Neurosymbolic Approaches for Generalizable, Explainable, and Scalable Knowledge Graph Integration: How can LLMs help?

Ensiyeh Raoufi
Polytech, University of Montpellier
IRD (Institut de recherche pour le développement)

20.04.2026

A research supervised by

Konstantin Todorov

Pierre Larmande

François Scharffe

Contents

Introduction

Terminology, Methods & Challenges

The Gaps

Analysis of Entity Alignment Datasets & Methods

Reasoning EA

CENLIEA: A Cross-Encoder NLI Framework Enhanced by LLM for Entity Alignment

Conclusion & Perspectives

Results & Vision

Introduction

Terminology, Methods & Challenges

The Gaps

Analysis of Entity Alignment Methods & Datasets

Reasoning EA

CENLIEA: A Cross-Encoder NLI Framework Enhanced by LLM for Entity Alignment

Conclusion & Perspectives

Publications & Future works

Knowledge Graph (KG)

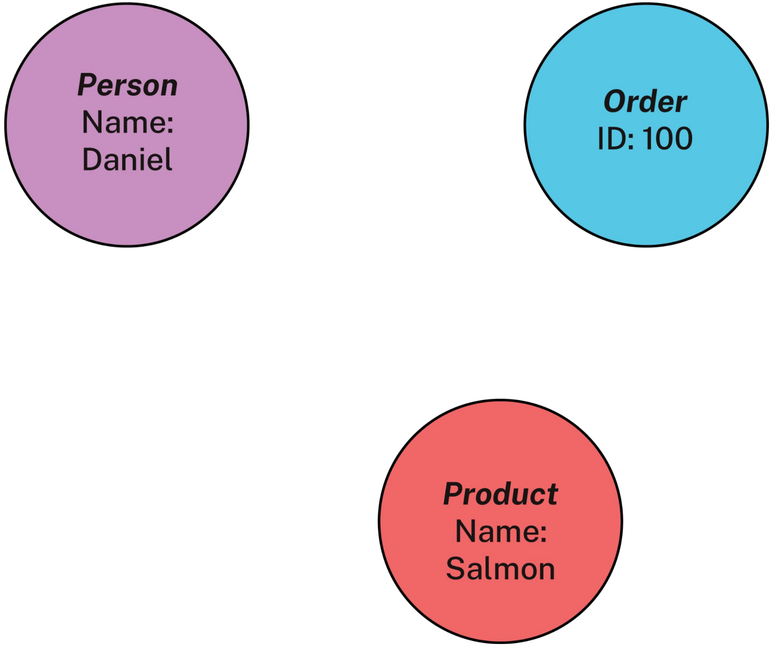
Idea: Facts represented as triples **(subject, predicate, object)**

Knowledge Graph (KG)

Idea: Facts represented as triples **(subject, predicate, object)**

Components:

- **Nodes** → denote entities (e.g., Person, Order)



From: https://neo4j.com/blog/knowledge-graph/what-is-knowledge-graph/

Knowledge Graph (KG)

Idea: Facts represented as triples **(subject, predicate, object)**

Components:

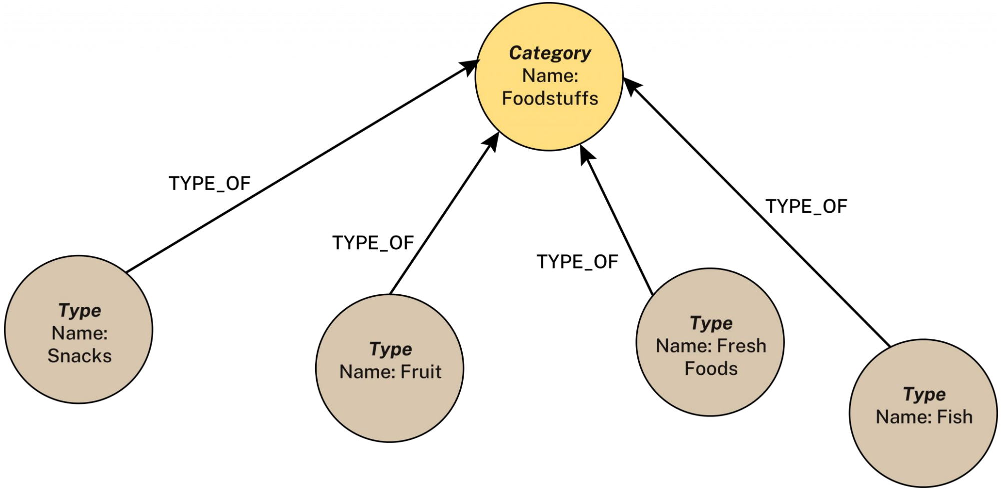
- **Nodes** → entities (e.g., Person, MusicalWork)
- **Edges (Relations)** → semantic links (e.g., "Placed_Order")

From: https://neo4j.com/blog/knowledge-graph/what-is-knowledge-graph/

Knowledge Graph (KG)

Components:

- **Nodes** → entities (e.g., Person, MusicalWork)
- **Edges (Relations)** → semantic links (e.g., "Composed")
- **Ontology / Schema** → types and taxonomies



From: https://neo4j.com/blog/knowledge-graph/what-is-knowledge-graph/

Knowledge Graph (KG)

Idea: Facts represented as triples **(subject, predicate, object)**

Components:

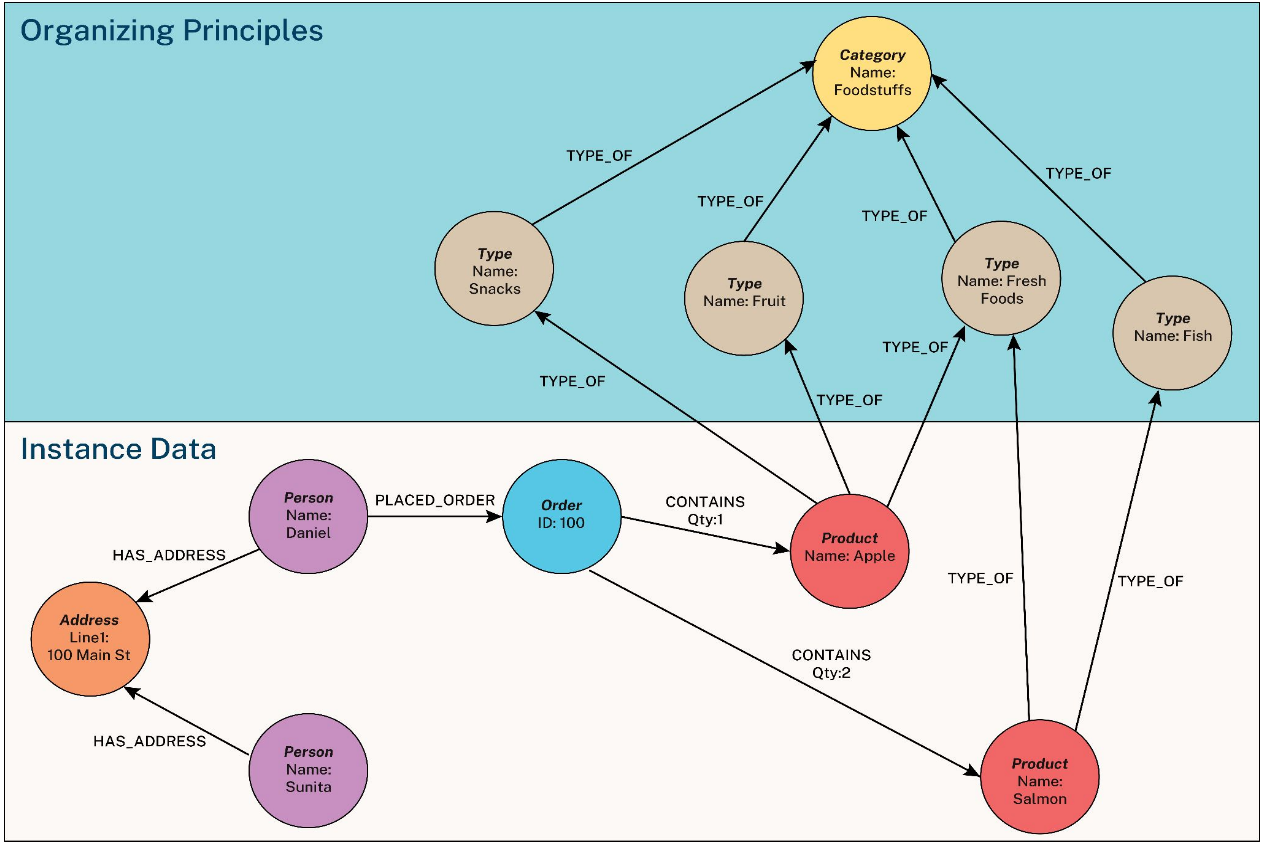
- Nodes → (e.g., Person, Order)
- Edges (Relations) → (e.g., "Placed_Order")
- Ontology / Schema

Example:
(Daniel, Placed_Order, Order_ID:100)

Reasoning over KGs

KGs enable **reasoning**.

For example:
Which persons have purchased fish?



From: https://neo4j.com/blog/knowledge-graph/what-is-knowledge-graph/

Knowledge Graph Integration

What is KG integration?

- Combine and merge KGs to have a unified KG

Knowledge Graph Integration

What is KG integration?

- Combine and merge KGs to have a unified KG

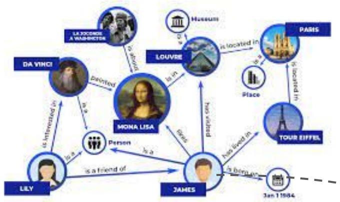
Why KG integration?

- Different KGs describe overlapping entities
- Richer semantic search and reasoning

Entity Alignment (EA)

Entity Alignment (EA) across Knowledge Graphs (KG)

- Large ever-growing KGs:



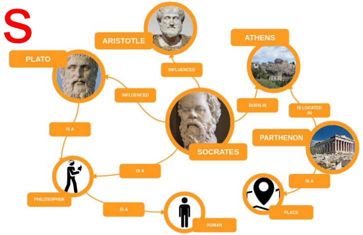
2B triples

yago

owl:sameAs

850M triples

DBpedia



source KG

Example:

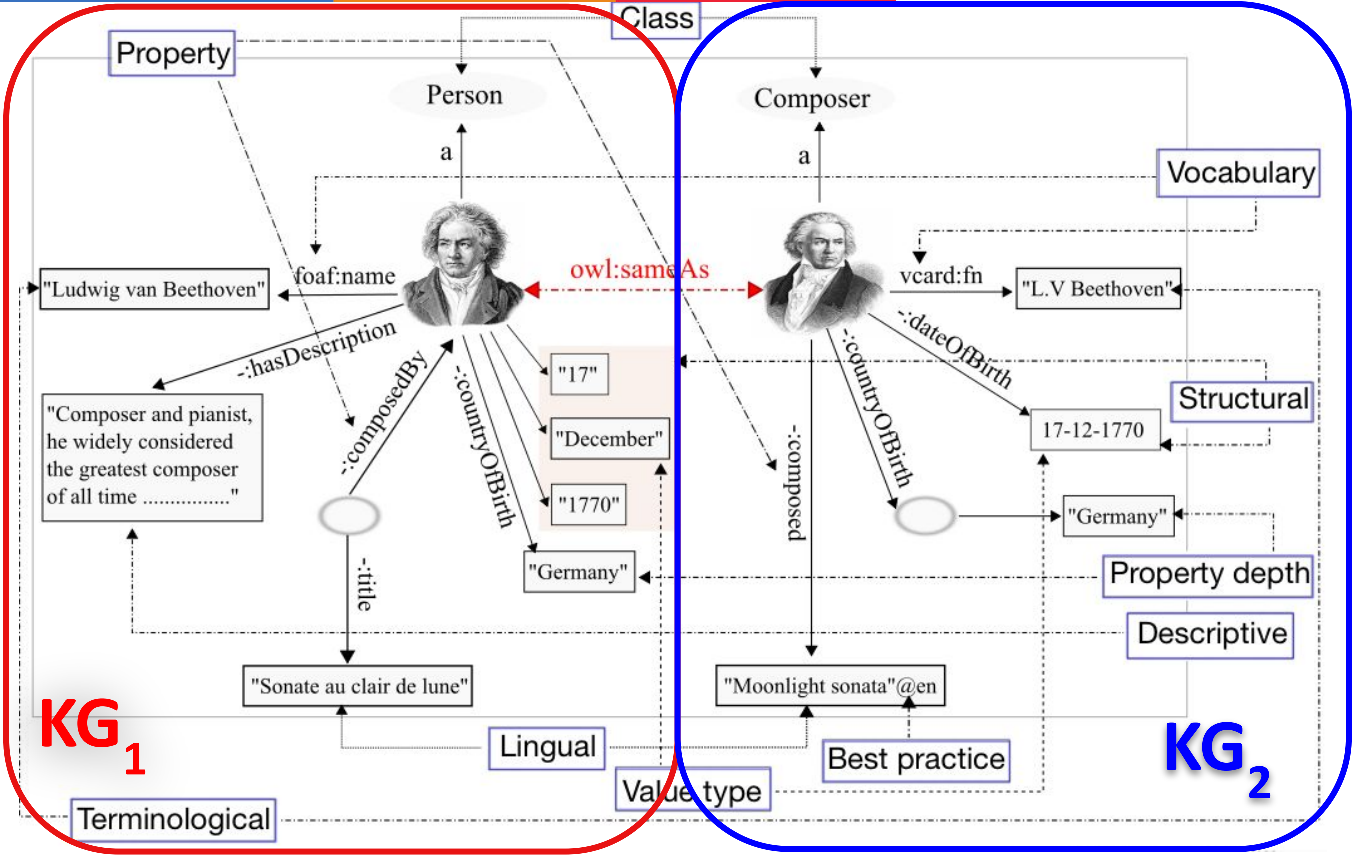
`http://yago-knowledge.org/resource/Ludwig_van_Beethoven,`
owl:sameAs, `http://dbpedia.org/resource/Ludwig_van_Beethoven`

target KG

Entity Alignment

Entity Alignment:

Identifying equivalent entities across KGs.

