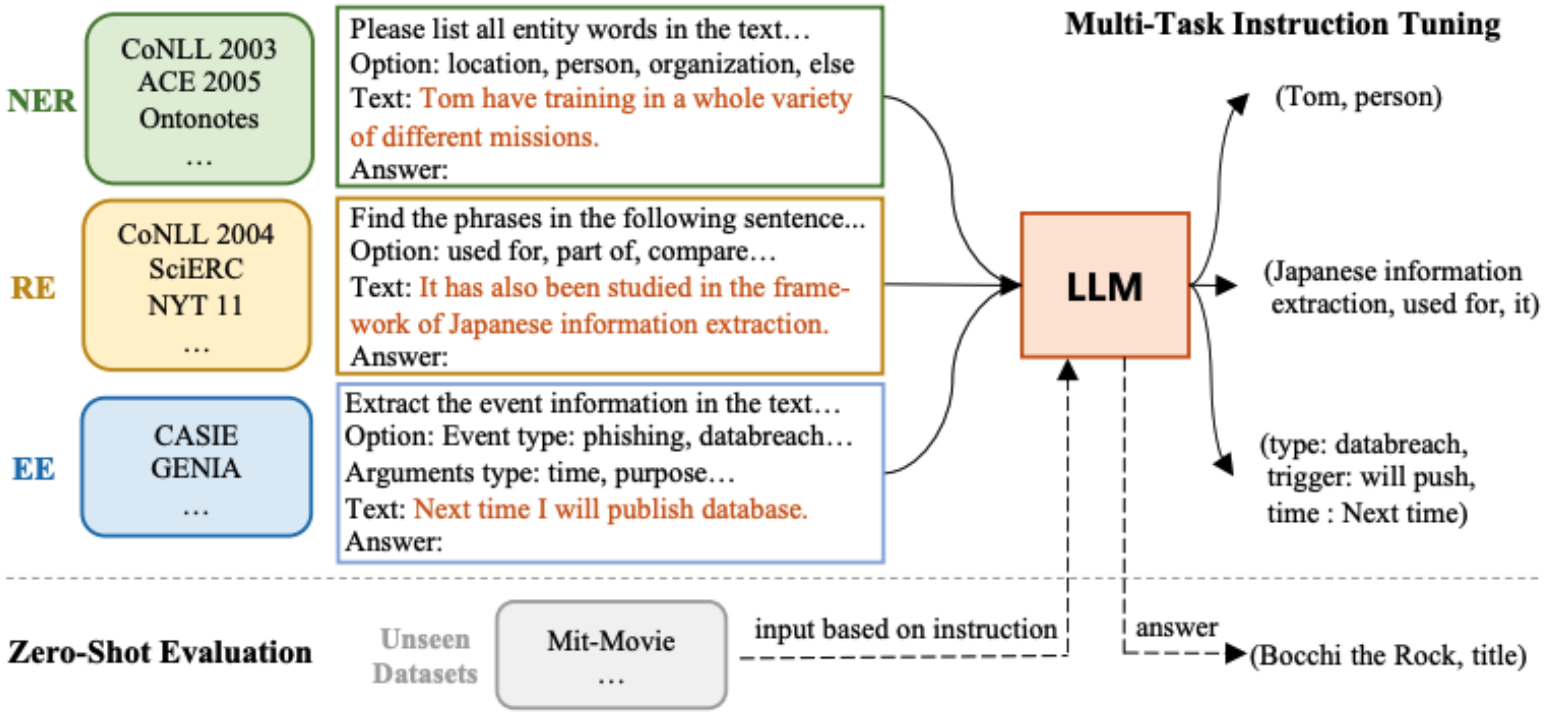
Input Sentence(s)

