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

Hye-Yoon Baek, Jinho Choi, Jimyeung Seo,
Xiongnan Jin, Dongcheon Lee, Byungkook Oh

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

hannah100@konkuk.ac.kr / hyeyoonbaek@gmail.com

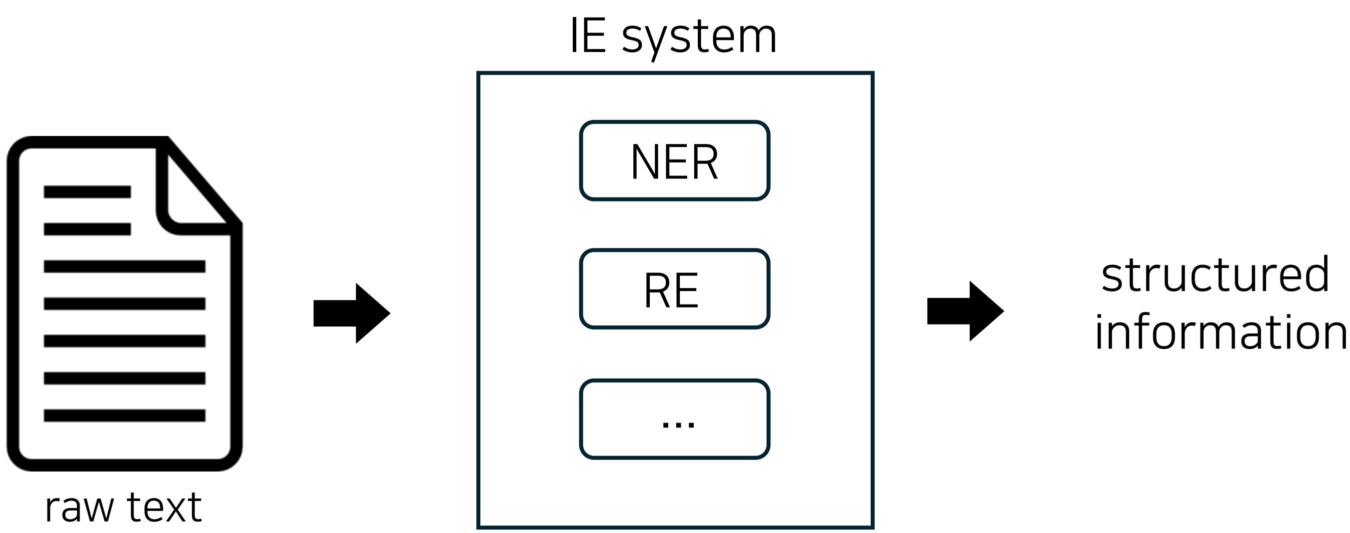
2025.11.13

CONTENTS

Background

Information Extraction

- Convert unstructured text into structured information



Background

Joint Information Extraction

- Convert unstructured text into structured information
- Simultaneously extract semantically valid triples in the form of `(span, relation, span)` from raw text

"*OpenAI, founded by Sam Altman, is located in San Francisco Bay.*"

⬇

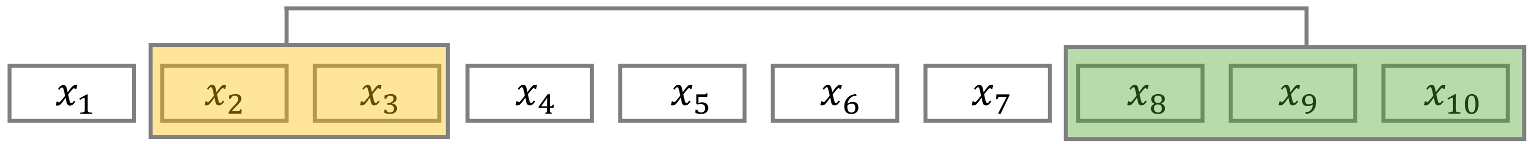
`(OpenAI, founded_by, Sam Altman)`

`(OpenAI, located_in, San Francisco Bay)`

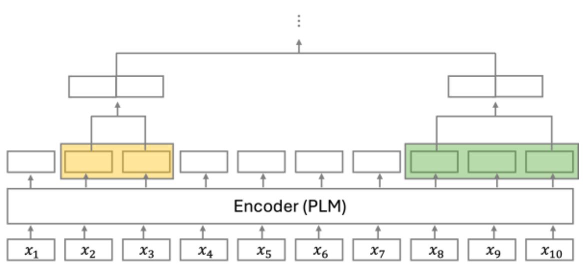
Background

Existing methods

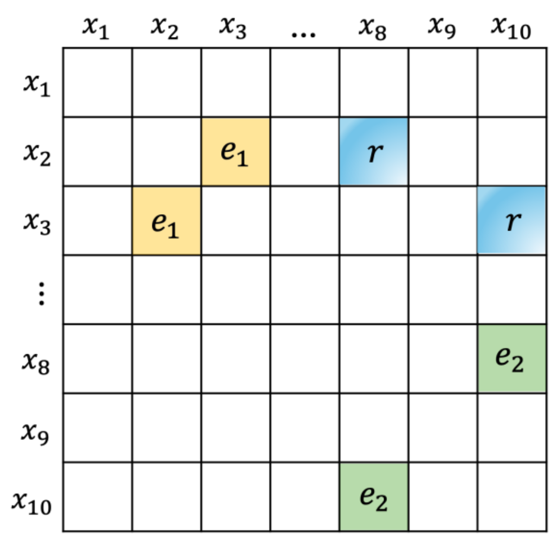
input sentence



Span-based



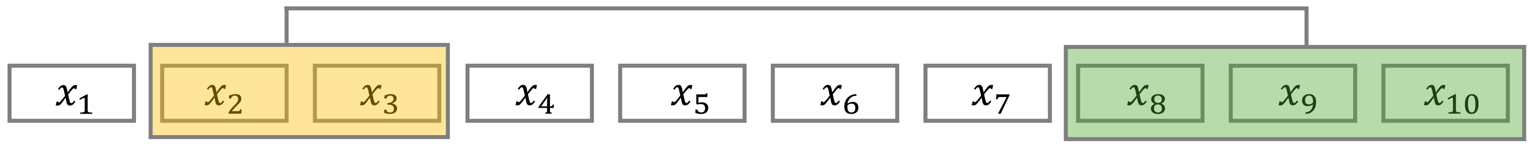
Token pair-based



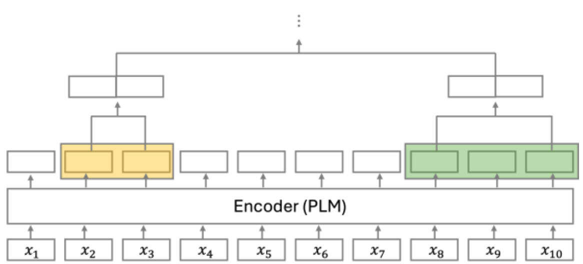
Background

Existing methods

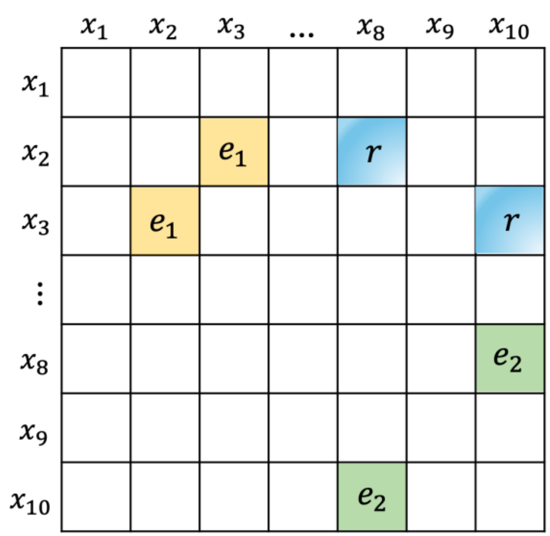
input sentence



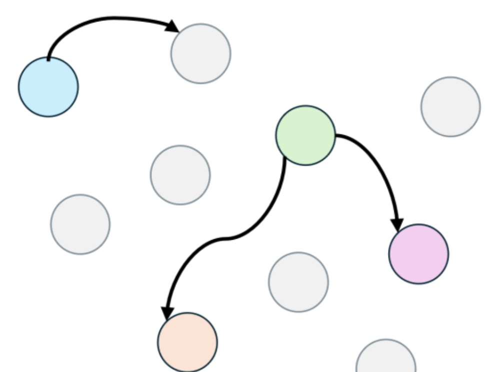
Span-based



Token pair-based



Text-to-graph



Challenges

Limitation

- #1. Static Span Formation
  - ✓ candidate spans are predefined and validated independently of relational context
  - ✓ span embeddings are typically constructed by concatenating only boundary tokens, ignoring intermediate tokens within multi-token spans

- #2. Relation-Agnostic Semantic Processing
  - ✓ assess span and span-pair validity without explicitly incorporating relation-specific contexts
  - ✓ prevents effective semantic grounding from LLMs
    - as LLMs cannot provide clear criteria for relation-agnostic validation

Challenges

Limitation

- #1. Static Span Formation
  - ✓ candidate spans are predefined and validated independently of relational context
  - ✓ span embeddings are typically constructed by concatenating only boundary tokens, ignoring intermediate tokens within multi-token spans

- #2. Relation-Agnostic Semantic Processing
  - ✓ evaluate spans and span-pairs without explicitly incorporating relation-specific contexts
  - ✓ prevents from leveraging effective semantic grounding from LLMs
    - as LLMs cannot provide clear criteria for relation-agnostic validation

Contributions

- **Relation-faceted text-to-graph framework**
  - ✓ explicitly integrates relation semantics and LLM-guided alignment to perform dynamic, context-aware span validation

- **Dynamic span composition with structural relations**
  - ✓ introduces auxiliary structural relations
  - ✓ preserve fine-grained intra-span information

- **Hierarchical Relation-faceted reasoning**
  - ✓ performs token-level to span-level triple validation
  - ✓ guided by LLM-based preference alignment for semantic consistency and robustness

CONTENTS

Proposed Method (RePooL)