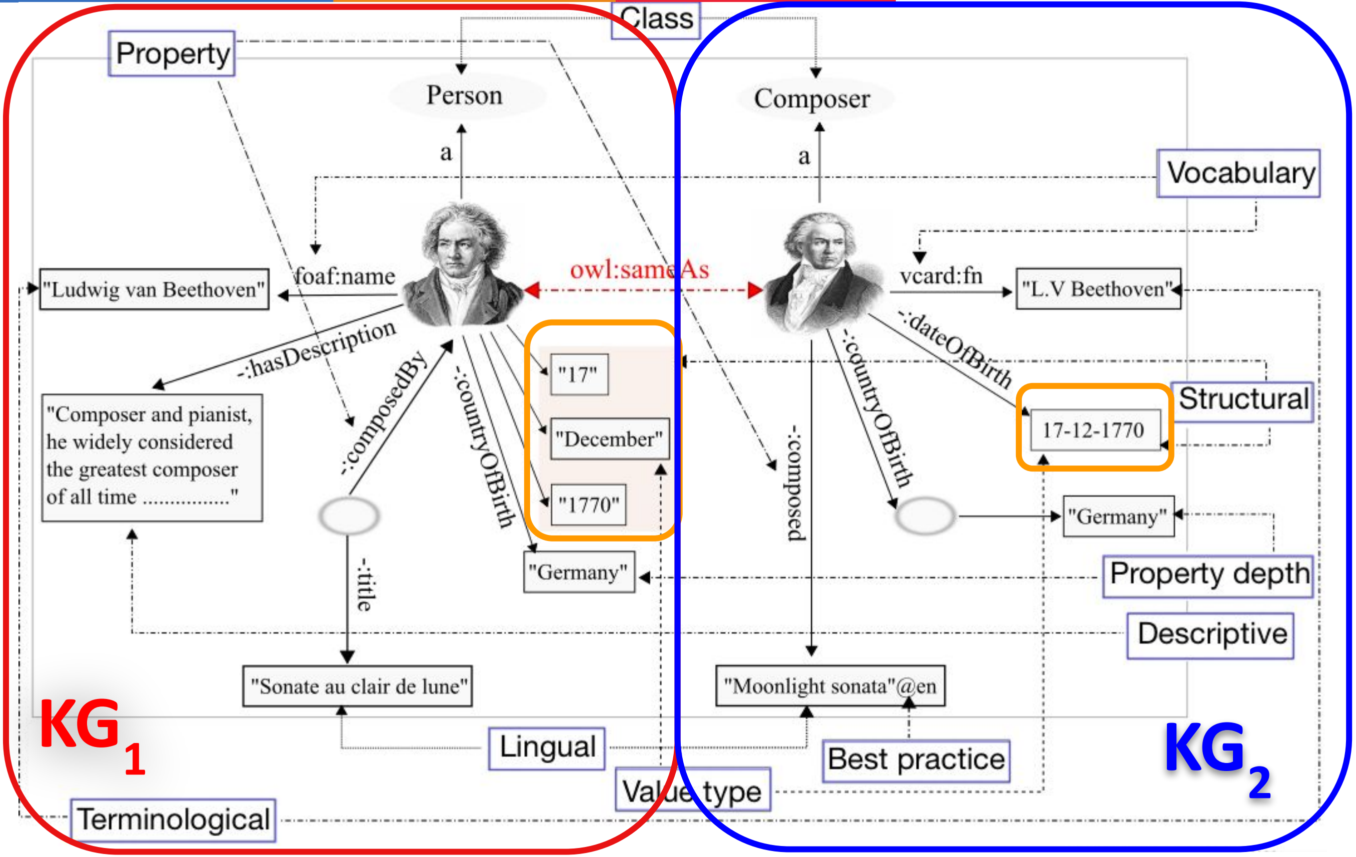


From [1], by Achichi et al.

Entity Alignment

Not an easy task…

- Structural difference (heterogeneity)

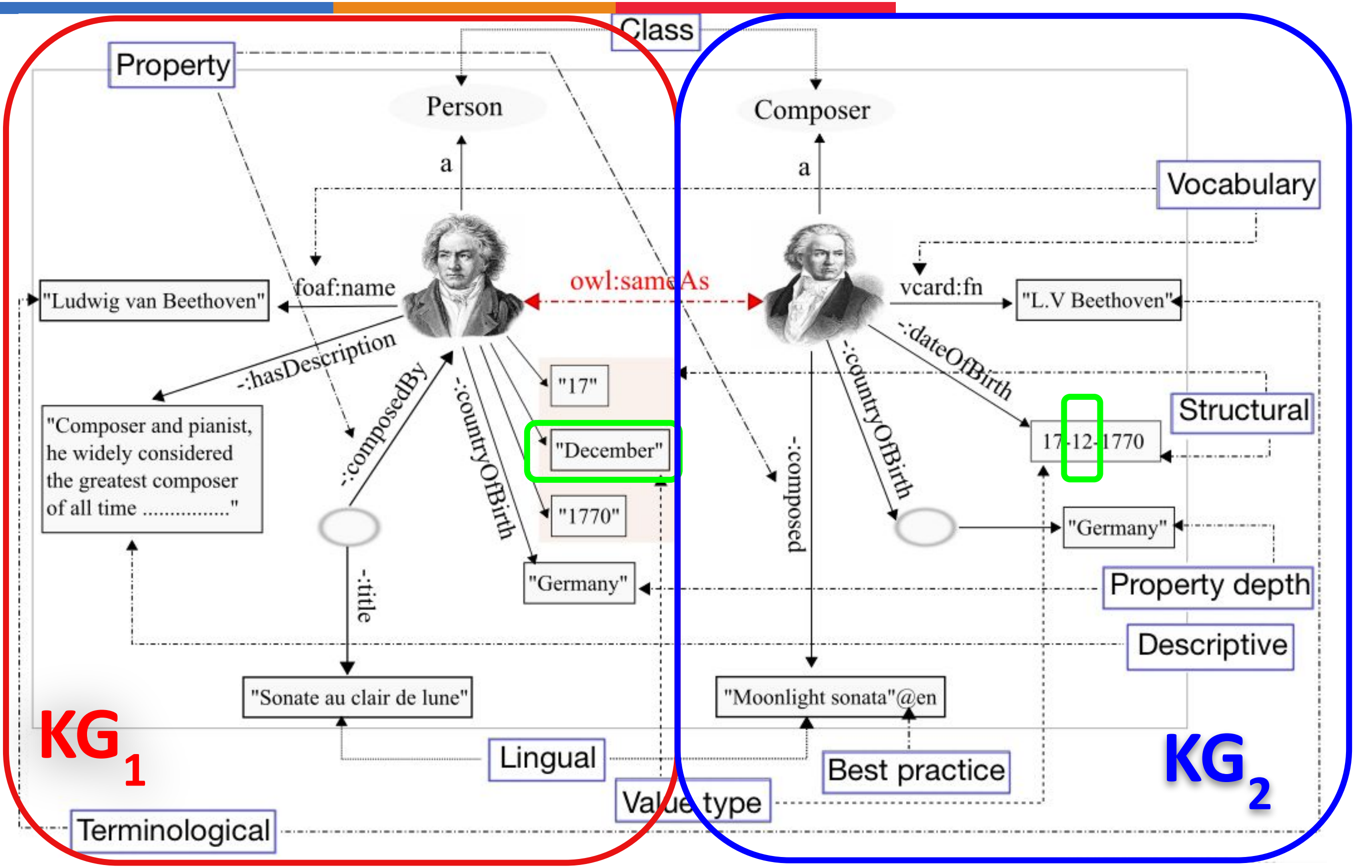


From [1], by Achichi et al.

Entity Alignment

Not an easy task…

- Syntactical heterogeneity

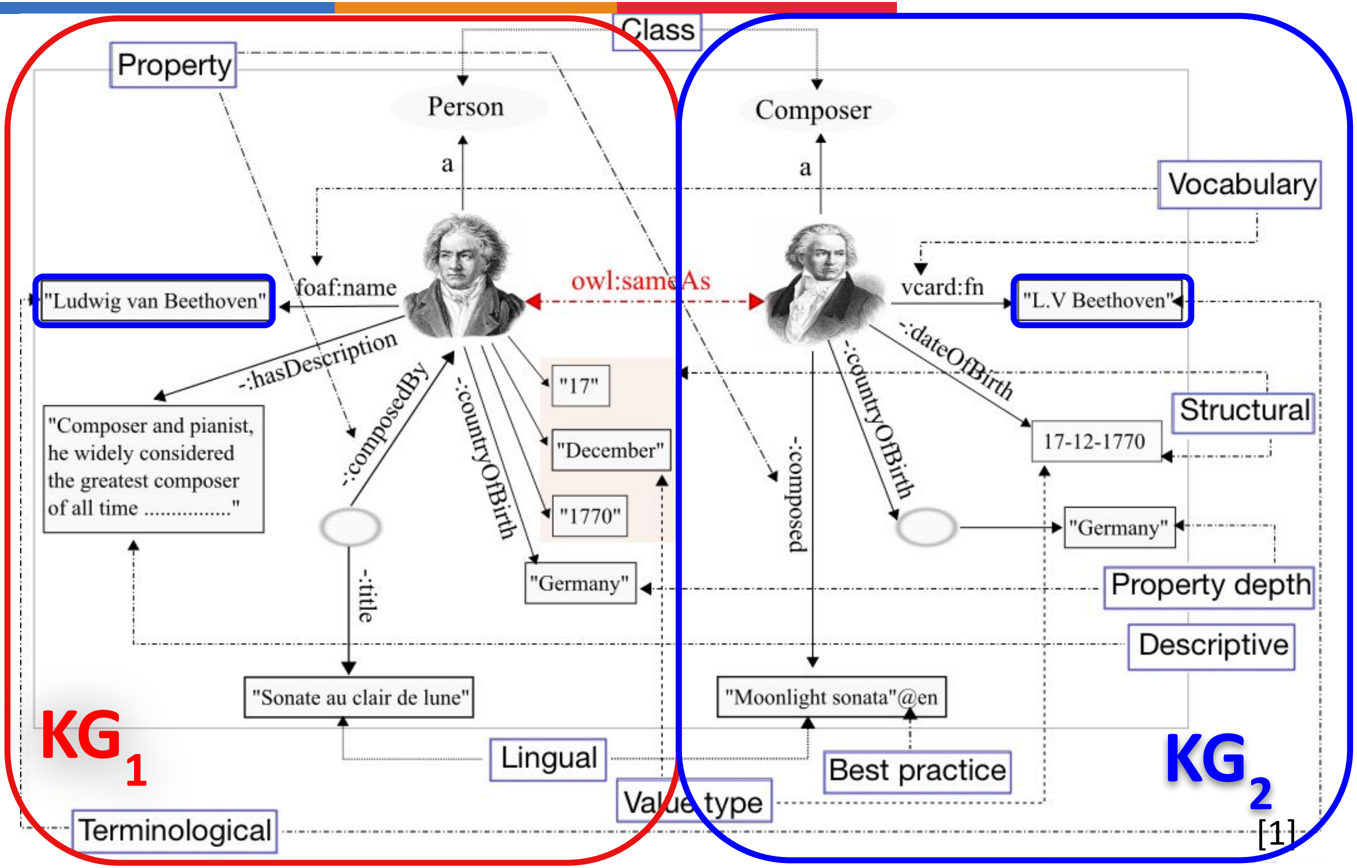


From [1], by Achichi et al.

Entity Alignment

Not an easy task…

- Terminological heterogeneity.



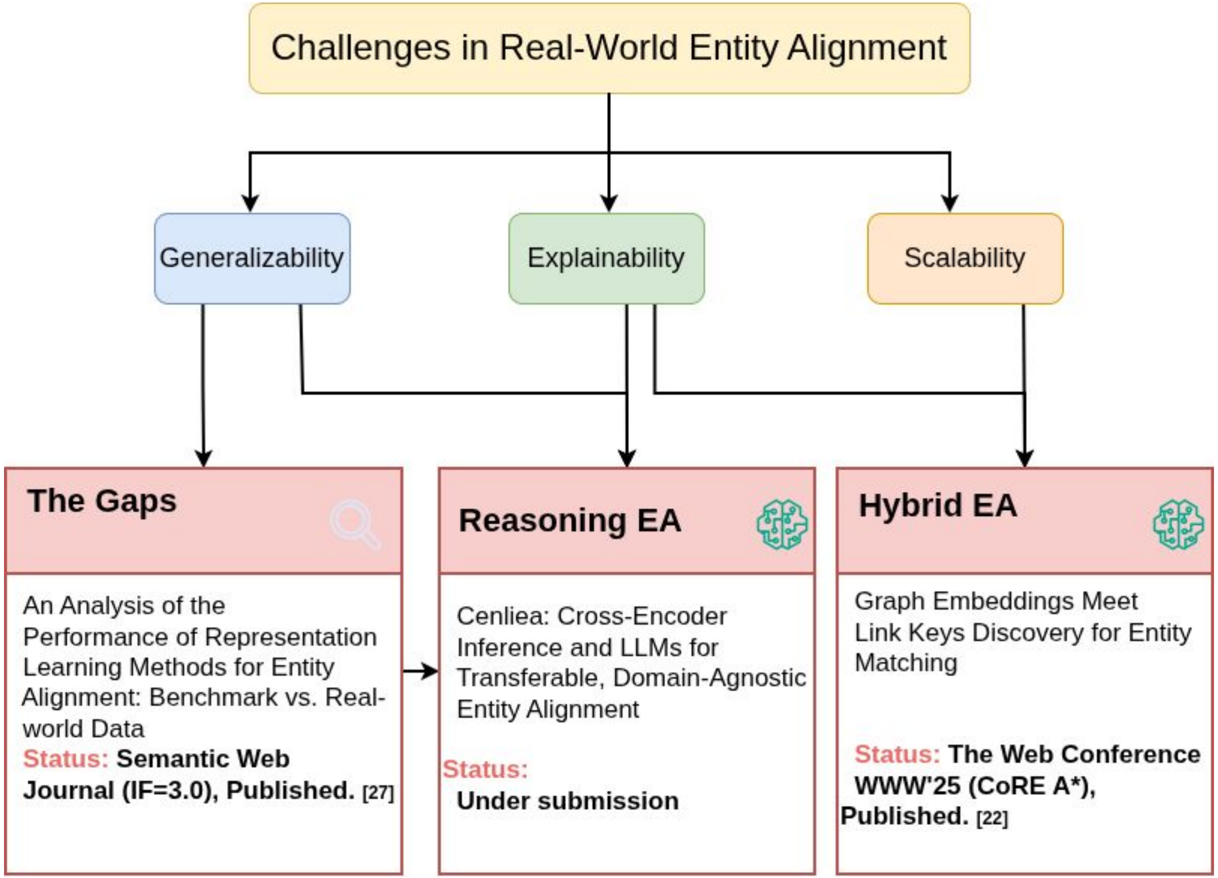
From [1], by Achichi et al.