LLM Output

Instruction Tuning

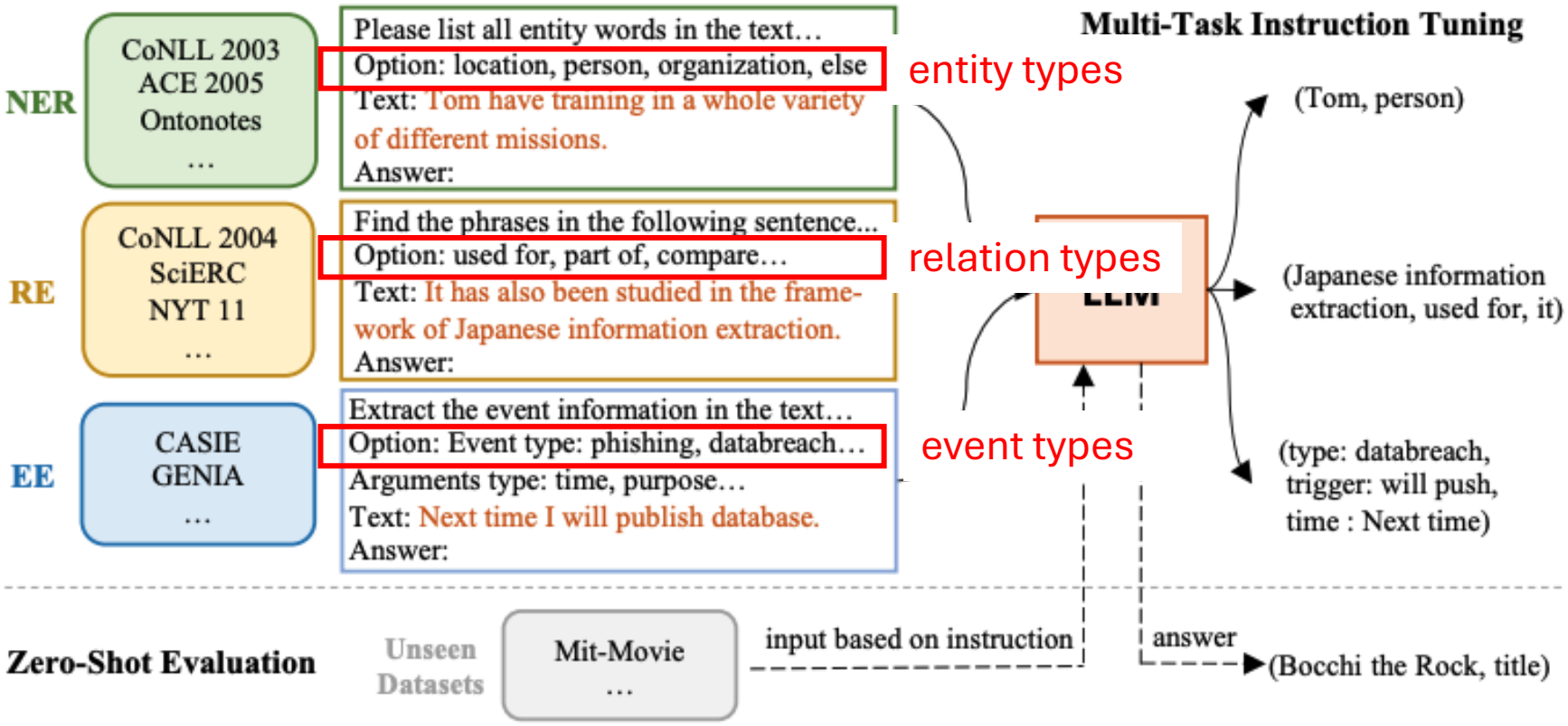
[2023][arXiv][InstructUIE]
Multi-task Instruction Tuning for Unified Information Extraction

- IE task들(NER, RE, EE)을 text-to-text 형식으로 재정의
- Instruction tuning을 통해 LLM의 출력을 유도함
- 32개의 IE 데이터셋으로 IE INSTRUCTIONS 벤치마크 데이터셋 구축