Problem Definition

- The primary goal is to identify and extract semantically valid triples, simultaneously classifying entity types within valid entity spans
  - ✓ head and tail spans are contextually appropriate entities for the given relation

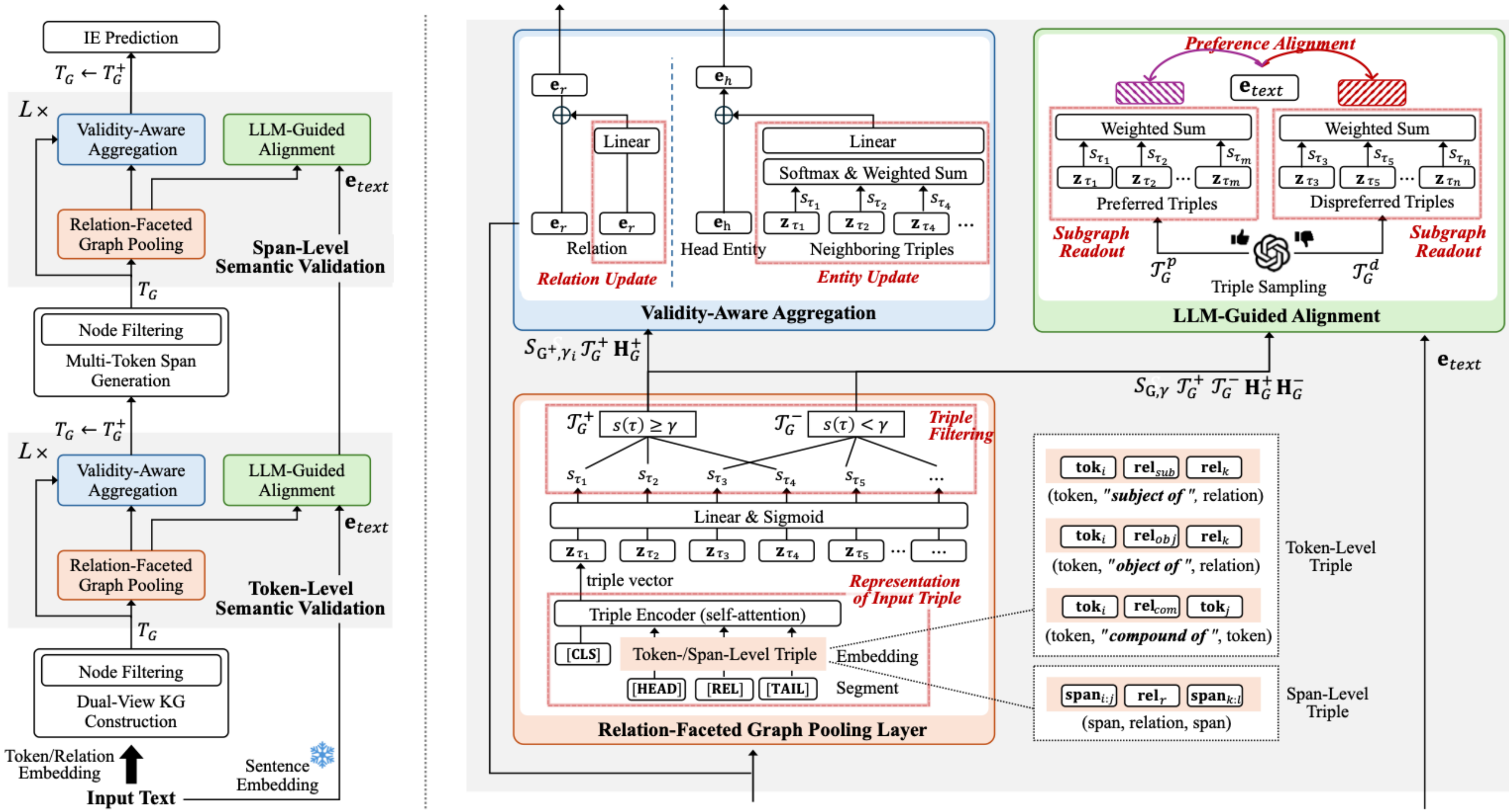
$$\mathcal{Y} = \{\langle h, r, t \rangle \mid h, t \in \{span_{i:j}\}, r \in \{rel_k\}\}$$

  - ✓ learn a scoring function that evaluates the likelihood of each candidate triple being contextually appropriate and semantically meaningful

$$\text{score}(\tau) : \mathcal{T} \rightarrow [0, 1]$$

Proposed Method (RePooL)

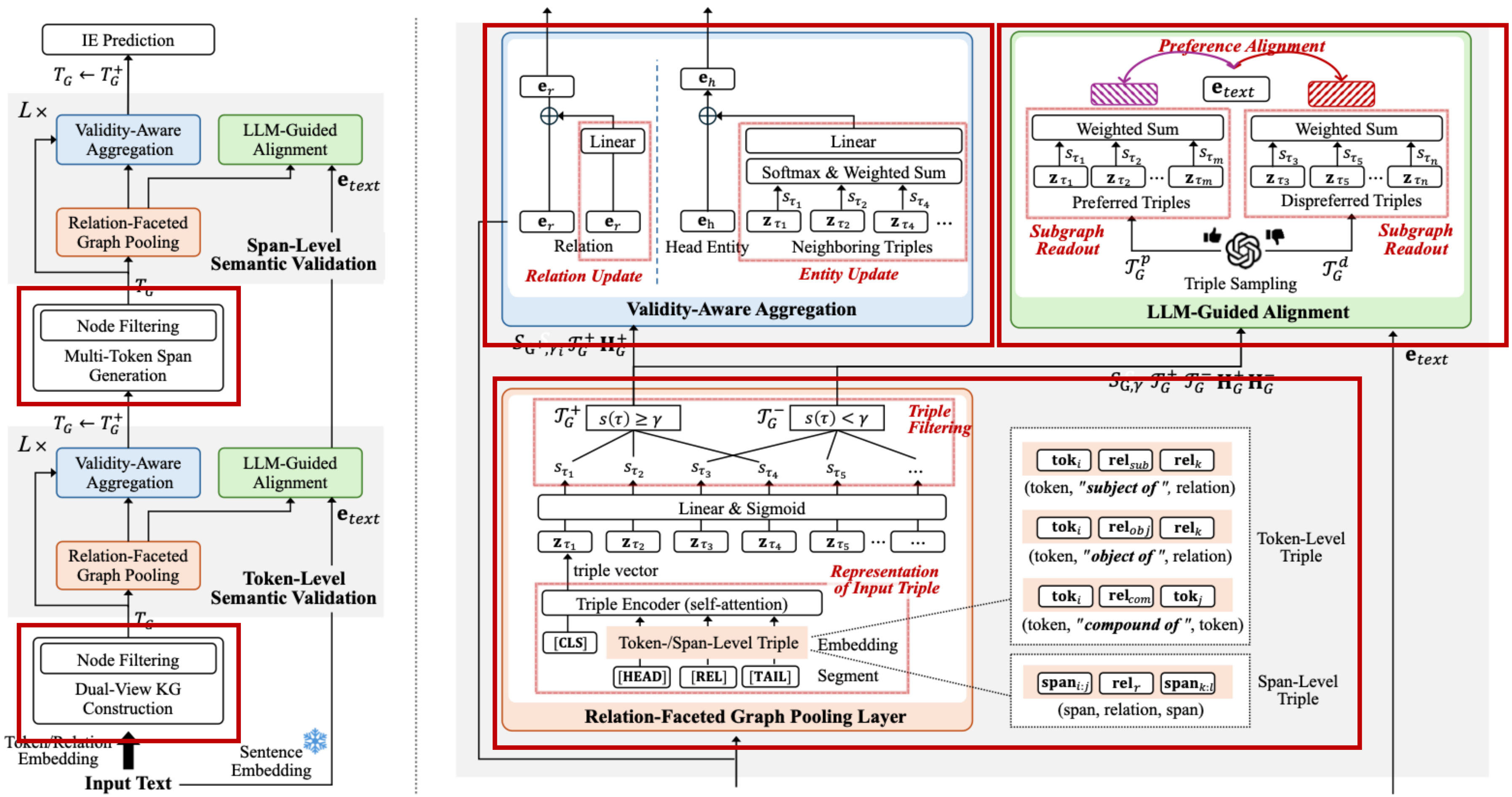
Overview



Proposed Method (RePooL)