EA: Some terminology

Dataset: a pair of a source and a target KG to be interlinked, together with a reference alignment.

Reference (or seed) alignment: a manually curated set of correspondences across the two KGs.

EA: Challenges & Contributions



Introduction

KG Integration & Entity Alignment, Methods & Challenges

The Gaps

Analysis of EA Datasets & Methods

Reasoning EA

CENLIEA: A Cross-Encoder NLI Framework Enhanced by LLM for Entity Alignment

Conclusion & Perspectives

Publications & Future works

EA datasets: benchmark vs. real-world

Datasets

1) **Benchmark Data**

SPIMBENCH[1]: Creative Works / General Web Media

DBP15K (Fr-En, Ja-En, Zh-En) [26]: General-purpose / Multilingual Encyclopedia

2) **Real-World Data**

DoReMus[2]: Classical Music / Cultural Heritage

AgroLD[2]: Plant Science / Bioinformatics

1. https://oaei.ontologymatching.org/2018/spimbench.html
2. https://github.com/DACE-DL/Create_Input_Data_to_EA_Models

Dataset Characteristics

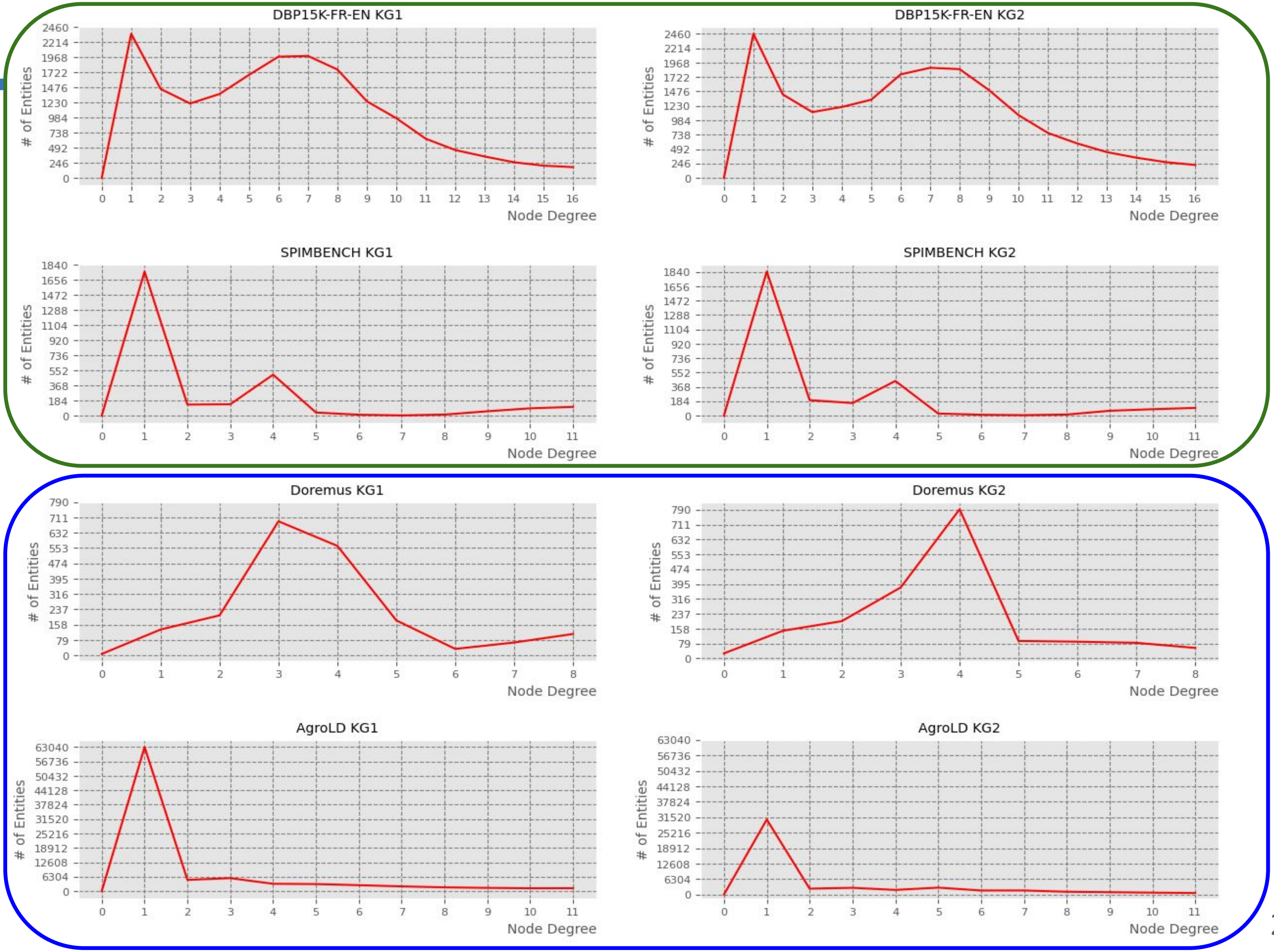
How similar are the two KGs in the datasets?

Benchmark

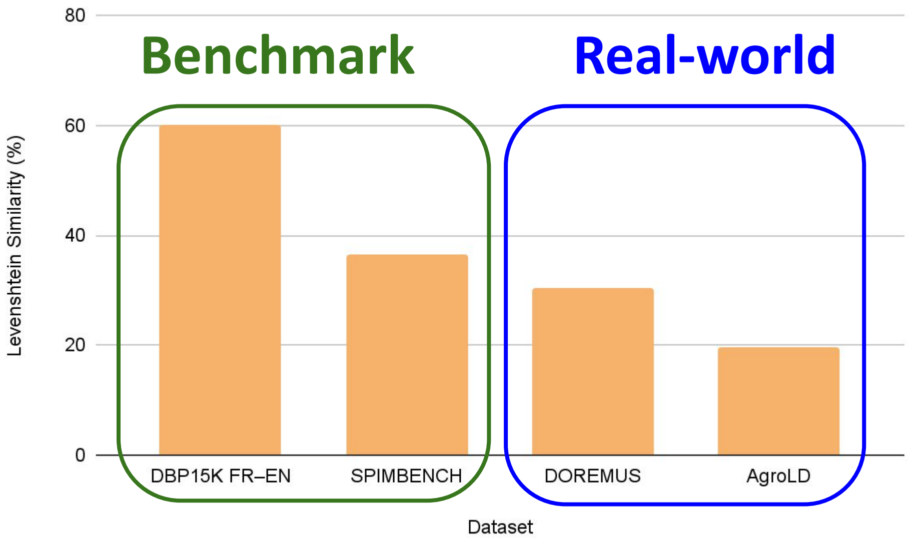
Structural Similarity:

Degree Distributions

Real-world



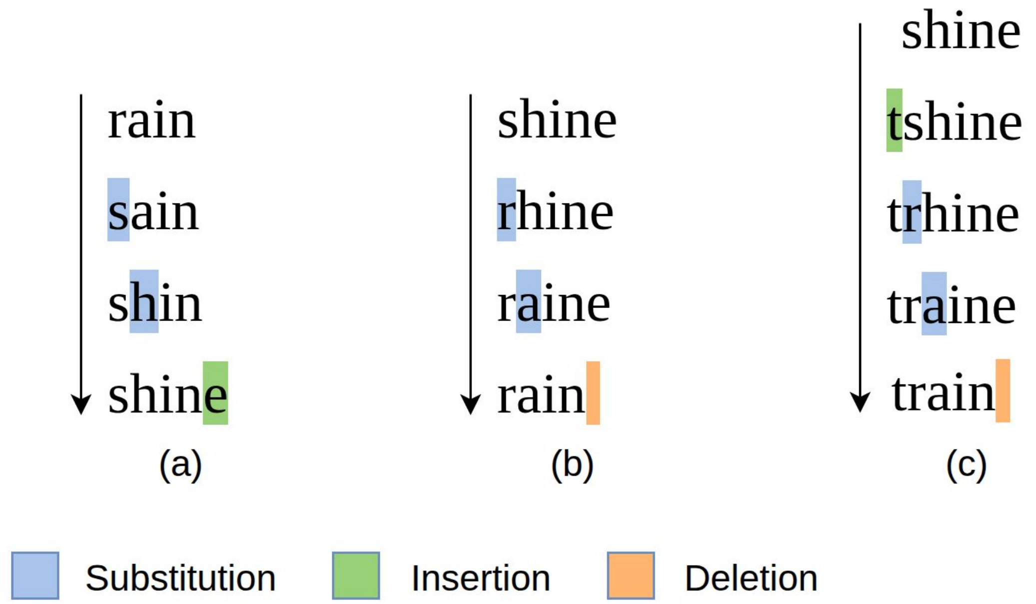
Dataset Characteristics: Textual Similarity



Normalized Levenshtein Similarity

SPIMBENCH from: https://oaei.ontologymatching.org/2018/spimbench.html
AgroLD & DoReMus from: https://github.com/DACE-DL/Create_Input_Data_to_EA_Models
DBP15K (Fr-En, Ja-En, Zh-En) from [26].

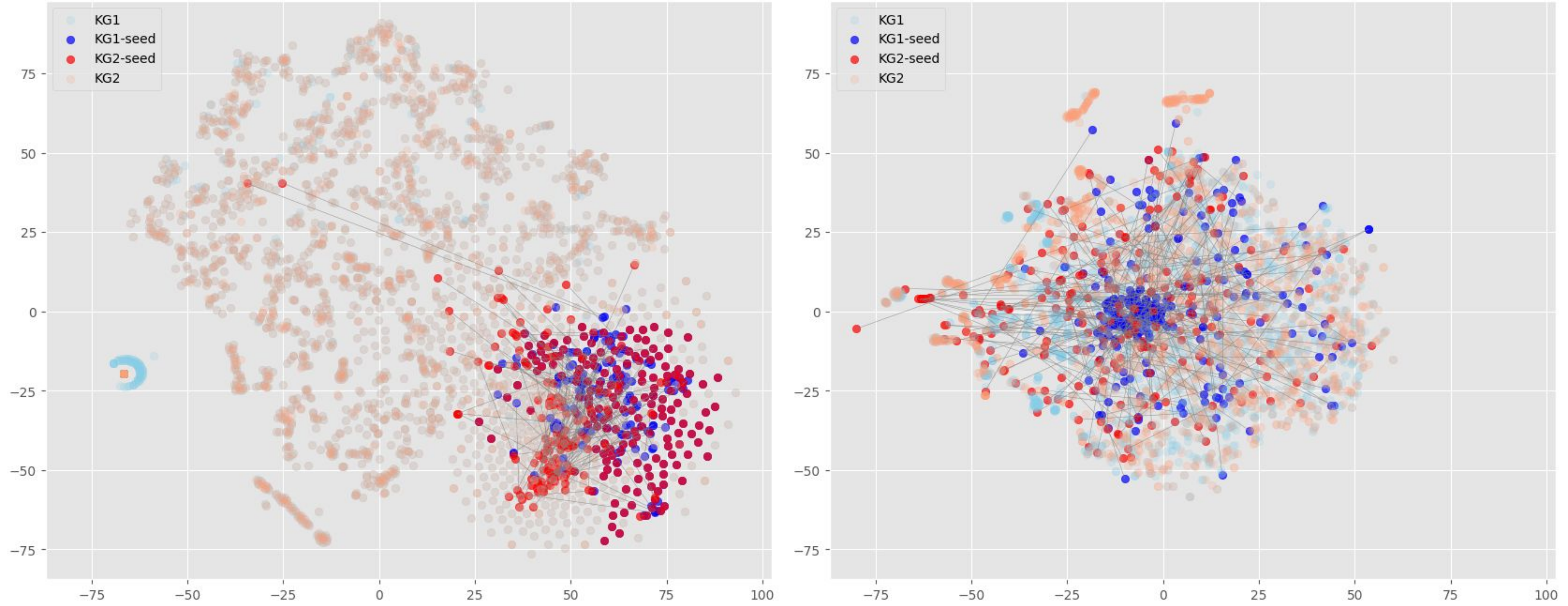
Textual Similarity: Levenshtein Distance



Levenshtein Distance. Minimum number of single-character edits.

Photo from: https://devopedia.org/levenshtein-distance

Dataset Characteristics: Embedding Spaces



SPIMBENCH (left) vs **DOREMUS (right):** Reduced-dimension BERT-based initial entity embeddings. **Seed pairs** connected by **gray** lines.

Dataset Characteristics: Result

- **Benchmarks:** semantically similar, structurally matched
- **Real-world:** structurally and semantically heterogeneous

State-of-the-Art EA methods

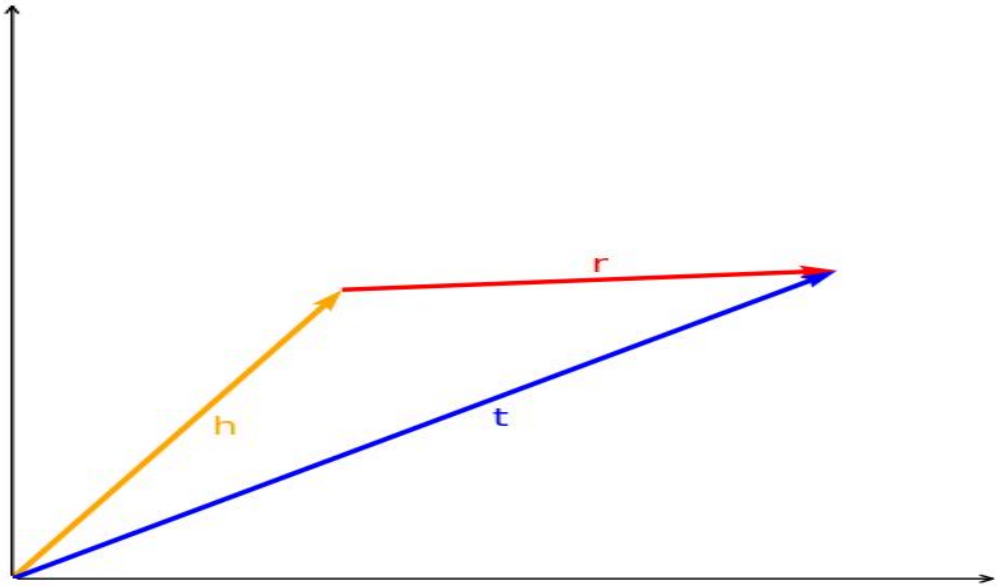
EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature	Selected Method
Translational	Encode triples as <s, r, o> with relation as a translation vector	Captures only **local edge patterns**; misses higher-order structure	**MultiKE**
GNN-based	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**; limited to k-hop	**RDGCN**
Graph Transformer	Self-attention across nodes (global or sampled subgraph)	Captures **long-range dependencies**; computationally heavy	**i-Align**
Cross-encoder	Embed entity pairs directly with local structure & attributes	No global graph embedding; focuses on **pairwise interactions**	**BERT-INT**

EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Translational	Encode triples as <h, r, t> with relation as a translation vector	Captures **local edge patterns**.