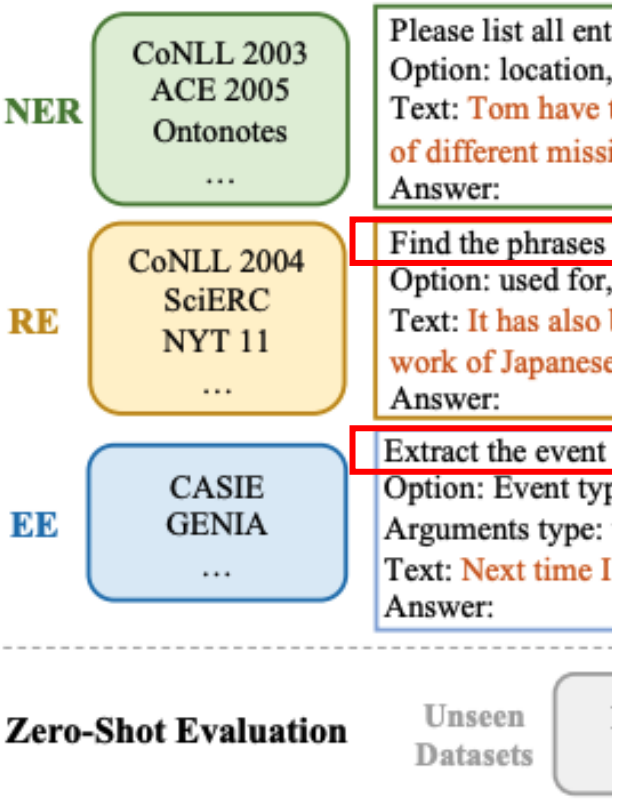
InstructUIE

- Task schema 구성: task instruction, option, text, output



InstructUIE

- Task schema 구성: task inst

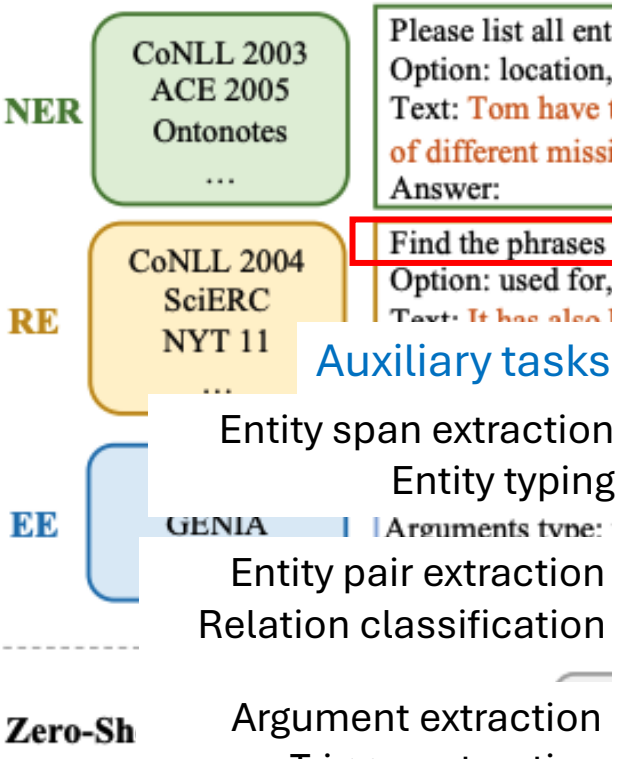


Task	Prompts
NER	Please list all entity words in the text that fit the category. Output format is "type1: word1; type2: word2". Please find all the entity words associated with the category in the given text. Output format is "type1: word1; type2: word2". Please tell me all the entity words in the text that belong to a given category. Output format is "type1: word1; type2: word2".
RE	Given a phrase that describes the relationship between two words, extract the words and the lexical relationship between them. The output format should be "relation1: word1, word2; relation2: word3, word4". Find the phrases in the following sentences that have a given relationship. The output format is "relation1: word1, word2; relation2: word3, word4". Given a sentence, please extract the subject and object containing a certain relation in the sentence according to the following relation types, in the format of "relation1: word1, word2; relation2: word3, word4".
EE	Locate the role in the text that participated in the event based on the event type and return it in the event list. Extract the event information in the text and return them in the event list.
ES	Please list all entity words in the text that fit the category. Output format is word1, word2.
ET	Given options, please tell me the categories of all the listed entity words.Output format is "type1: word1; type2: word2".
EP	Please list all entity pairs containing a certain relationship in the given options.Output format is "word1, word2; word3, word4".
EPR	Given options, please tell me the relationships of all the listed entity pairs.Output format is "relation1: word1, word2; relation2: word3, word4".
EEA	Given event type and trigger, please tell me the arguments of all the listed option. Output format is "name: role".
EET	Please tell me event type and its trigger word from given type options. Output format is "event type: trigger".