Overview

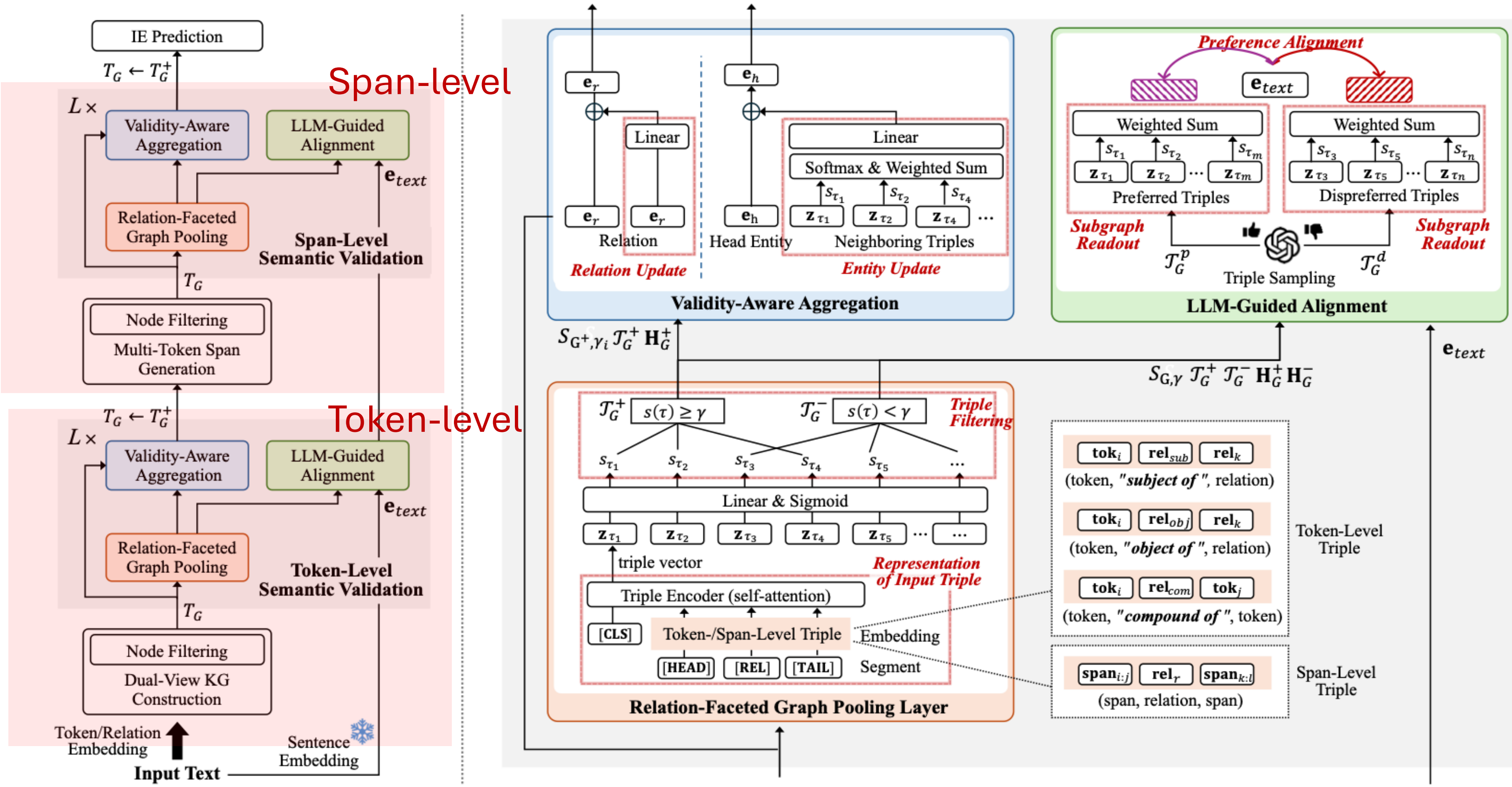
- Consists of five main modules



Proposed Method (RePooL)

Overview

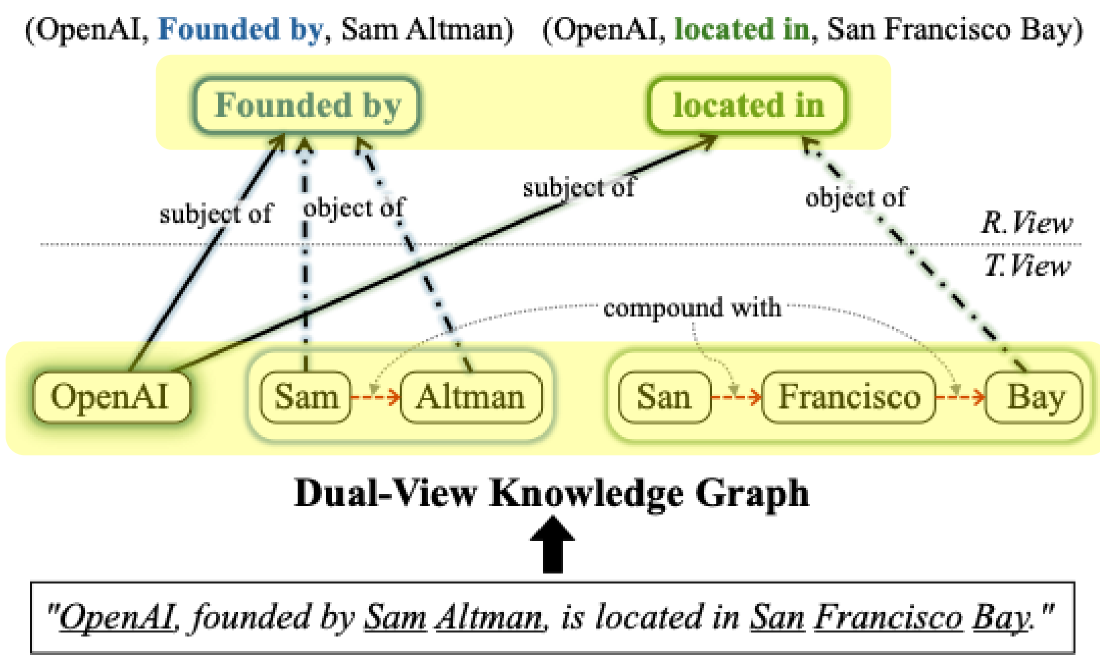
- Hierarchical validation across token and span levels

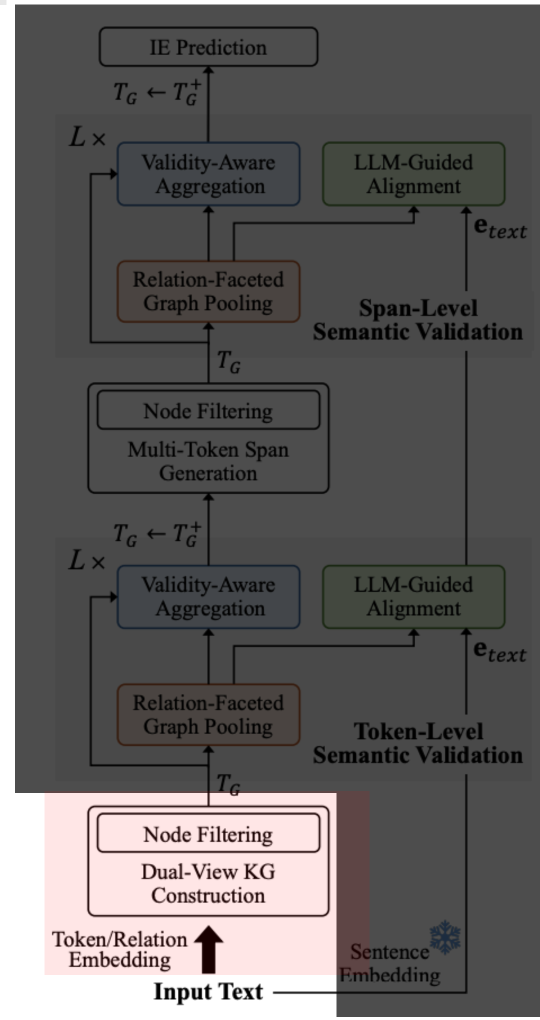


Token-level Stage

Dual-view KG Construction

- Dual-view knowledge graph
  - ✓ Node: token / relation
  - ✓ Edge: auxiliary structural relations

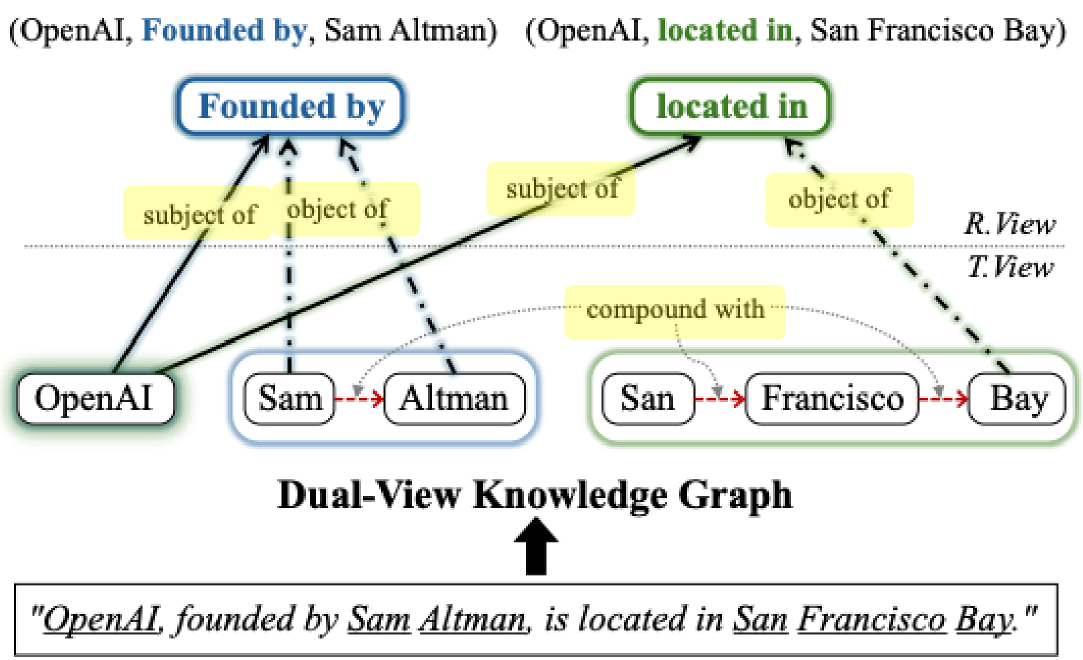




Token-level Stage

Dual-view KG Construction

- Dual-view knowledge graph
  - ✓ Node: token / relation
  - ✓ Edge: auxiliary structural relations (subject of, object of, compound with)
    - subject of/object of: links a token to a relation as a potential subject/object
    - compound with: connects adjacent tokens that may belong to the same entity span, which enables multi-token span formation