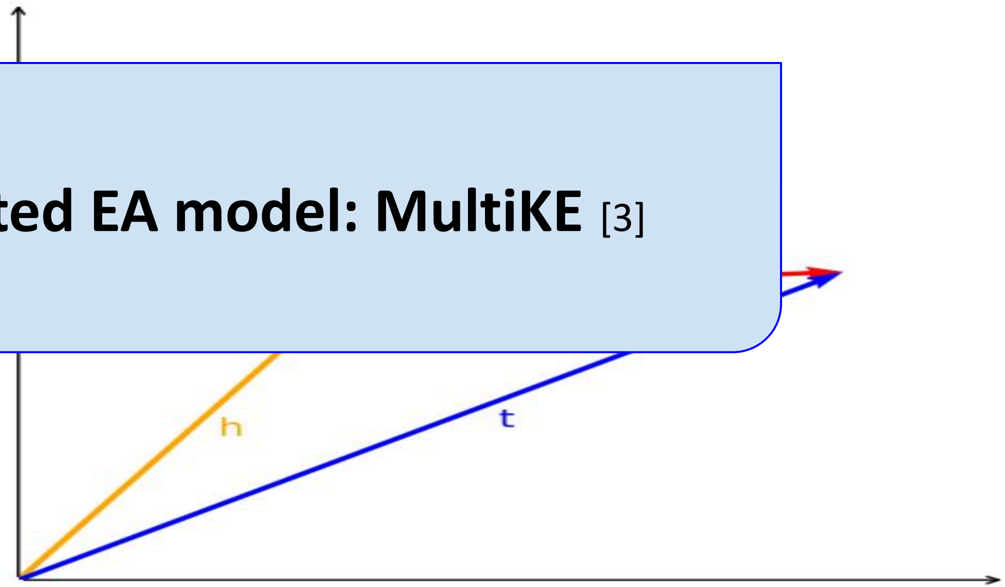
head entity (h)
relation (r)
tail entity (t)



EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Translational	Encode triples as <h, r, t> with relation as a translation vector	Captures **local edge patterns**.

head entit
relation (r)
tail entity (t)



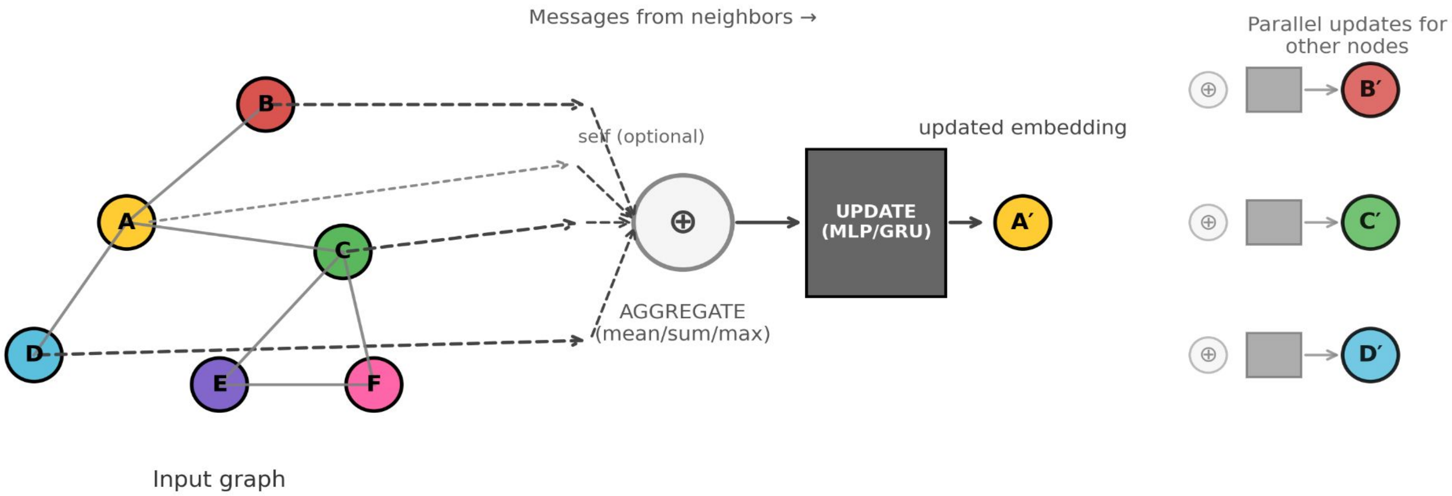
Selected EA model: MultiKE [3]

EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Translational	Encode triples as <h, r, t> with relation as a translation vector	Captures **local edge patterns**.
GNN-based	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**.

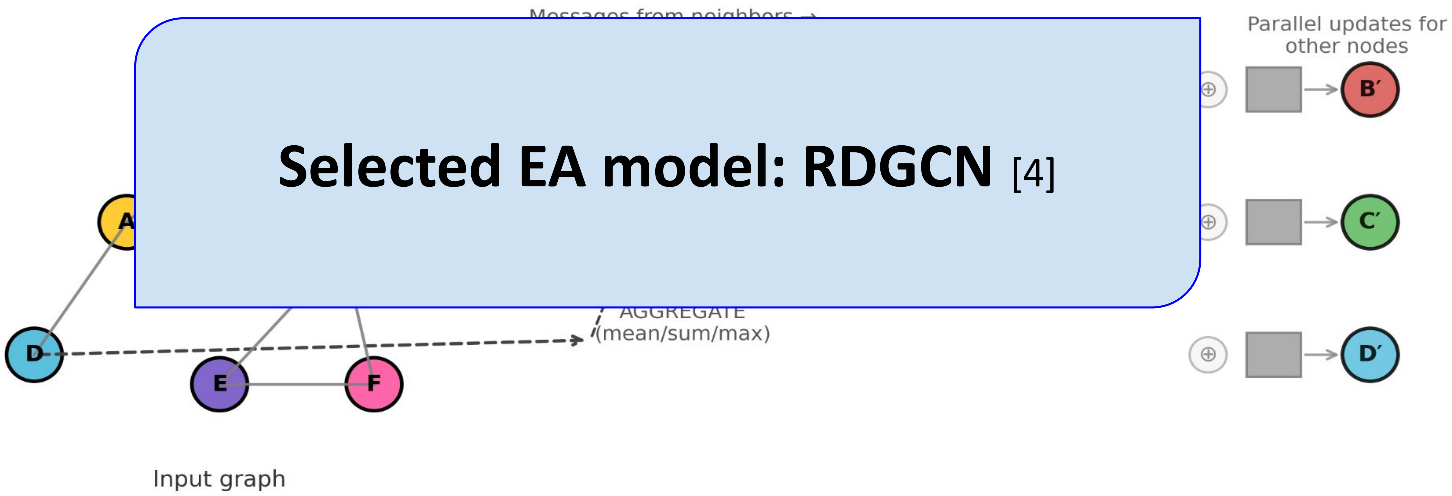
EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
GNN	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**; limited to k-hop



EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
GNN	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**; limited to k-hop



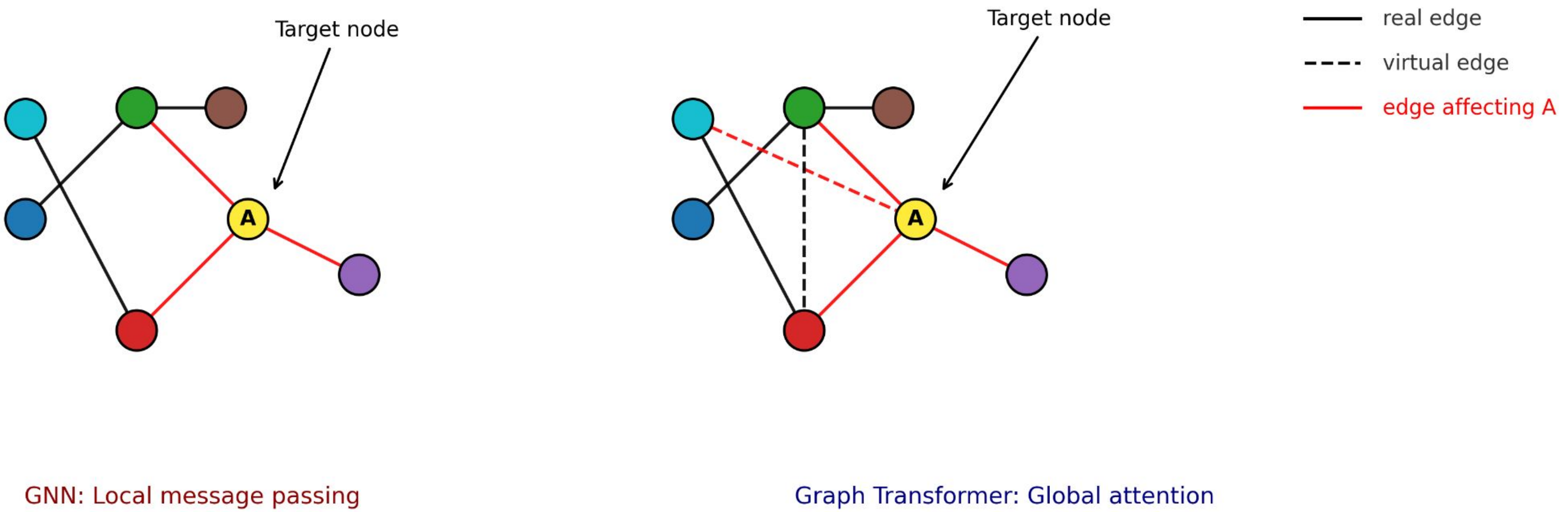
Selected EA model: RDGCN [4]

EA embedding-based methods

Method Family	**Structural Embedding Mechanism**	**Key Feature**
Translational	Encode triples as <h, r, t> with relation as a translation vector	Captures only **local edge patterns**
GNN-based	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**
Graph Transformer	Self-attention across nodes (global or sampled subgraph)	Captures **long-range dependencies**

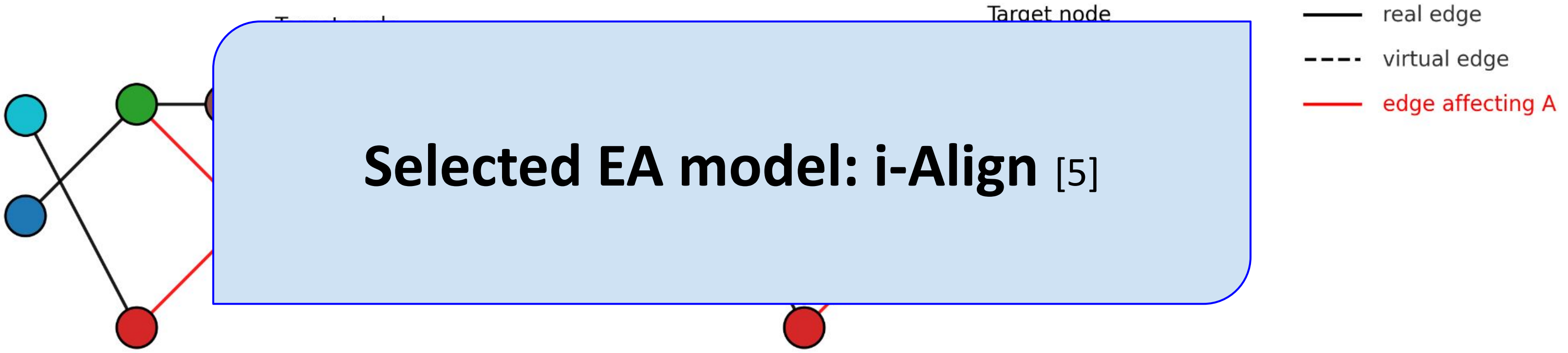
EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Graph Transformer	Self-attention across nodes (global or sampled subgraph)	Captures **long-range dependencies**; computationally heavy



EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Graph Transformer	Self-attention across nodes (global or sampled subgraph)	Captures **long-range dependencies**; computationally heavy



Selected EA model: i-Align [5]