Table 8: Instructions for different tasks.

InstructUIE

- Task schema 구성: task inst



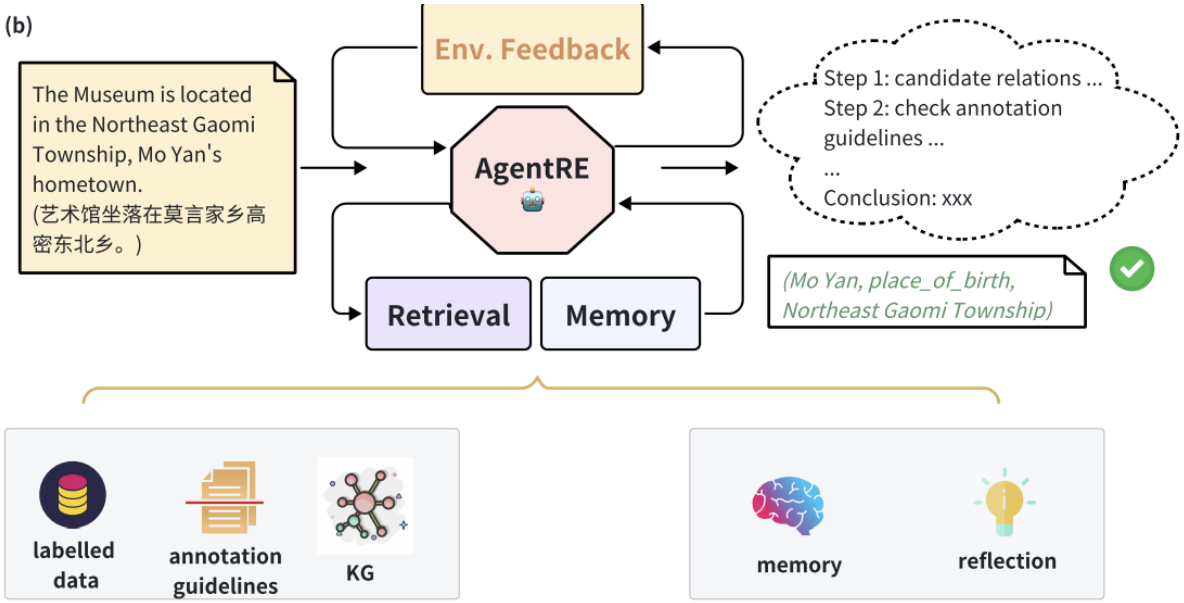
Task	Prompts
NER	Please list all entity words in the text that fit the category. Output format is "type1: word1; type2: word2". Please find all the entity words associated with the category in the given text. Output format is "type1: word1; type2: word2". Please tell me all the entity words in the text that belong to a given category. Output format is "type1: word1; type2: word2".
RE	Given a phrase that describes the relationship between two words, extract the words and the lexical relationship between them. The output format should be "relation1: word1, word2; relation2: word3, word4". Find the phrases in the following sentences that have a given relationship. The output format is "relation1: word1, word2; relation2: word3, word4". Given a sentence, please extract the subject and object containing a certain relation in the sentence according to the following relation types, in the format of "relation1: word1, word2; relation2: word3, word4".
EE	Locate the role in the text that participated in the event based on the event type and return it in the event list. Extract the event information in the text and return them in the event list.
ES	Please list all entity words in the text that fit the category. Output format is word1, word2.
ET	Given options, please tell me the categories of all the listed entity words.Output format is "type1: word1; type2: word2".
EP	Please list all entity pairs containing a certain relationship in the given options.Output format is "word1, word2; word3, word4".
EPR	Given options, please tell me the relationships of all the listed entity pairs.Output format is "relation1: word1, word2; relation2: word3, word4".
EEA	Given event type and trigger, please tell me the arguments of all the listed option. Output format is "name: role".
EET	Please tell me event type and its trigger word from given type options. Output format is "event type: trigger".