Token-level Stage

Dual-view KG Construction

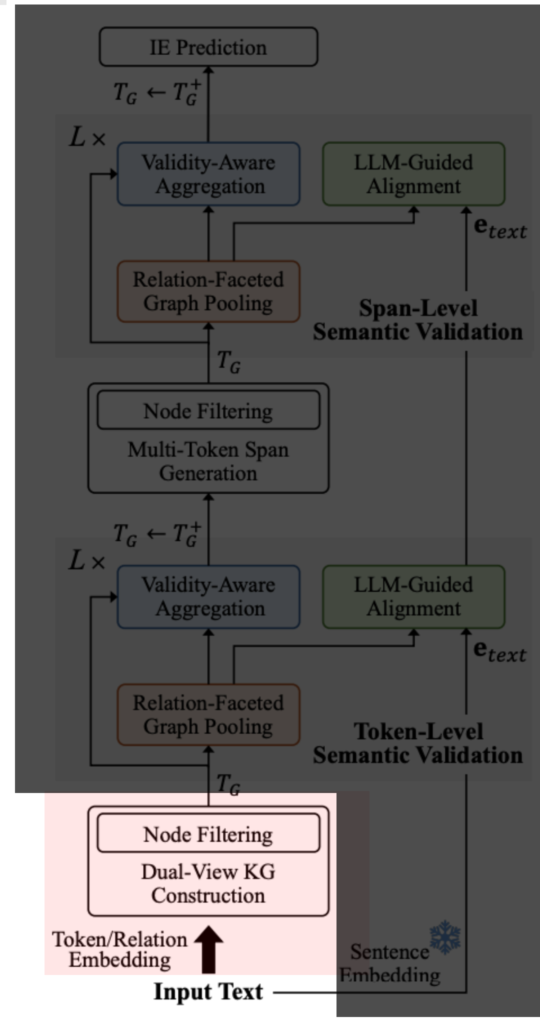
- Token-Level Node Filtering
  - ✓ each token receives relevance score through token specific scoring function:

$$\mathrm{score}_t(tok_i) = \sigma(\mathbf{w}_t^T \mathbf{tok}_i),$$

  - ✓ top-K selection strategy to retain the most relevant tokens

$$V_T \leftarrow \mathrm{TopK}(\{tok_i \mid \mathrm{score}_t(tok_i)\}, K).$$

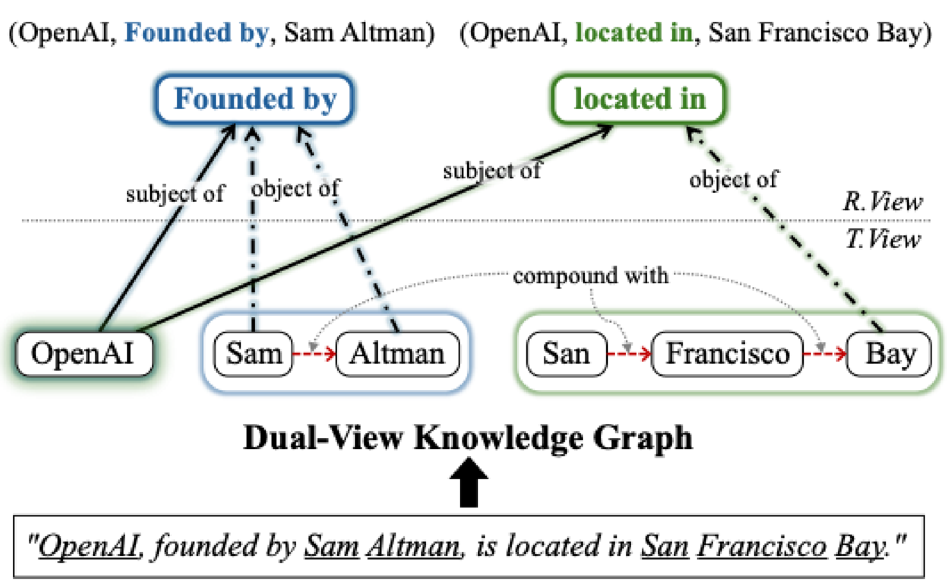
  - ✓ prunes noisy or irrelevant tokens

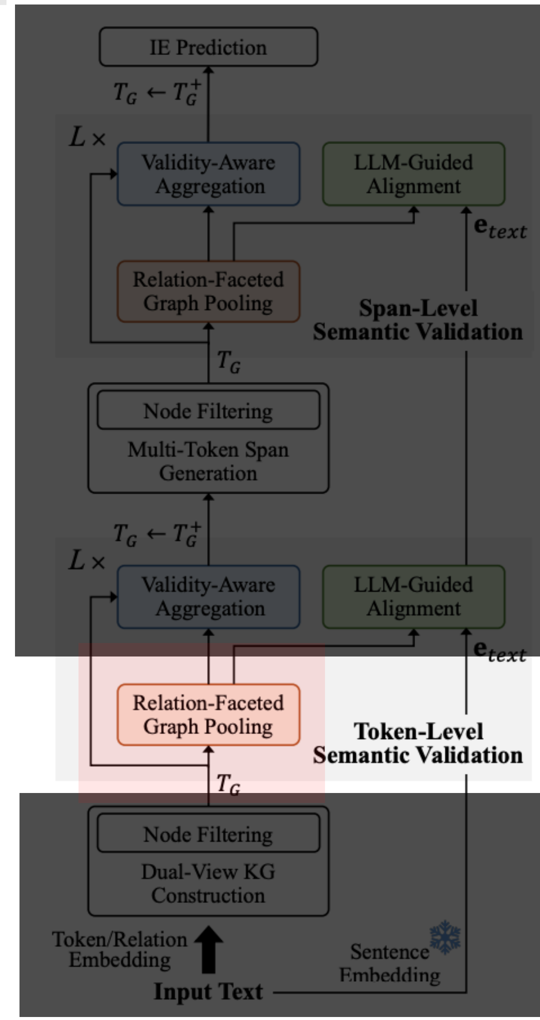


Token-level Stage

Relation-Faceted Graph Pooling

- Triple construction from the dual-view KG
  $(tok_i, rel_{sub/obj/com}, rel_k)$
- Use auxiliary structural relations
  - ✓ (OpenAI, subject_of, Founded_by)
  - ✓ (Sam, compound_with, Altman)

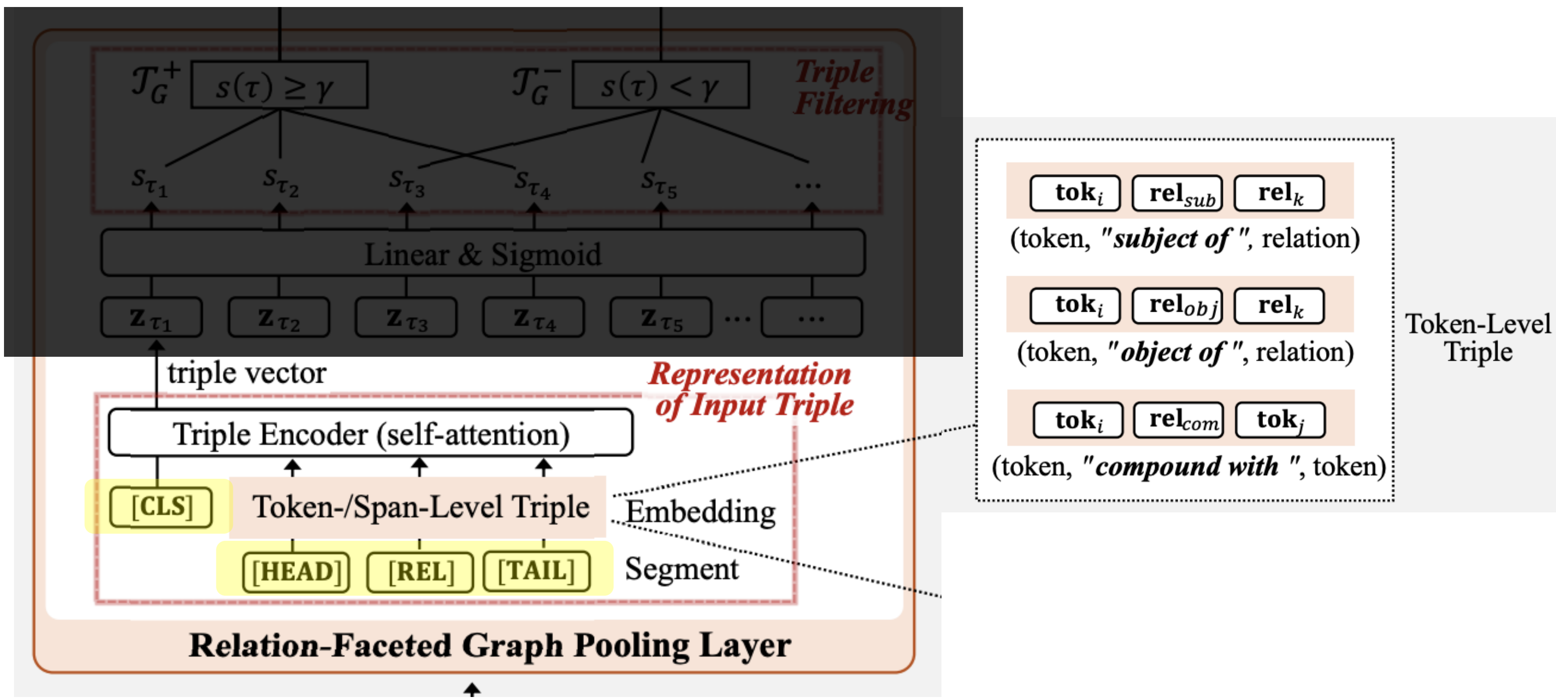




Token-level Stage

Relation-Faceted Graph Pooling

- Triple representation
  - ✓ segmentation token [HEAD], [REL], [TAIL]
  - ✓ [CLS] token summarizes the overall semantics of the triple



Token-level Stage

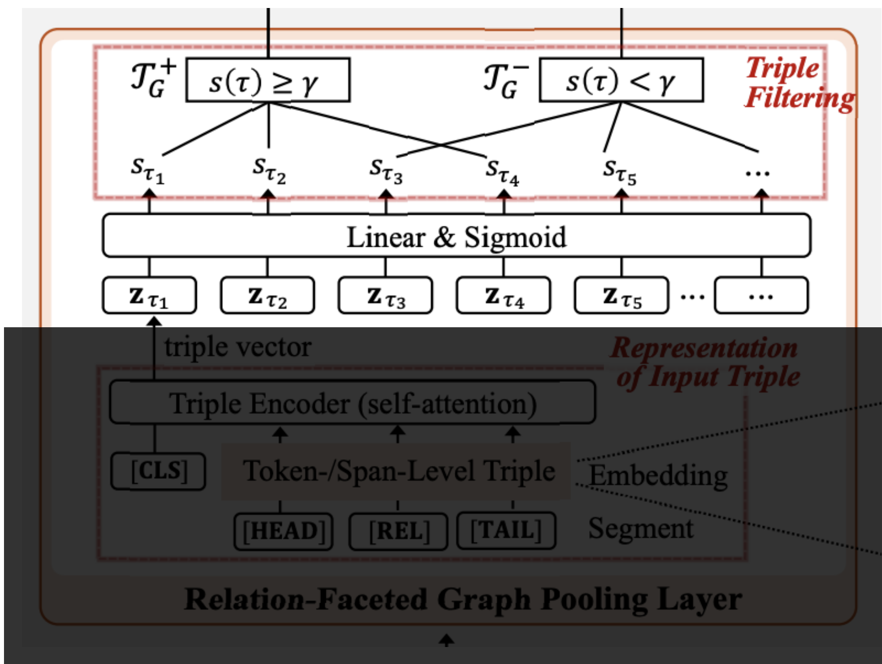
Relation-Faceted Graph Pooling

- Triple filtering
  - ✓ assign validity score for each triple

$$\text{score}_\tau(\tau) = \sigma(\mathbf{w}_\tau^T \mathbf{z}_\tau + b_\tau),$$