GNN: Local message passing

Graph Transformer: Global attention

Contribution 1

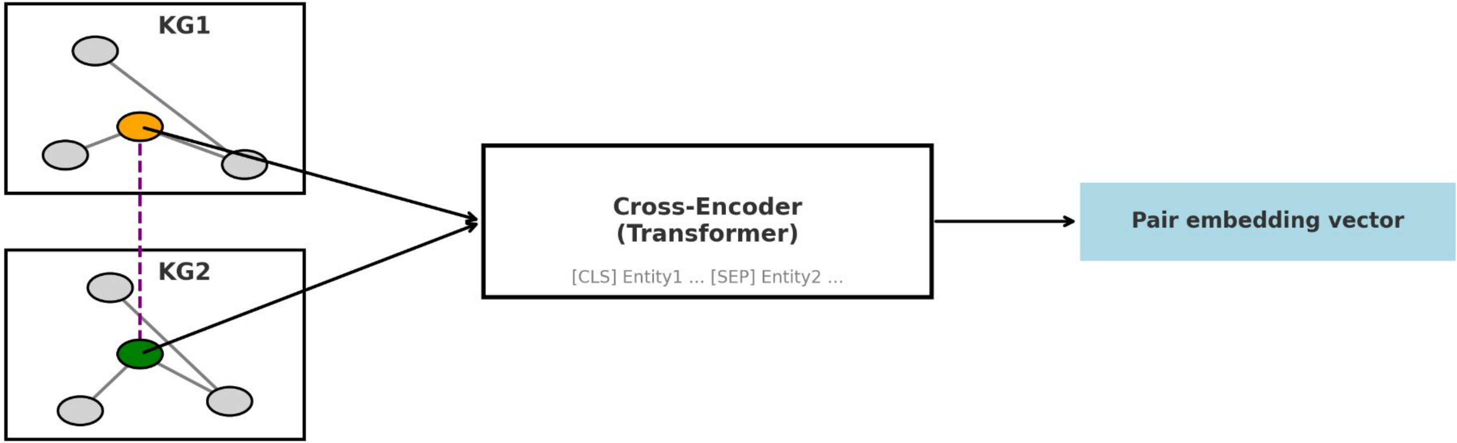
EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Translational	Encode triples as <h, r, t> with relation as a translation vector	Captures only **local edge patterns**; misses higher-order structure
GNN-based	Message passing: aggregate features from neighbors (k-hop)	Captures **local neighborhoods**; limited to k-hop
Graph Transformer	Self-attention across nodes (global or sampled subgraph)	Captures **long-range dependencies**; computationally heavy
Cross-encoder	Embed entity pairs directly with local structure & attributes	No global graph embedding; focuses on **pairwise interactions**

Contribution 1

EA embedding-based methods

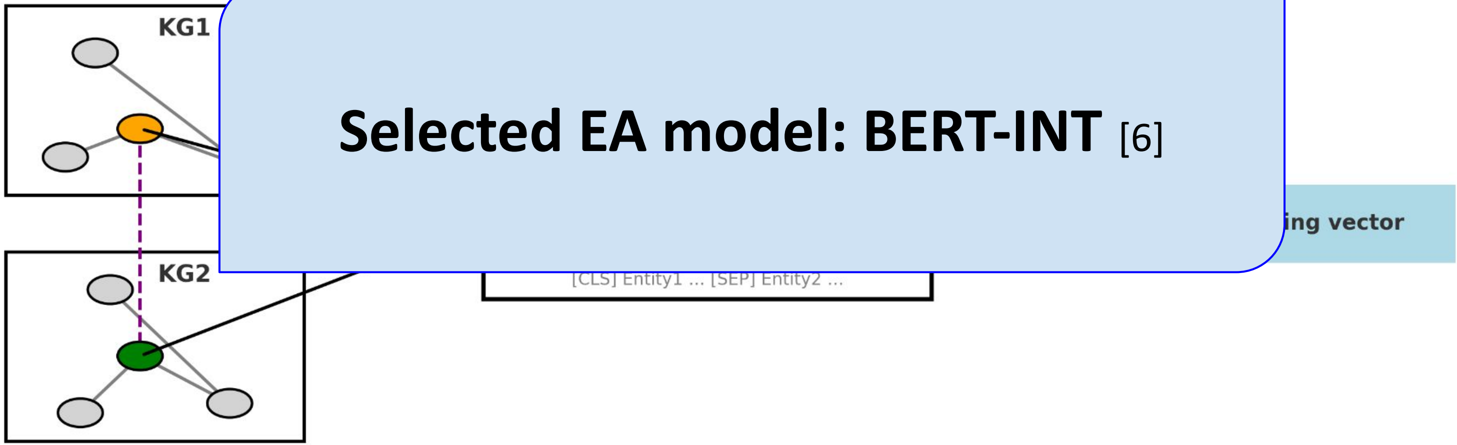
Method Family	Structural Embedding Mechanism	Key Feature
Cross-encoder	Embed entity pairs directly with local structure & attributes	No global graph embedding; focuses on **pairwise interactions**



Contribution 1

EA embedding-based methods

Method Family	Structural Embedding Mechanism	Key Feature
Cross-encoder	Embed entity pairs directly with local structure & attributes	No global graph embedding; focuses on **pairwise interactions**



Selected EA model: BERT-INT [6]

The Gaps

Results: Performance on Benchmark vs Real-world

Datasets		Methods							
		BERT-INT		RDGCN		MultiKE		i-Align	
		Hit@1	Hit@10	**Hit@1**	Hit@10	**Hit@1**	Hit@10	**Hit@1**	Hit@10
Benchmark	$DBP15K_{FR-EN}$	**99.3**	99.8	88.6*	95.7*	37.5	43.6	26.6	43.2
	SPIMBENCH	**82.4**	82.4	77.7	94.7	57.1	57.1	75.0	86.5
Real-world	DOREMUS	47.9	64.1	1.33	5.92	2.70	8.70	53.1	68.0
	AgroLD	21.1	33.2	0.02	0.3	2.30	5.7	4.4	12.1

Input preparation: https://github.com/DACE-DL/Create_Input_Data_to_EA_Models

Takeaways*

- Benchmark overfitting, where models struggle with generalization to unseen, real-world data.

- Cross-encoders e.g. BERT-INT transfer the best on real-world datasets.

* **These findings are published as:** *An analysis of the performance of representation learning methods for entity alignment: Benchmark vs. real-world data.*
Raoufi, E., Happi, B. G. H., Larmande, P., Scharffe, F., & Todorov, K. – **Semantic Web Journal**

Introduction

KG Integration & Entity Alignment, Methods & Challenges

The Gaps

Analysis of Entity Alignment Methods & Datasets

Reasoning EA

CENLIEA: A Cross-Encoder NLI Framework Enhanced by LLM for EA

Conclusion & Perspectives

Publications & Future works

Designing a Novel EA Method

- **Inductive** Learning
- **Transferable** on real-world data
- **Cross-Encoder** Model

Designing a Novel EA Method

- Inducti
- Transf
- Cross-

Our EA Method: Cenliea

Cenliea

- We use a pre-trained Natural Language Inference (NLI) model*:
  - It has fine-tuned on 2.7 million **hypothesis-premise pairs** in 27 languages.
  - The model predicts the probability of **"entailment"**, **"neutral"**, and **"contradiction"** in a given text pair.

* https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7

NLI Dataset Entailment Example

Premise*	Hypothesis	Label
How to take care of ducklings Find a **brooding box.** After ducklings have hatched from their shells and spent about 24 hours getting used to their new surroundings, they're ready to move to a brooder. A plastic storage **container**, sturdy cardboard box, or large glass aquarium can all work for this purpose.	Just about any **container** would work as a **brooder**	**entailment**

* Sample from: https://huggingface.co/datasets/MoritzLaurer/multilingual-NLI-26lang-2mil7

Reasoning EA

NLI Dataset Neutrality Example

Premise*	Hypothesis	Label
Paul Beard (4 August 1901 – 22 April 1989) **was** an English **violinist,** known particularly as leader of Sir Thomas Beecham's original London Philharmonic Orchestra and Sir Adrian Boult's BBC Symphony Orchestra.	**Paul Beard** has the highest **salary** in the **orchestra**	**neutral**

* Sample from: https://huggingface.co/datasets/MoritzLaurer/multilingual-NLI-26lang-2mil7

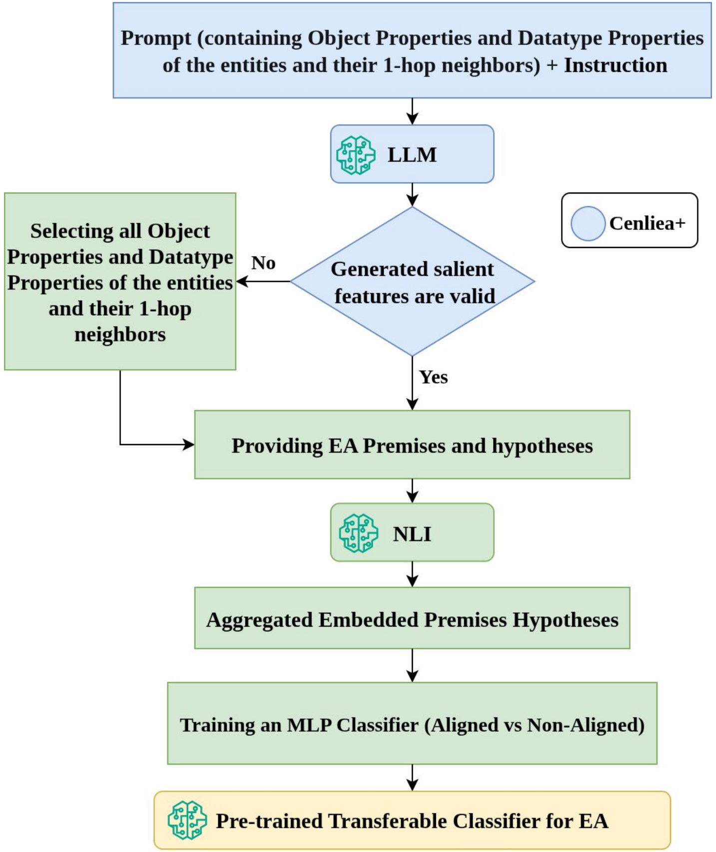
NLI Dataset Contradiction Example

Premise*	Hypothesis	Label
How to make a healthy breakfast Scramble some eggs in the microwave. Believe it or not, **you can** actually **make** really good scrambled **eggs** in the **microwave.**	A **microwave will not** cook **eggs.**	**contradiction**

* Sample from: https://huggingface.co/datasets/MoritzLaurer/multilingual-NLI-26lang-2mil7

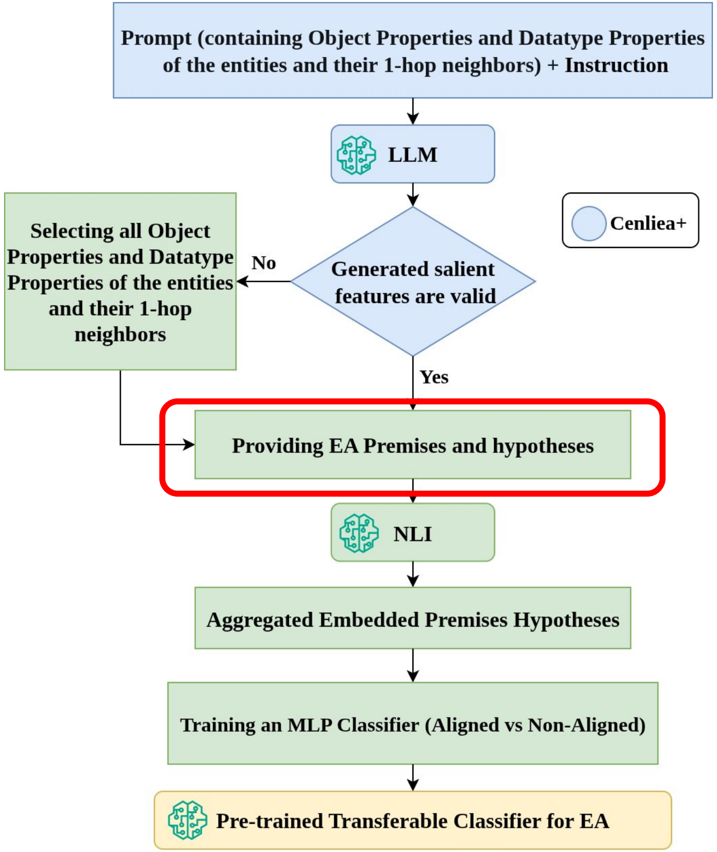
Cenliea: NLI-Based Model

- How to employ the NLI model for EA?



Cenliea: NLI-Based Model

- How to employ the NLI model for EA?



How to construct

EA premise-hypothesis?

Cenliea: The Structured Input

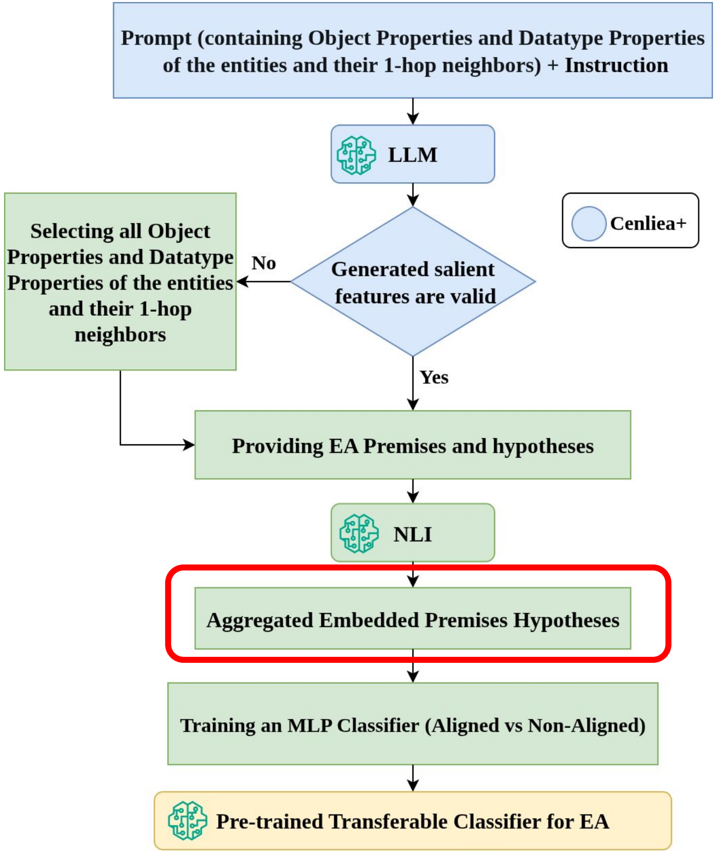
We extract a restricted **subset of subgraph** containing the given entity.