Table 8: Instructions for different tasks.

Agent

[2024][CIKM][AgentRE]
An Agent-Based Framework for Navigating Complex Information Landscapes in Relation Extraction

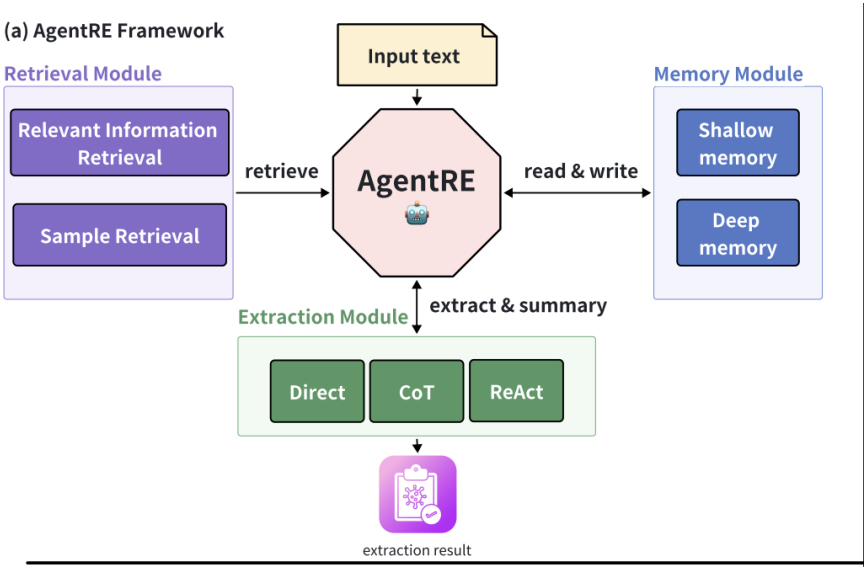
- LLM을 agent로 보고, 다양한 외부 정보에 접근하여 다중 단계 추론을 수행
  - ✓ 단일 text-in, text-out 방식이 아닌, 다중 라운드의 reasoning을 통해 정보를 탐색하고 축적
  - ✓ Retrieval, memory, extraction 모듈 도입
  - ✓ 추론 경로를 기록하여 소형 모델 학습에 활용 (Distillation)
    _ Reasoning Trajectory를 학습 데이터로 전환



Agent

[2024][CIKM][AgentRE]
An Agent-Based Framework for Navigating Complex Information Landscapes in Relation Extraction

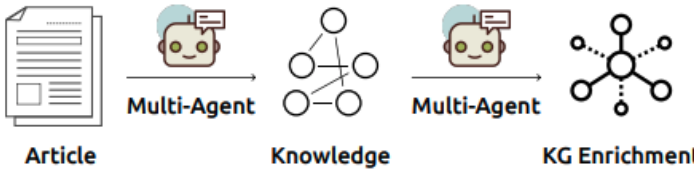
- 주요 모듈
  - ✓ Retrieval: static knowledge를 다루며, 훈련 샘플, 주석 가이드 등 외부 정적 정보를 검색함
  - ✓ Memory: dynamic knowledge를 다루며, 과거 추출 경험을 기록하고 반성(reflect)하며, 추론 과정에서의 행동을 요약 후 반성함
  - ✓ Extraction: 직접 추출, 단계적 추출, CoT 기반 등 다양한 추론 방식으로 관계 추출 수행
  - ✓ LLM은 이 모듈들을 호출하며, 능동적으로 정보 탐색 및 reasoning을 수행함