  - ✓ apply top-k filtering to retain the most semantically valid triples

$$\mathcal{T}_{TL}^{+} = \text{TopK}(\{\tau \mid \text{score}(\tau), \tau \in \mathcal{T}_{TL}\}, K),$$



Token-level Stage

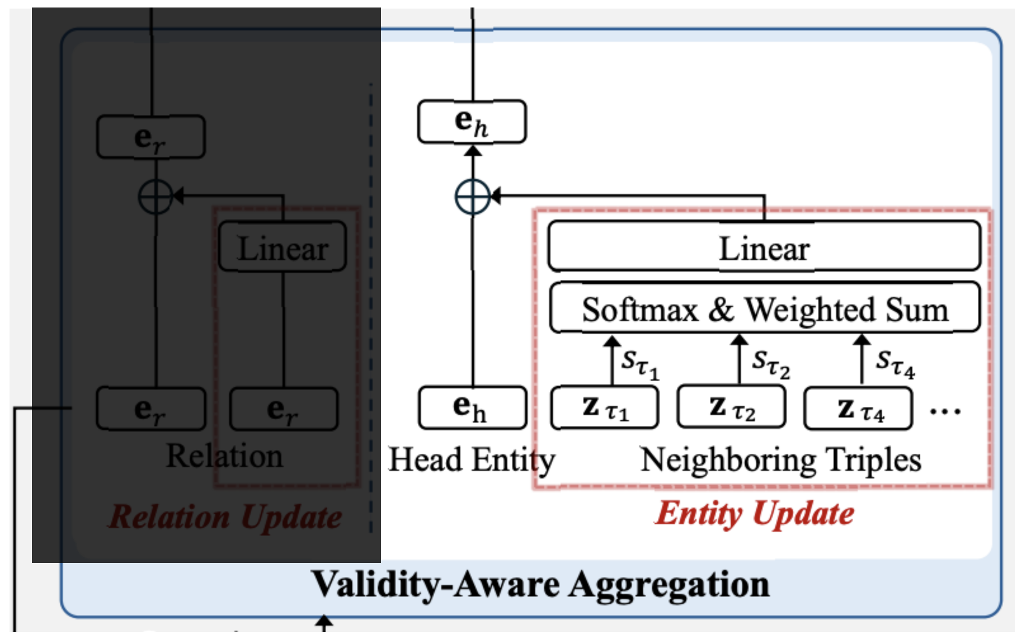
Validity-Aware Aggregation

- Entity Node Update
  - ✓ by aggregating information from connected triples

$$\mathbf{v} \leftarrow \mathbf{v} + \mathrm{Linear}\left(\sum_{\tau \in \mathcal{N}(v)} \mathrm{softmax}(s_\tau) \cdot \mathbf{z}_\tau\right)$$

  - ✓ weighted by their semantic validity scores
  - ✓ aggregates information from three structural triple types of $r_{sub}, r_{obj}, r_{com}$
- Relation Node Update
  - ✓ directly updated through a linear transformation

$$\mathbf{e} \leftarrow \mathbf{e} + \mathrm{Linear}(\mathbf{e}).$$



Token-level Stage

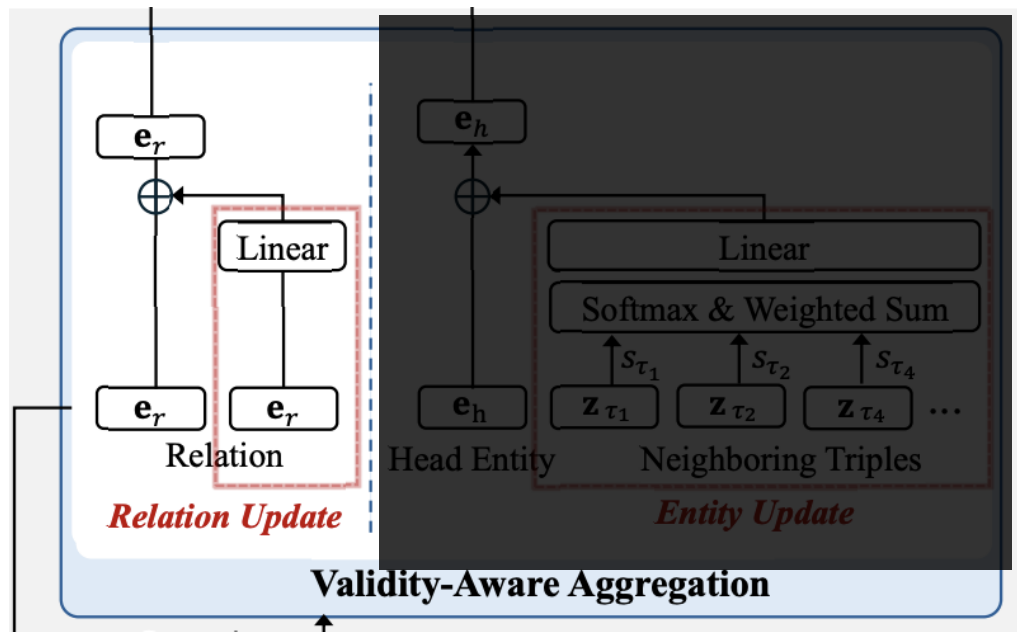
Validity-Aware Aggregation

- Entity Node Update
  - ✓ by aggregating information from connected triples

$$\mathbf{v} \leftarrow \mathbf{v} + \mathrm{Linear}\left(\sum_{\tau \in \mathcal{N}(v)} \mathrm{softmax}(s_\tau) \cdot \mathbf{z}_\tau \right)$$

  - ✓ weighted by their semantic validity scores
  - ✓ aggregates information from three structural triple types of $r_{sub}, r_{obj}, r_{com}$
- Relation Node Update
  - ✓ directly updated through a linear transformation

$$\mathbf{e} \leftarrow \mathbf{e} + \mathrm{Linear}(\mathbf{e}).$$

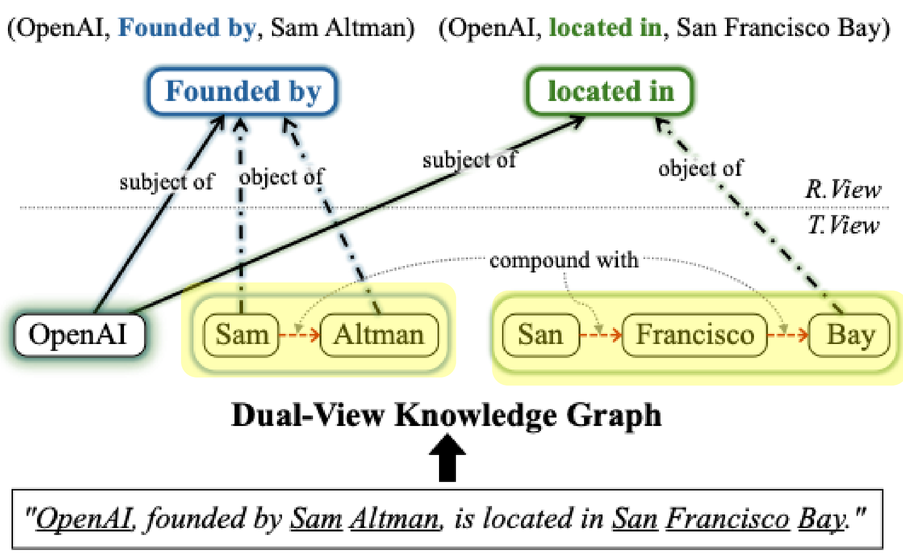


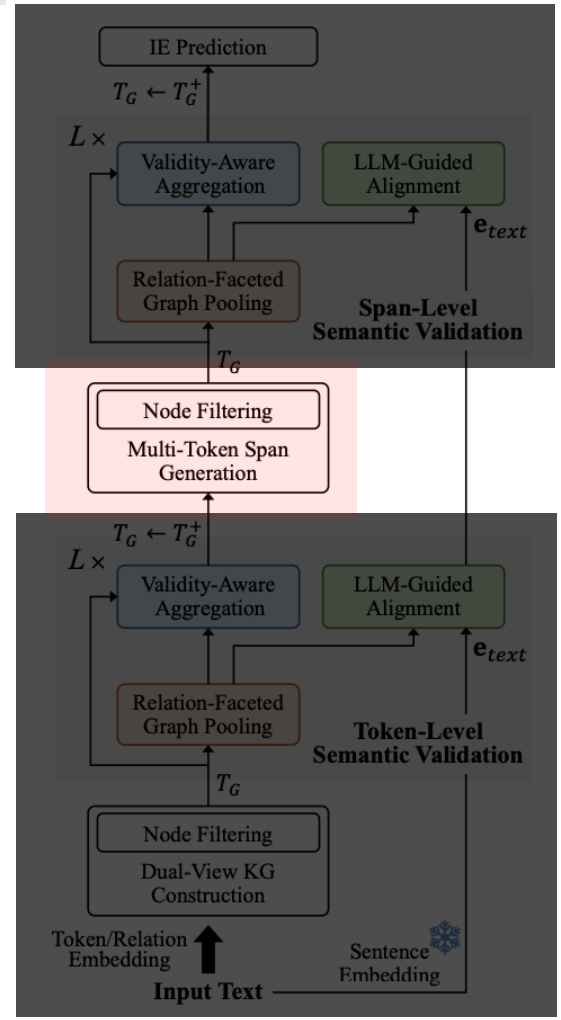
Span-level Stage

Multi-Token Span Generation

- Validated tokens are composed into multi-token spans through compound_with relation r_{com}
- Ensure each one is contiguous and semantically meaningful

$$V_S = \{span_{i:j} \mid \{(tok_i, rel_{com}, tok_{i+1}), (tok_{i+1}, rel_{com}, tok_{i+2}), \ldots, (tok_{j-1}, rel_{com}, tok_j)\} \subseteq \mathcal{T}_{TL},\ i < j\}.$$