Premise*	Hypothesis
Entity 1 direct features: **1.** text: it is a time of chaos. a mysteriousnarmy led by, **2.** core altlabel: rise of the rakghouls, **3.** award3pub: gametrailers, **Entity 1 neighbor's features:** **R-1.1.** crew -> rdf schema comment: kurt futrell worked as a animator for bioware a, **R-2.1.** cast -> name: josh keaton, **R-2.2.** cast -> occupation: actor,	**Entity 2 direct features:** **a.** wikipagewikilinktext: star wars: the old republic, **b.** release: 2011-12-20, **Entity 2 neighbor's features:** **R-a.1.** developer -> core altlabel: bioware austin, **R-a.2.** developer -> abstract: bioware is a canadian ship building company fou,

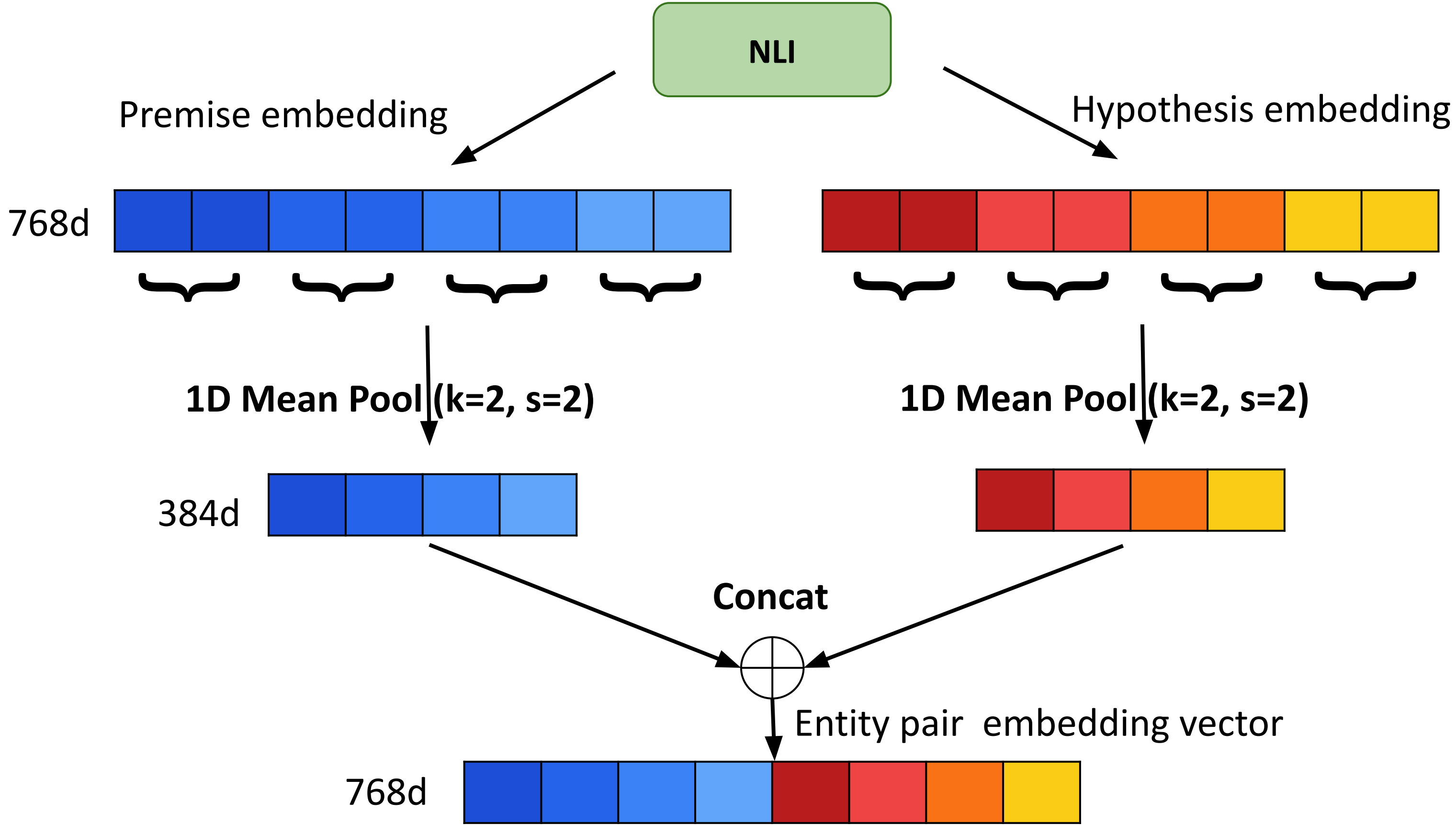
* Sample input generated by us: A positive sample from **2023StarwarsSWTOR** dataset in **KGtrack** (https://oaei.ontologymatching.org/2021/knowledgegraph/index.html)

Cenliea: Embedding and Aggregation

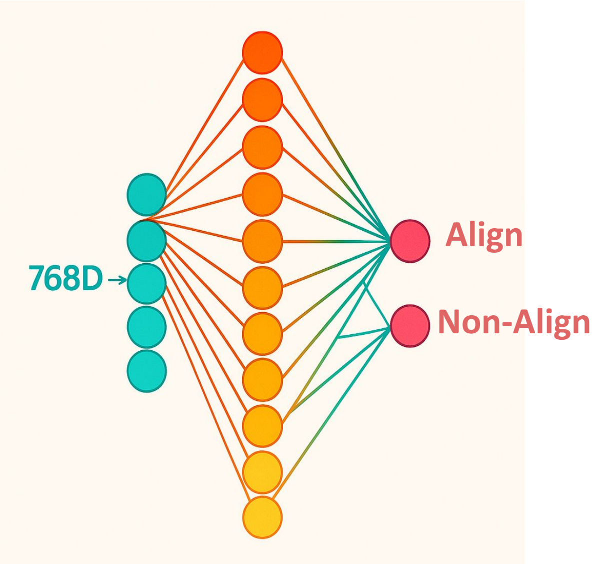
- How to embed entity pairs using NLI model?

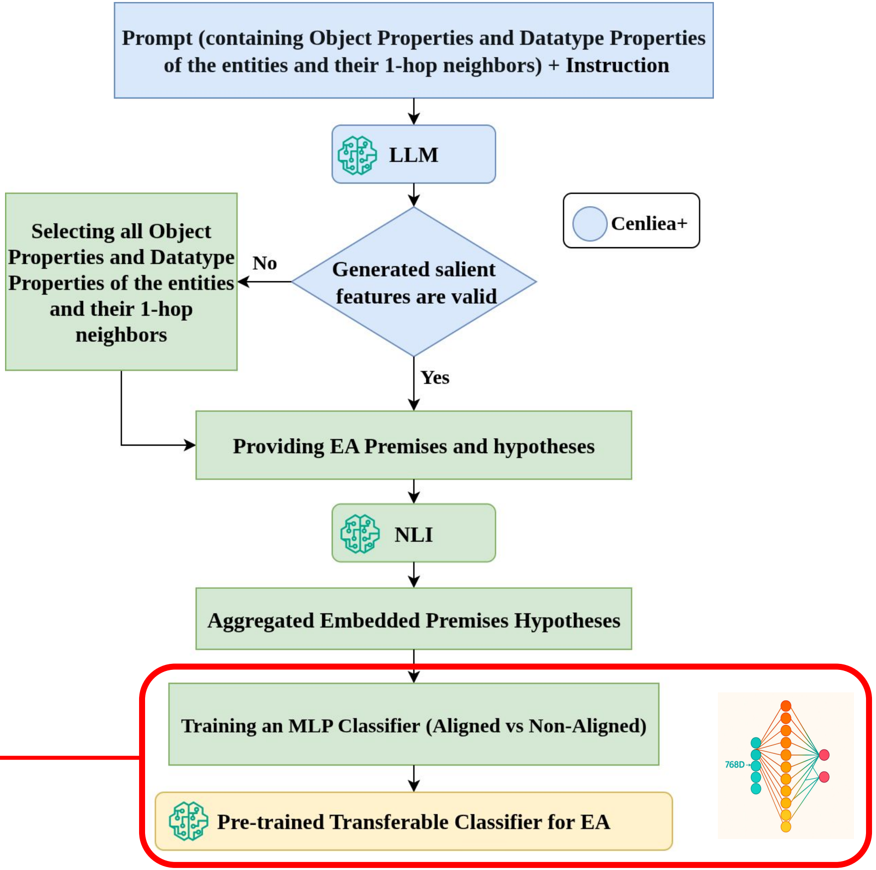


Cenliea: Embedding and Aggregation



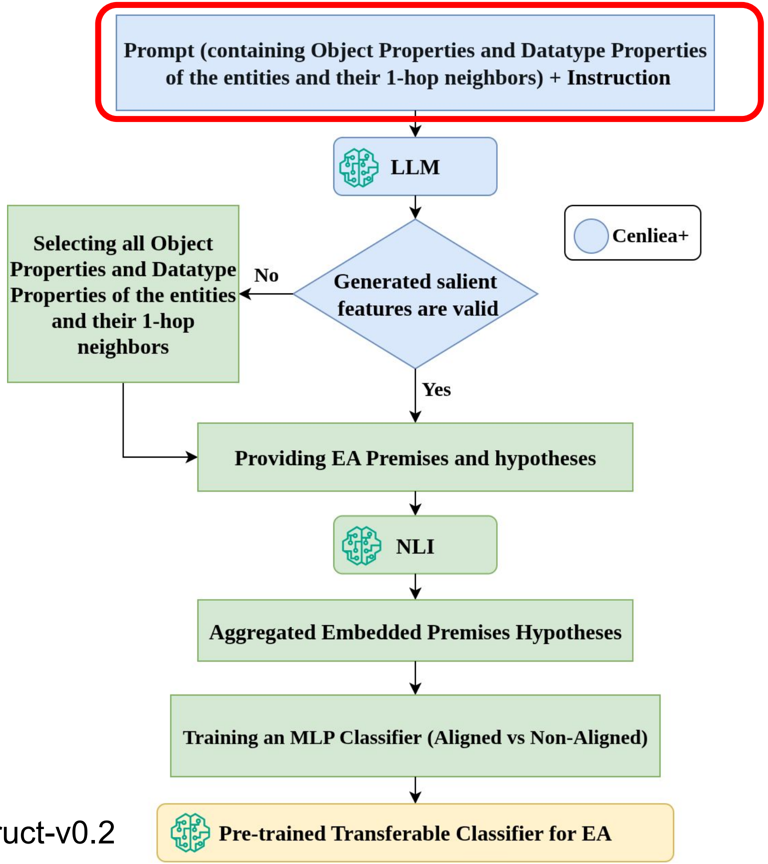
Cenliea: EA Classifier





Cenliea+: LLM-Enhanced Premise-Hypothesis Generation

- We use a **Mistral 7B*** model as the LLM.



* https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2

Cenliea+: Prompt

Structured input features of entity pair

- Instruction:

+

Use numerical-alphabetical enumerators given in the prompt and give me tuples of features of entity 1 & 2 e.g. (2, c) which represent their salient similarities.

Cenliea+: Prompt

Structured input features of entity pair

- Instruction:

+

Use numerical-alphabetical enumerators given in the prompt and give me tuples of features of entity 1 & 2 e.g. (2, c) which represent their salient similarities.

- 1-shot prompt: The demonstration output:

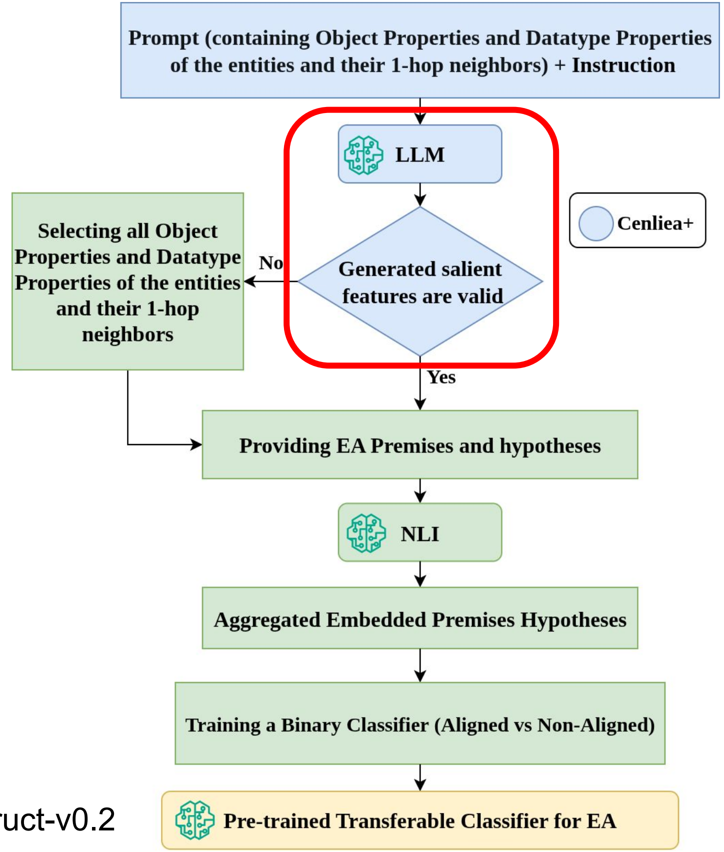
+

(2, a): Both entities have a similar title, although with English to French translation.

(R-3, R-a): Entities have the same authors.

Cenliea+: Response Generation by LLM

- Generating responses using Mistral 7B*.



* https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2

Cenliea+: Prompt & LLM Response

<table>
<tr><th>Prompt</th><th>Response generated by Mistral7B</th></tr>
<tr><td>

Entity 1 direct features:
1. p102 has title: cycle du feu,
2. p3 has note: le 17 décembre 1992 à l'espace de projection de,
Entity 2 direct features:
a. p3 has note: pour orchestre de chambre et projection audio-n,
b. p102 has title: feuillages,
Entity 1 neighbor's features:
R-1.1. ontology u13 has casting -> p3 has note: flûte, hautbois, clarinette en la, cor, trompet,
R-2.1. ontology u12 has genre -> p1 is identified by: musique contemporaine,
Entity 2 neighbor's features:
R-a.1. ontology u13 has casting -> p3 has note: orchestre de chambre,
Use numerical-alphabetical enumerators given in the prompt and give me tuples of features of entity 1 & 2 e.g. (2, c) which represent their salient similarities:

</td><td>

(1, b): Both entities have titles related to nature (1. p102 has title: cycle du feu, AND b. p102 has title: feuillages).

(R-2.1, R-a.1): Both entities belong to the contemporary music genre (2. ontology u12 has genre -> p1 is identified by: musique contemporaine, AND a. ontology u13 has casting -> p3 has note: orchestre de chambre).