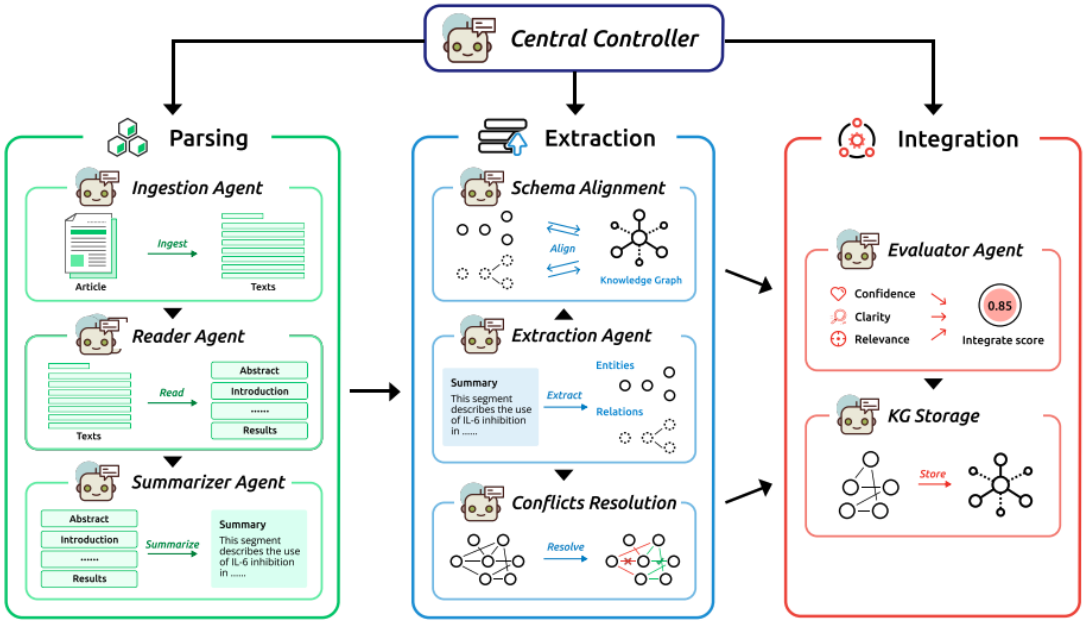
Multi-Agent

[2025][arXiv][KARMA]
Leveraging Multi-Agent LLMs for Automated Knowledge Graph Enrichment

- Multi-agent 시스템을 활용하여 knowledge graph enrichment 수행
  - ✓ 인풋 문서로부터 새로운 트리플을 추출하여 기존의 KG에 추가



  - ✓ 총 9개의 에이전트로 구성되며, 그중 Central Controller Agent가 이들을 통제함



CONTENTS

논문 _ 심사 중

[RePooL] Relation-Faceted Graph Pooling with LLM Guidance for Dynamic Span-Aware Information Extraction

- 기존 방법론의 한계)
  - ✓ Static span formation:

    후보 스팬은 relational context 앞단에 결정되며, 스팬 임베딩은 보통 boundary 토큰 임베딩을 concatenate 하는 형태로 구성되어, 멀티토큰 엔티티의 경우 중간 토큰 임베딩은 무시됨
  - ✓ Relation-agnostic semantic processing:

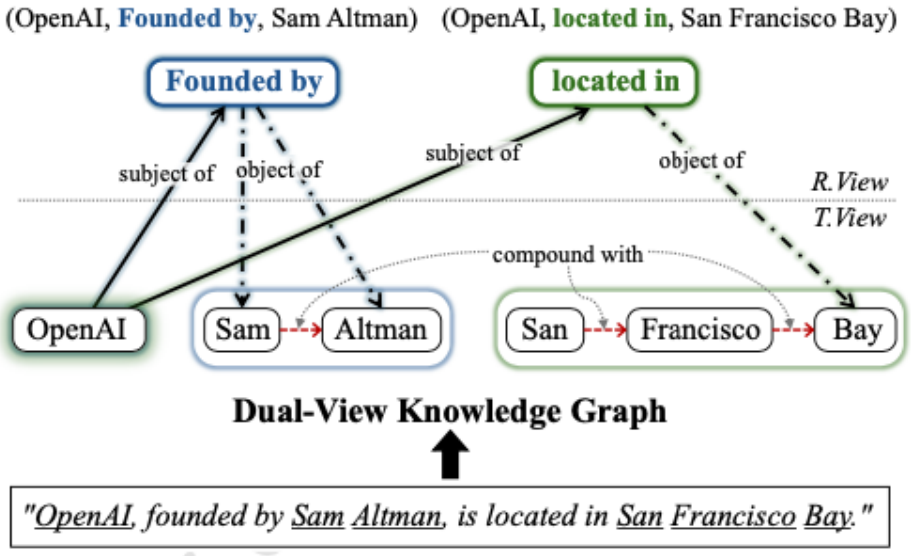
    Relation-specific context에 대한 고려 없이 span의 유효성을 평가함

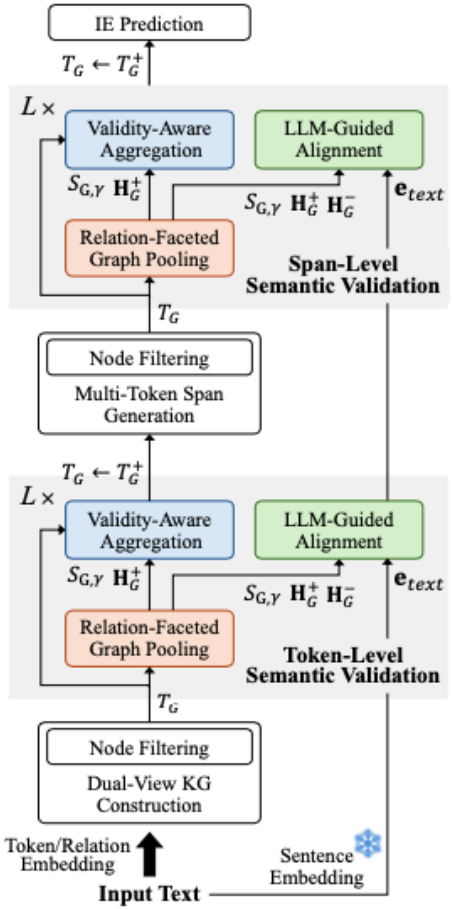
    유효성 평가에 명시적인 기준이 없기 때문에 LLM의 semantic guidance를 받을 수 없음

논문 _ 심사 중

[RePooL] Relation-Faceted Graph Pooling with LLM Guidance for Dynamic Span-Aware Information Extraction

- 제안 모델: RePooL)
  - ✓ Dual-view knowledge graph _ token and relation nodes
  - ✓ Auxiliary structural relations (edge):
    'subject_of', 'object_of', 'compound_with'
  - ✓ LLM preference alignment





현재 진행 중인 연구

- 기본 baseline: generative JointIE (ATG 모델)
- Constraint decoding
  - ✓ 온톨로지 정보
  - ✓ Type 간의 관계성
    - Entity type – relation type (연결 빈도 등)

  …
- 구현 방법
  - ✓ Instruction tuning
  - ✓ LLM agent
  - ✓ Embedding 학습

  …