Span-level Stage

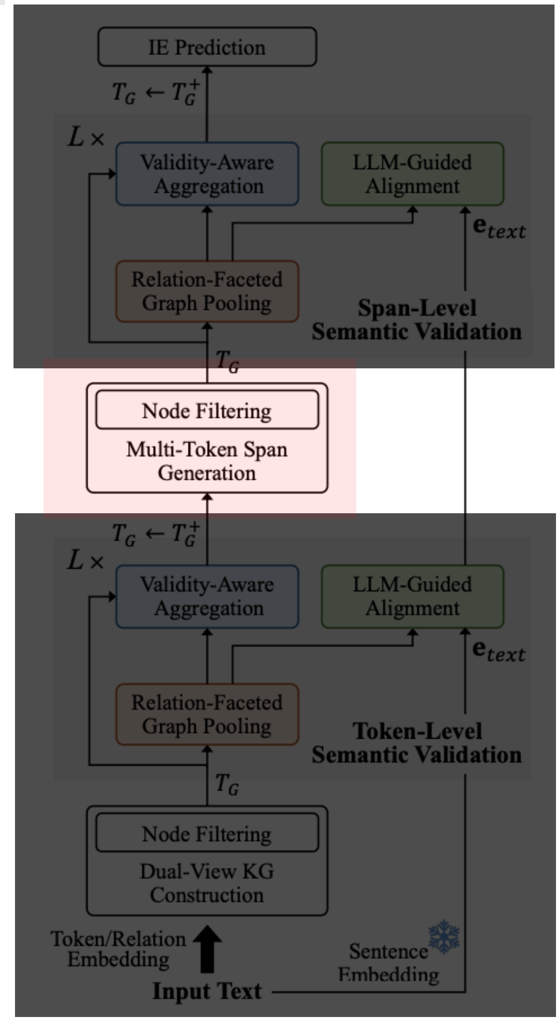
Multi-Token Span Generation

- Validated tokens are composed into multi-token spans through compound_with relation r_{com}
- Ensure each one is contiguous and semantically meaningful

$$V_S = \{span_{i:j} \mid \{(tok_i, rel_{com}, tok_{i+1}), (tok_{i+1}, rel_{com}, tok_{i+2}), \ldots, (tok_{j-1}, rel_{com}, tok_j)\} \subseteq \mathcal{T}_{TL},\ i < j\}.$$

- Each span is represented by pooling its token embeddings

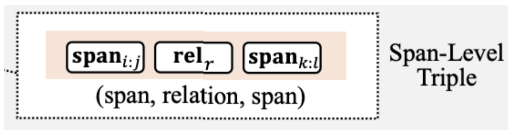
$$\mathbf{span}_{i:j} = \mathrm{Pool}(\{\mathbf{tok}_i, \mathbf{tok}_{i+1}, \ldots, \mathbf{tok}_j\})$$



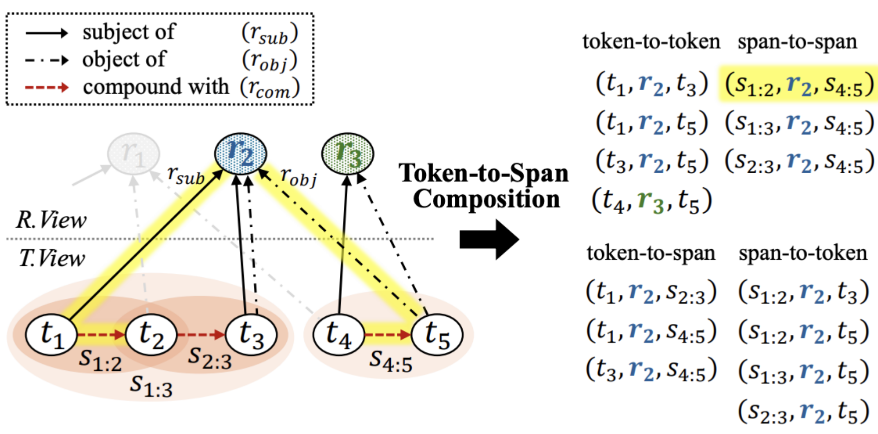
Span-level Stage

Span-level semantic validation

- Relation-Faceted Graph Pooling

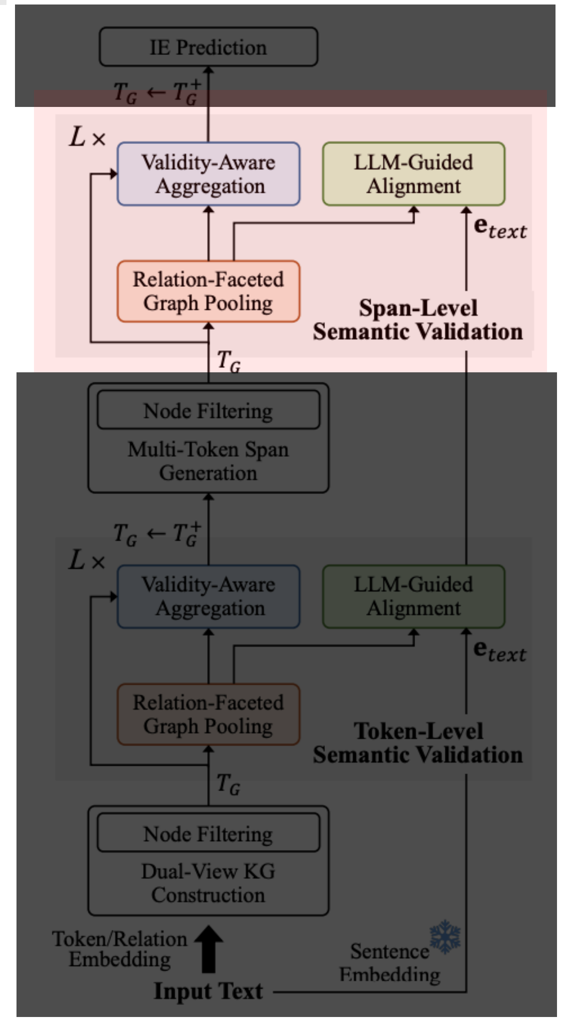


  - ✓ triple composition
  - ✓ threshold-based filtering
- Validity-Aware Aggregation
- LLM-Guided Alignment



Token-Level Triples
(Dual-View KG)