Sample is taken from DoReMus

</td></tr>
</table>

Cenliea+: Prompt & LLM Response

Prompt	Response generated by Mistral7B
Entity 1 direct features: **1.** p102 has title: cycle du feu, **2.** p3 has note: le 17 décembre 1992 à l'espace de projection de, **Entity 2 direct features:** **a.** p3 has note: pour orchestre de chambre et projection audio-n, **b.** p102 has title: feuillages, **Entity 1 neighbor's features:** R-1.1. ontology u en la, cor, trompet R-2.1. ontology u contemporaine, **Entity 2 neighbo** R-a.1. ontology u **Use numerical-al** **me tuples of feat** **salient similaritie**	**(1, b):** Both entities have titles related to nature (1. p102 has title: cycle du feu, AND b. p102 has title: feuillages). **(R-2.1, R-a.1):** Both entities belong to the contemporary music genre (2. ontology u12 has genre -> p1 is identified by: musique contemporaine, AND a. ontology u13 has casting -> p3 has note: orchestre de chambre).

Response Sanity Check

Cenliea+: Prompt & LLM Response

Prompt	Response generated by Mistral7B
Entity 1 direct features: **1.** p102 has title: cycle du feu, **2.** p3 has note: le 17 décembre 1992 à l'espace de projection de, **Entity 2 direct features:** **a.** p3 has note: pour orchestre de chambre et projection audio-n, **b.** p102 has title: feuillages, **Entity 1 neighbor's features:** R-1.1. ontology u en la, cor, trompet R-2.1. ontology u contemporaine, **Entity 2 neighbo** R-a.1. ontology u **Use numerical-al** **me tuples of feat** **salient similaritie**	**(1, b):** Both entities have titles related to nature (1. p102 has title: cycle du feu, AND b. p102 has title: feuillages). **(R-2.1, R-a.1):** Both entities belong to the contemporary music genre (2. ontology u12 has genre -> p1 is identified by: musique contemporaine, AND a. ontology u13 has casting -> p3 has note: orchestre de chambre).

✅ **(1, b)** →

Premise 1: **1.** p102 has title: cycle du feu,

Hypothesis 1: **b.** p102 has title: feuillages,

Cenliea+: Prompt & LLM Response

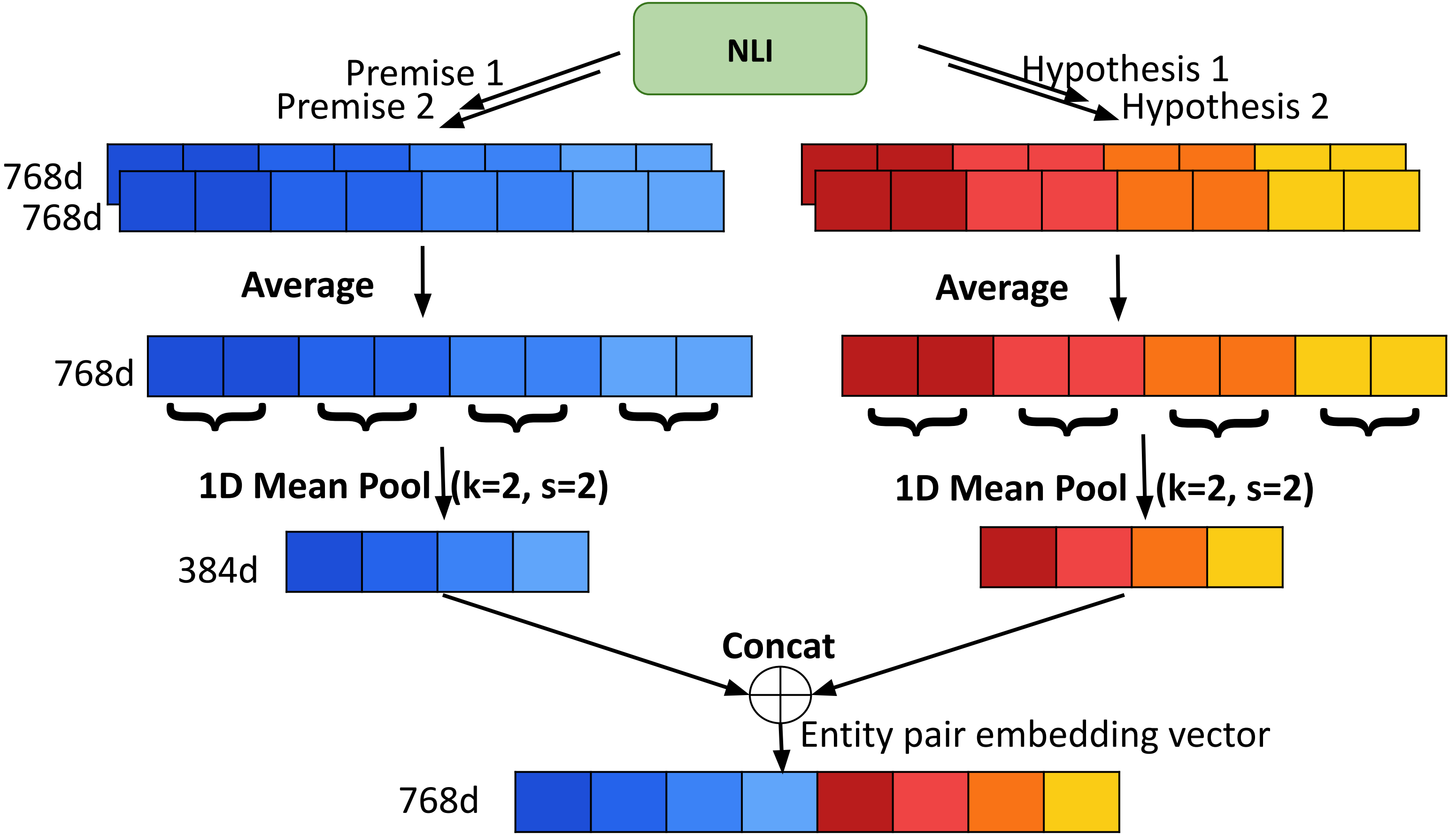
Prompt	Response generated by Mistral7B
Entity 1 direct features: **1.** p102 has title: cycle du feu, **2.** p3 has note: le 17 décembre 1992 à l'espace de projection de, **Entity 2 direct features:** **a.** p3 has note: pour orchestre de chambre et projection audio-n, **b.** p102 has title: feuillages, **Entity 1 neighbor's features:** R-1.1. on en la, cor, R-2.1. on contempo **Entity 2 n** R-a.1. on **Use num** **me tuples** **salient si**	**(1, b):** Both entities have titles related to nature (1. p102 has title: cycle du feu, AND b. p102 has title: feuillages). **(R-2.1, R-a.1):** Both entities belong to the contemporary music genre (2. ontology u12 has genre -> p1 is identified by: musique contemporaine, AND a. ontology u13 has casting -> p3 has note: orchestre de chambre).

✅(2, a) →

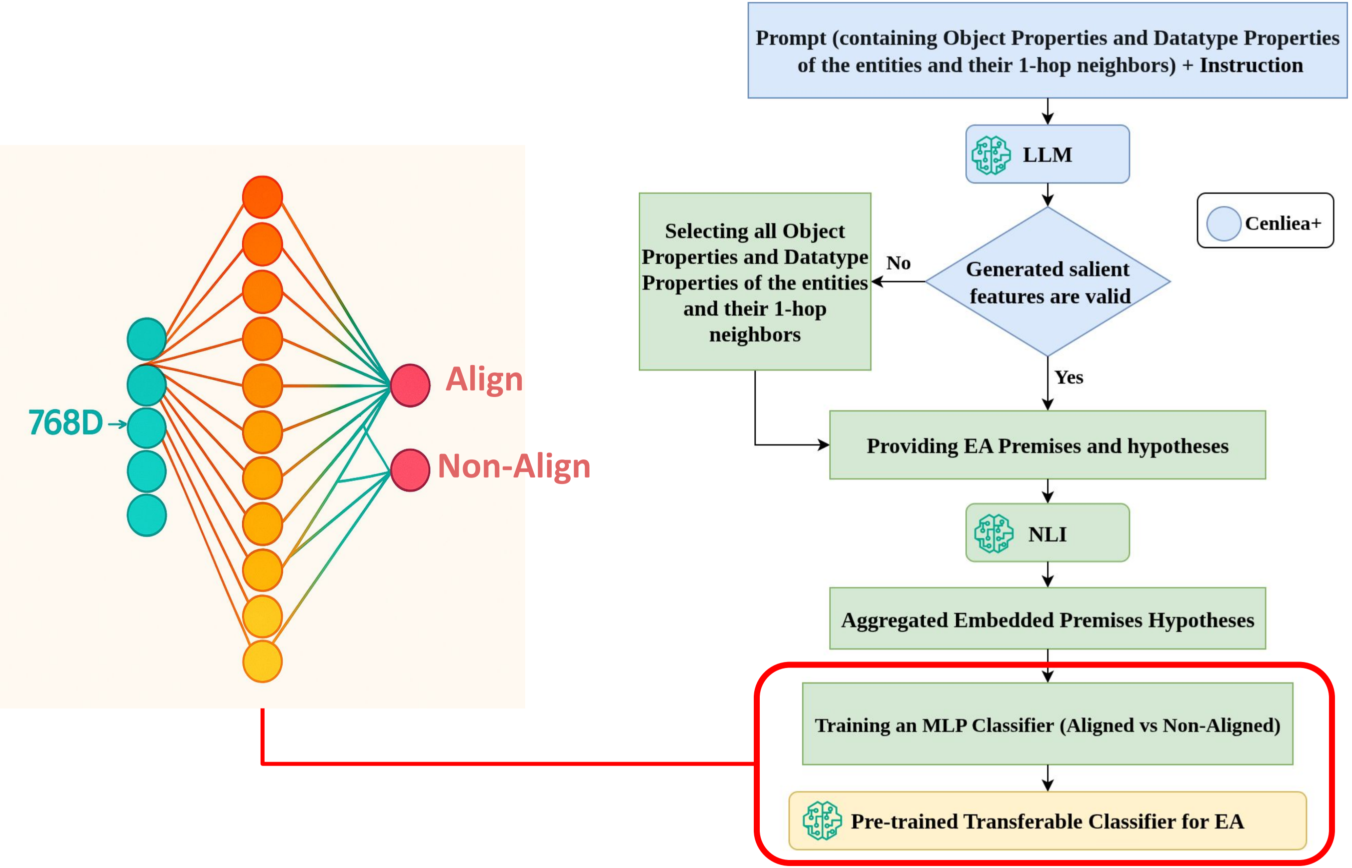
Premise 2: **2.** p3 has note: le 17 décembre 1992 à l'espace de projection de,

Hypothesis 2: **a.** p3 has note: pour orchestre de chambre et projection audio-n,

Cenliea+: Embedding and Aggregation

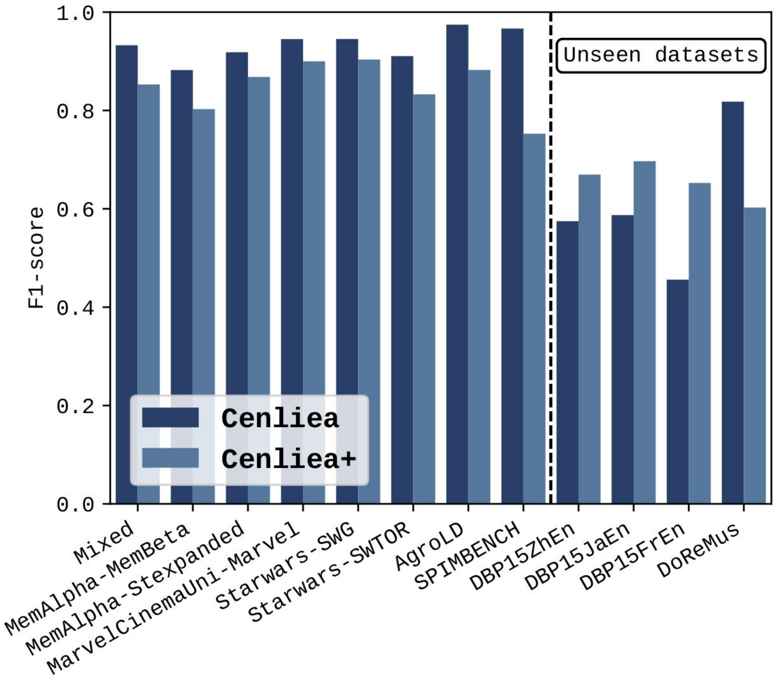


Cenliea+: EA Classifier



Experiments

Experiments: Comparing CENLIEA and CENLIEA+



Comparison between the performance of final binary classifiers in Cenliea and Cenliea+.
Note: **Unseen** datasets are evaluated using **zero-shot** inference.

Mixed: merged pairs from AgroLD, SPIMBENCH, and five OAEI KGTrack datasets.