Span-Level Triples

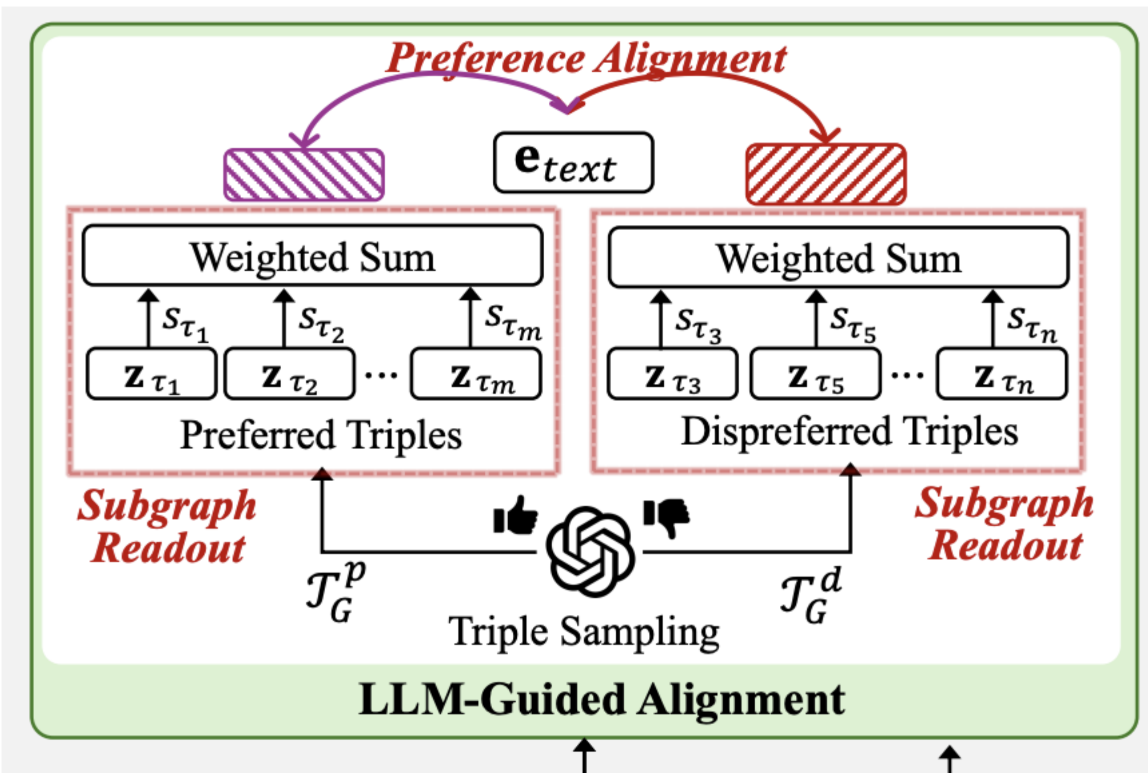


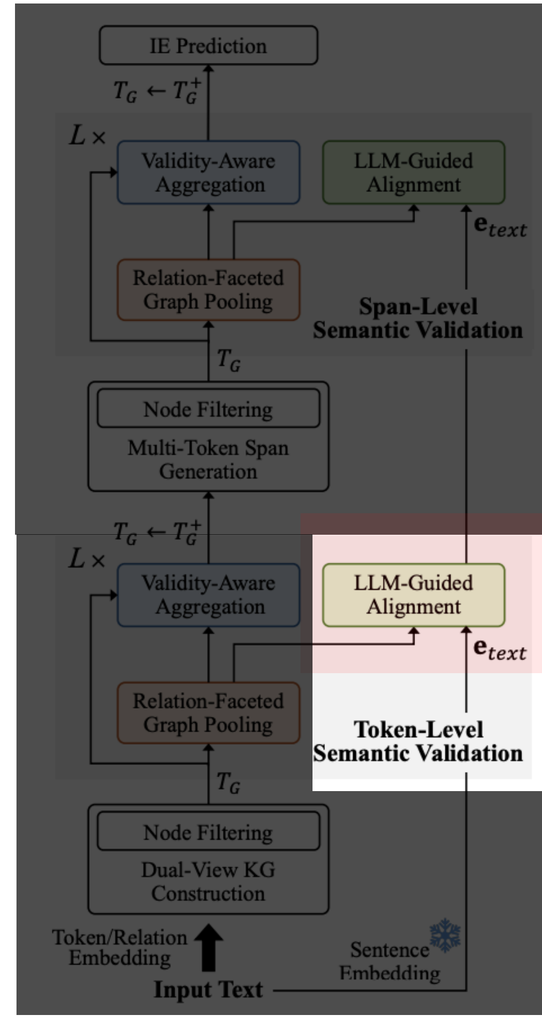
LLM-guided Alignment

- Model-driven validation may lack deep semantic awareness
  - ✓ model can assign unreasonably low scores to valid triples,
  - ✓ filtering them out and creating a kind of negative feedback loop during training

LLM-guided Alignment

- Query LLM to compare candidate triple sets and decide which one better matches the input text

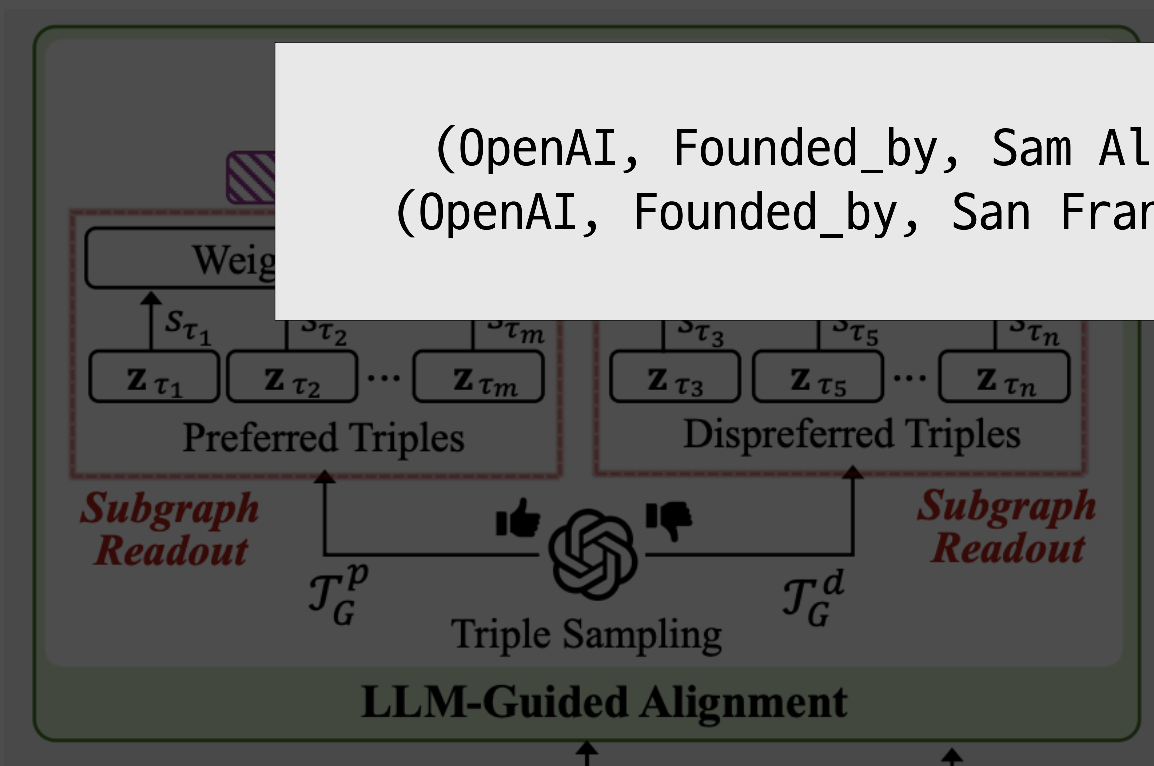


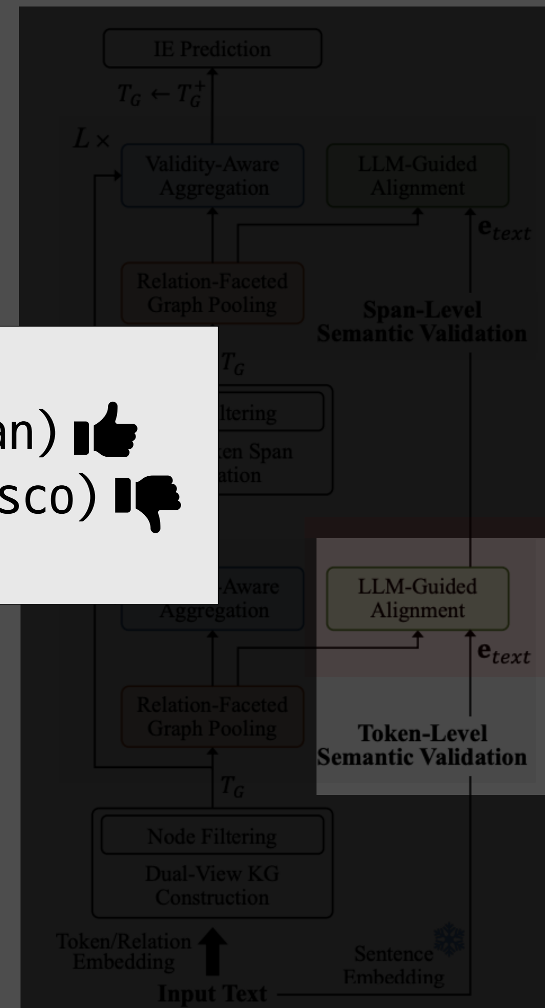


LLM-guided Alignment

- Query LLM to compare candidate triple sets and decide which one better matches the input text

(OpenAI, Founded_by, Sam Altman)
(OpenAI, Founded_by, San Francisco)

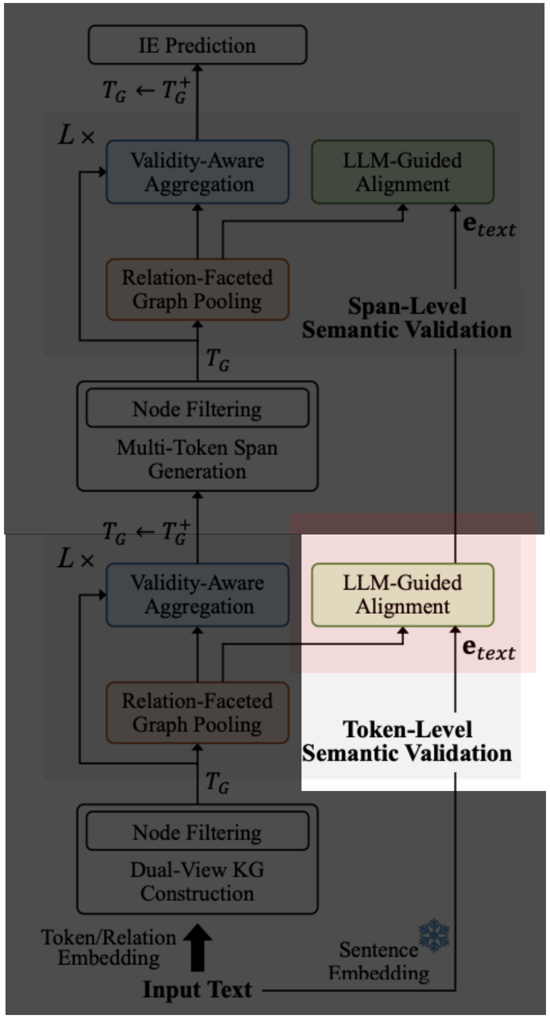




LLM-guided Alignment

- Query LLM to compare candidate triple sets and decide which one better matches the input text
- Apply Bayesian Personalized Ranking(BPR) loss
  - ✓ encourages to prioritize triples that LLM judges as more semantically consistent with the input text

$$\mathcal{L}_{align} = -\log \sigma(\cos(\mathbf{text}, \mathbf{z}_{\mathcal{T}^+}) - \cos(\mathbf{text}, \mathbf{z}_{\mathcal{T}^-}))$$

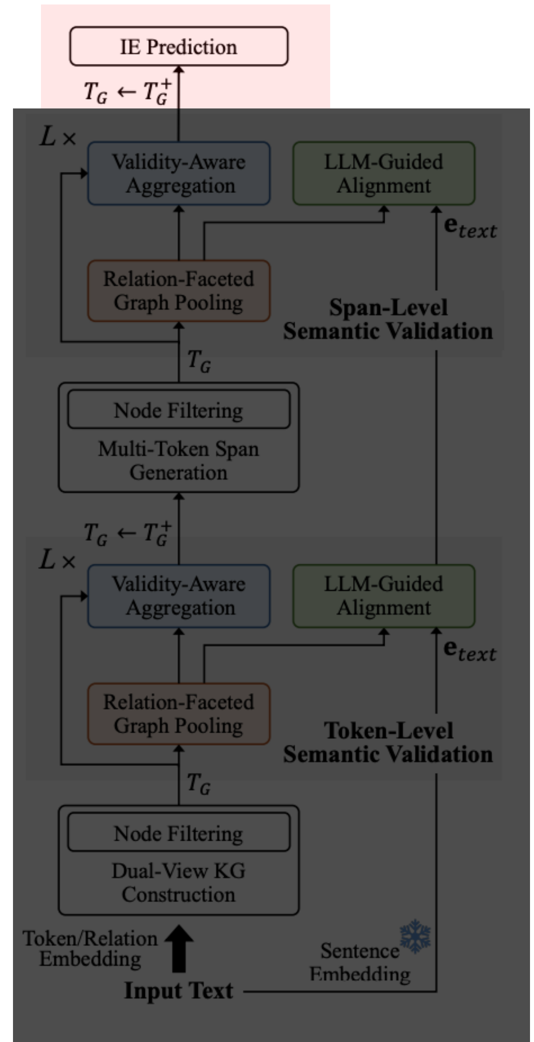


Prediction & Training

- Prediction
  - ✓ rank all candidate triples based on validity scores
  - ✓ top-ranked triples are selected as final outputs

- Training Objective
  - ✓ each corresponds to prefiltering, LLM-guided alignment, and prediction

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prefilter}} + \mathcal{L}_{\text{align}} + \mathcal{L}_{\text{prediction}}$$



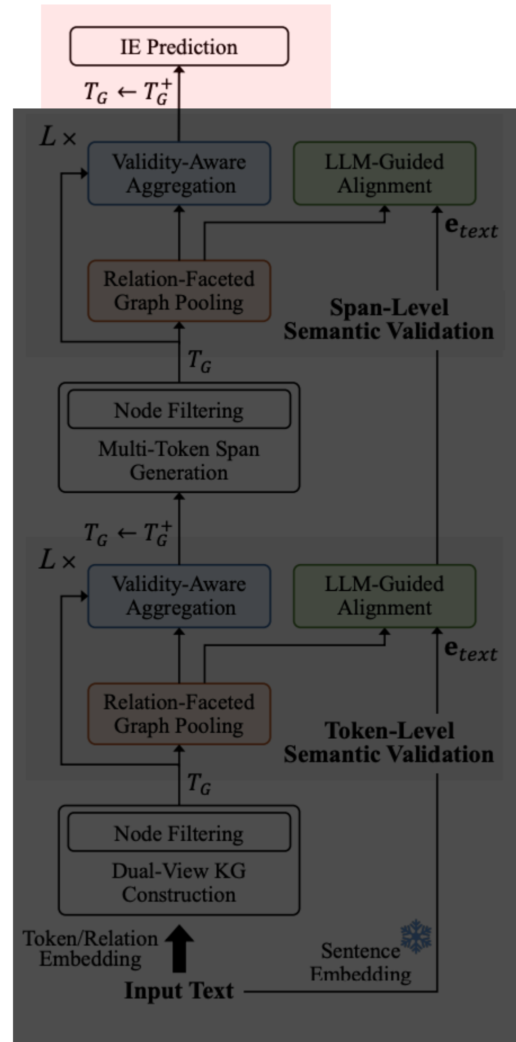
Prediction & Training

- Prediction
  - ✓ rank all candidate triples based on validity scores
  - ✓ top-ranked triples are selected as final outputs

- Training Objective
  - ✓ each corresponds to prefiltering, LLM-guided alignment, and prediction

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prefilter}} + \mathcal{L}_{\text{align}} + \mathcal{L}_{\text{prediction}}$$

- $\mathcal{L}_{prefilter}$: filter out noisy or irrelevant tokens and spans
- $\mathcal{L}_{align}$: transfer LLM's semantic preferences through pairwise supervision
- $\mathcal{L}_{prediction}$: optimize triple-level accuracy for entity and relation extraction



CONTENTS

Experiments

Datasets

- CoNLL04
  : Newswire sentences with general entity&relation types
- SciERC
  : Scientific abstracts from AI conferences, with scientific entities & relations
- ACE05
  : Diverse domains (news, forums, broadcast) with annotated entities & relations

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	# Train	# Dev	# Test
CoNLL04	4	5	922	231	288
SciERC	6	7	1,861	275	551
ACE05	7	6	10,051	2,424	2,050

Table 1: The statistics of datasets.

Experiments

Evaluation Metrics

- ENT
  : correct if span boundaries and entity type match gold
- RE
  : correct if relation type and head/tail entity spans match gold
- RE+
  : correct if relation type, head/tail entity spans and entity types all correct

Experiments

Overall Performance Comparison

- Evaluate on Joint IE task
- RePooL consistently outperforms baselines on CoNLL04, SciERC, and ACE05
- Achieves higher F1 scores for ENT, RE, and RE+

Models	CoNLL04			SciERC			ACE05		
	ENT	RE	RE+	ENT	RE	RE+	ENT	RE	RE+
SpERT [8]	89.25	71.33	71.09	55.26	37.72	36.59	-	-	-
UTC-IE [40]	89.04	65.13	64.53	64.45	40.73	33.73	84.91	63.83	62.57
UniRE [38]	-	-	-	67.36	-	37.23	89.19	-	62.66
PFN [42]	88.79	-	71.32	67.82	-	37.57	89.59	-	67.21
DYGIE++ [36]	-	-	-	67.79	47.37	-	83.52	60.47	-
GraphER [48]	89.07	64.15	63.18	62.06	37.81	35.06	80.48	63.93	61.53
ATG [47]	89.63	70.53	70.53	63.79	40.83	30.89	81.41	64.96	61.79
HGERE [41]	-	-	-	66.61	44.20	30.51	83.72	64.07	60.83
RePooL (our model)	**90.17**	**72.35**	**72.15**	**68.03**	**47.51**	**38.02**	**89.91**	**68.83**	**67.79**

Table 2: Overall performance comparison of joint IE methods, reported in ENT, RE and RE+ F1 scores. Best scores are highlighted in bold and second-best scores are underlined. All experiments are based on results from their GitHub repo or our reproduction, and were reproducibly conducted on the datasets used in the original work, following the original experimental settings.

Experiments

Overall Performance Comparison

- Relation-faceted modeling contributes to extracting more coherent and semantically meaningful triples
  - ✓ provide stronger benefits for relation extraction than for entity extraction
  - ✓ larger gains on datasets with clear and distinct relation types

    -> indicates that explicit relation modeling is especially effective when relations are distinct

Models	CoNLL04			SciERC			ACE05		
	ENT	RE	RE+	ENT	RE	RE+	ENT	RE	RE+
SpERT [8]	89.25	71.33	71.09	55.26	37.72	36.59	-	-	-
UTC-IE [40]	89.04	65.13	64.53	64.45	40.73	33.73	84.91	63.83	62.57
UniRE [38]	-	-	-	67.36	-	37.23	89.19	-	62.66
PFN [42]	88.79	-	71.32	67.82	-	37.57	89.59	-	67.21
DYGIE++ [36]	-	-	-	67.79	47.37	-	83.52	60.47	-
GraphER [48]	89.07	64.15	63.18	62.06	37.81	35.06	80.48	63.93	61.53
ATG [47]	89.63	70.53	70.53	63.79	40.83	30.89	81.41	64.96	61.79
HGERE [41]	-	-	-	66.61	44.20	30.51	83.72	64.07	60.83
RePooL (our model)	**90.17**	**72.35**	**72.15**	**68.03**	**47.51**	**38.02**	**89.91**	**68.83**	**67.79**

Table 2: Overall performance comparison of joint IE methods, reported in ENT, RE and RE+ F1 scores. Best scores are highlighted in bold and second-best scores are underlined. All experiments are based on results from their GitHub repo or our reproduction, and were reproducibly conducted on the datasets used in the original work, following the original experimental settings.

Experiments

Overall Performance Comparison

- Relation-faceted modeling contributes to extracting more coherent and semantically meaningful triples
  - ✓ provide stronger benefits for relation extraction than for entity extraction
  - ✓ larger gains on datasets with clear and distinct relation types

    -> indicates that explicit relation modeling is especially effective when relations are distinct

Models	CoNLL04			SciERC			ACE05		
	ENT	RE	RE+	ENT	RE	RE+	ENT	RE	RE+
SpERT [8]	89.25	<u>71.33</u>	71.09	55.26	37.72	36.59	-	-	-
UTC-IE [40]	89.04	65.13	64.53	64.45	40.73	33.73	84.91	63.83	62.57
UniRE [38]	-	-	-	67.36	-	37.23	89.19	-	62.66
PFN [42]	88.79	-	<u>71.32</u>	<u>67.82</u>	-	<u>37.57</u>	<u>89.59</u>	-	<u>67.21</u>
DYGIE++ [36]	-	-	-	67.79	<u>47.37</u>	-	83.52	60.47	-
GraphER [48]	89.07	64.15	63.18	62.06	37.81	35.06	80.48	63.93	61.53
ATG [47]	<u>89.63</u>	70.53	70.53	63.79	40.83	30.89	81.41	<u>64.96</u>	61.79
HGERE [41]	-	-	-	66.61	44.20	30.51	83.72	64.07	60.83
RePooL (our model)	**90.17**	**72.35**	**72.15**	**68.03**	**47.51**	**38.02**	**89.91**	**68.83**	**67.79**

Table 2: Overall performance comparison of joint IE methods, reported in ENT, RE and RE+ F1 scores. Best scores are highlighted in bold and second-best scores are underlined. All experiments are based on results from their GitHub repo or our reproduction, and were reproducibly conducted on the datasets used in the original work, following the original experimental settings.

CONTENTS

Conclusions

- Key Contributions
  - ✓ proposed RePooL: Relation-Faceted Graph Pooling with LLM-guided Semantic Alignment
  - ✓ improves joint IE through explicit relation modeling and LLM-based semantic supervision

- Results
  - ✓ outperforms competitive baselines on multiple benchmarks

- Future work
  - ✓ extend to open-world or few-shot IE
  - ✓ explore LLM-based knowledge distillation

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

Thank you for your attention!
Any questions?

Hye-Yoon Baek (hannah100@konkuk.ac.kr / hyeyoonbaek@gmail.com)

Konkuk University, Seoul, Republic of Korea

2025.11.13