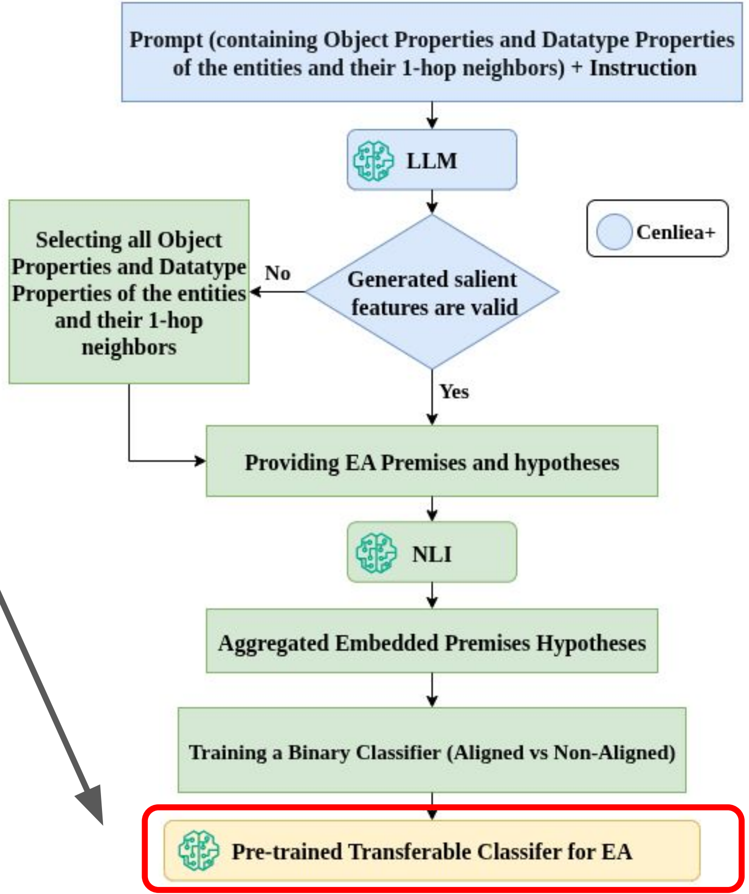
Complementarity of Cenliea & Cenliea+

Cenliea Confidence Threshold	Fallback Cases	Cenliea+ Correct	Cenliea Lost	Only-Cenliea+ Recoveries	Net Gain	Macro F1 on Mixed dataset
0.60	932	601	160	253	**+93**	0.9270
0.65	1 489	985	245	394	**+149**	0.9287
0.70	2 041	1 364	334	515	**+181**	0.9296
0.75	2 702	1 806	451	611	**+160**	0.9290

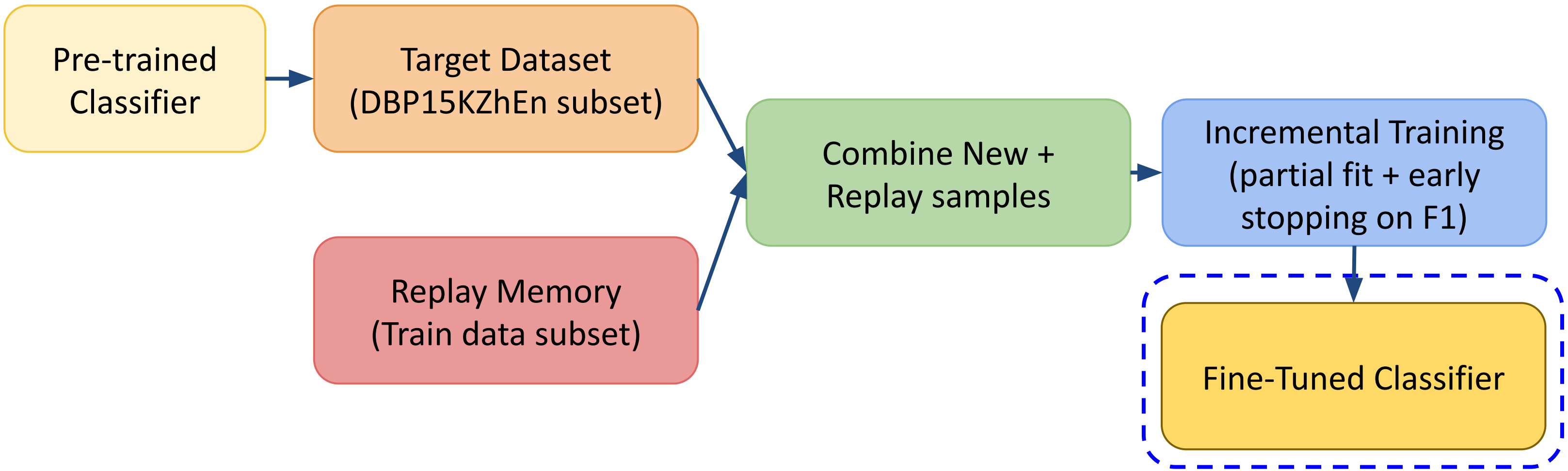
Performance of the Cenliea+ fallback strategy at different confidence thresholds on Cenliea.

Experiments: Fine-Tuning (FT)

Goal: Fine-Tuning the classifier on a proportion of unseen data



Experiments: Replay-Based FT



Result:

Fine-tuning on DBP15ZhEn **boosts** Cenliea's **recall** by up to **+60%** on unseen datasets.

Replay-Based FT for Unseen/Partially-seen Domains

Dataset	Δ Recall (%)
DBP15ZhEn	+60
DBP15KFrEn	+66
DBP15KJaEn	+60
DoReMus	+20
Starwars–SWTOR	+9
AgroLD	–18

Average positive-class **recall** improvements from replay-based fine-tuning of Cenliea, across datasets.

Experiments: Selected Baselines

We choose the following baselines for comparison:

- BERT-INT [6]
- i-Align [5]
- **TEA** [9]: Measures similarity (i.e. how much one entity's information entails another's) via MLM[1] or NSP[2].

1. MLM: Masked Language Modelling
2. NSP: Next Sentence Prediction

Experiments: Selected Baselines

We choose the following baselines for comparison:

- BERT-INT [6]
- i-Align [5]
- TEA [9]
- **DAEA** [10]: **Transfer learning** by choosing an enriched source dataset (e.g. DBP15K-FrEn) with a distribution similar to the real-world test data (e.g. AgroLD, DoReMus).

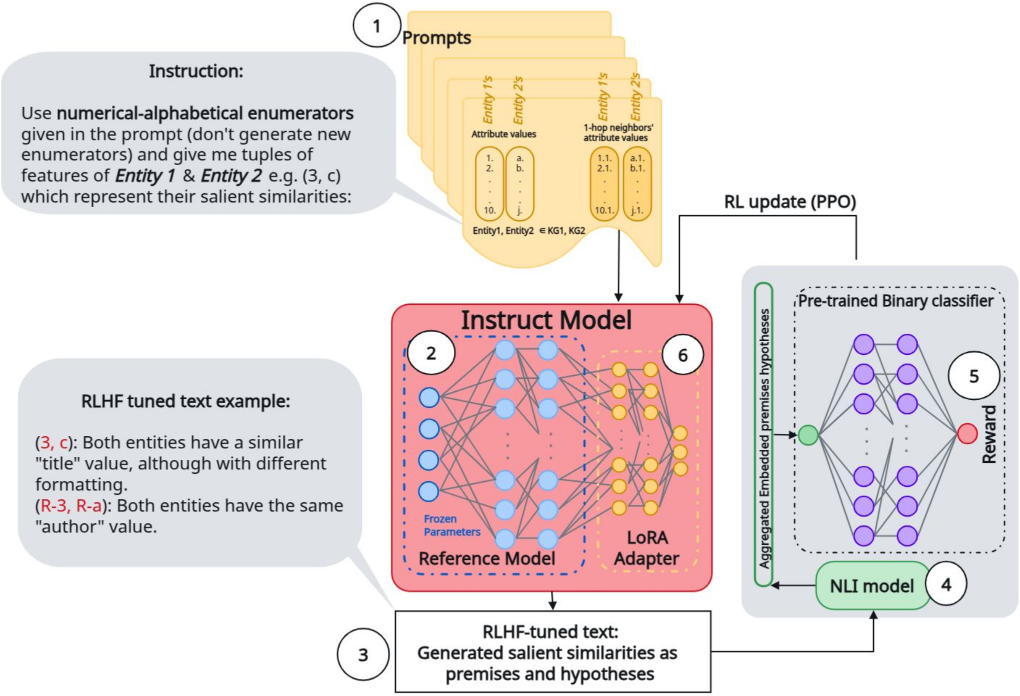
Experiments: Generalization across Datasets

	Benchmark						Real-world					
	SPIMBENCH			DBP15FrEn			AgroLD			DoReMus		
Method	Rec.	Prec.	F_1	Rec.	Prec.	F_1	Rec.	Prec.	F_1	Rec.	Prec.	F_1
BERT-INT	82.4	"	"	99.3	"	"	21.1	"	"	47.9	"	"
i-Align	75	"	"	26.6	"	"	4.4	"	"	53.1	"	"
TEA	84.8	"	"	98.7	"	"	1.12	"	"	24.4	"	"
DAEA	91.9	"	"	98.8	"	"	27.1	"	"	77.8	"	"
Cenliea+	70.5	67.3	68.9	91.0*	80.6*	85.5*	88.3	87.9	88.1	88.6*	66.4*	75.9*
Cenliea	86.8	89.2	88	94.1*	95.3*	94.9*	96.7	98.9	97.8	91.6*	87.9*	89.8*

Results (positive-class scores) on datasets comparing baselines with Cenliea and Cenliea+. We highlight the best and the **second best** recalls of each dataset.

* indicates result of the **fine-tuned** model on **30%** of the data (**Not a zero-shot inference**).

Experiments: LLM fine-tuning Framework



LLM fine-tuning framework using reinforcement learning. A reward model guides the adaptation of a pre-trained reference model into an instruct model by updating only a low-rank subset of parameters.

Takeaways

- Cenliea **outperforms** baselines on **real-world** datasets.
- **Unsupervised applicability** of our pre-trained classifiers.
- Cenliea+ improves **multilingual transfer**.
- A need for a Meta-classifier for **Cenliea** ↔ **Cenliea+** integration

Introduction

Terminology,
Methods &
Challenges

The Gaps

Analysis of Entity
Alignment
Datasets &
Methods

Reasoning EA

CENLIEA: A
Cross-Encoder NLI
Framework Enhanced
by LLM for Entity
Alignment

Conclusion & Perspectives

Conclusion &
Future works

Conclusion

- Systematic evaluation → revealed **generalization limits** [27]

Conclusion

- Systematic evaluation → revealed **generalization limits**
- **Cenliea** and **Cenliea+** (NLI + LLM)→ **Transferable** & **explainable** across domains

Conclusion

- Systematic evaluation → revealed **generalization limits**
- **Cenliea** and **Cenliea+** (NLI + LLM)→ **Transferable** & **explainable** across domains
- **HMatch** [22] (hybrid embeddings + link keys) → **Scalable** & **explainable** alignments

Conclusion

- Systematic evaluation → revealed **generalization limits**
- **Cenliea** and **Cenliea+** (NLI + LLM)→ **Transferable** & **explainable** across domains
- **HMatch** (hybrid embeddings + link keys) → **Scalable** & **explainable** alignments

👉 Toward: **EA systems that are scalable, explainable, and generalizable**

Perspective

- Toward **continuous learning EA** (dynamic, evolving KGs)

Perspective

- Toward **continuous learning EA** (incremental, evolving KGs)
- **Confidence calibration** for EA model’s self-updates

Perspective

- Toward **continuous learning EA** (incremental, evolving KGs)
- **Confidence calibration** for EA model's self-updates
- **Resource-efficient models** (pruning, distillation, distributed)

Perspective

- Toward **continuous learning EA** (incremental, evolving KGs)
- **Confidence calibration** for EA model’s self-updates
- **Resource-efficient models** (pruning, distillation, distributed)

Vision: **autonomous, self-adaptive EA assistants** for large-scale KG integration

Open to Collaboration

Looking for R&D opportunities in:

- Applied AI / LLMs
- Knowledge Graphs & Data Engineering
- Scalable inference & fine-tuning (GPU / batching / pipelines)

I’m interested in:

- R&D projects
- Tech transfer & co-development
- Productized ML systems

Contact:

Email: Ensiyeh.raoufi@umontpellier.fr

LinkedIn: https://www.linkedin.com/in/ensiyeh-raoufi/

Thank you for listening.

“Out beyond ideas of wrongdoing and rightdoing, There is a field. I'll meet you there.”
Rumi

References

1. Achichi, M., Bellahsene, Z., Ellefi, M. B., & Todorov, K. (2019). Linking and disambiguating entities across heterogeneous RDF graphs. *Journal of Web Semantics*, *55*, 108-121.
2. Baek, Jinheon, Dong Bok Lee, and Sung Ju Hwang. "Learning to extrapolate knowledge: **Transductive few-shot out-of-graph** link prediction." *Advances in neural information processing systems 33 (2020): 546-560.*
3. Zhang, Qingheng, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu."**Multi-view** knowledge graph embedding for entity alignment." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. 2019: 5429–5435.
4. Wu, Yuting, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. "**Relation-aware** entity alignment for heterogeneous knowledge graphs." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. 2019: 5278–5284.
5. Trisedya, Bayu Distiawan, et al. "**i-Align**: an interpretable knowledge graph alignment model." *Data Mining and Knowledge Discovery 37.6 (2023)*: 2494-2516.
6. Tang, Xiaobin, et al. "**BERT-INT**: A BERT-based interaction model for knowledge graph alignment." *interactions 100 (2020)*: e1.
7. Leone, Manuel, et al. "**A critical re-evaluation** of neural methods for entity alignment." *Proceedings of the VLDB Endowment 15.8 (2022)*: 1712-1725.
8. **Sun**, Zequn, et al. **"A Benchmarking Study** of Embedding-Based Entity Alignment for Knowledge Graphs." *Proceedings of the VLDB Endowment, vol. 13, no. 12, VLDB Endowment, July 2020*, pp. 2326–2340.
9. Zhao, Yu, et al. "From **Alignment to Entailment: A Unified Textual Entailmen**t Framework for Entity Alignment." Findings of the Association for Computational Linguistics: ACL 2023, edited by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, *Association for Computational Linguistics, 2023*, pp. 8795–8806.
10. Yang, Linyan, et al. "**DAEA**: Enhancing Entity Alignment in Real-World Knowledge Graphs Through Multi-Source Domain Adaptation." Proceedings of the 31st International Conference on Computational Linguistics, edited by Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert, *Association for Computational Linguistics, 2025*, Abu Dhabi, UAE, pp. 5890–5901. ACL Anthology
11. Rodrigues, D., & da Silva, A. (2021). A study on machine learning techniques for the schema matching network problem. *Journal of the Brazilian Computer Society*, 27(1), 14.
12. Jiménez-Ruiz, E., & Cuenca Grau, B. (2011). **LogMap**: Logic-based and scalable ontology matching. *The Semantic Web – ISWC 2011*, 273–288. Springer.
13. Ngonga Ngomo, A. C., & Auer, S. (2011). **LIMES**: A time-efficient approach for large-scale link discovery on the web of data. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011),* 2312–2317.
14. Soru, Tommaso, Edgard Marx, and Axel-Cyrille Ngonga Ngomo. "**ROCKER**: A refinement operator for key discovery." *Proceedings of the 24th International Conference on World Wide Web*. 2015.
15. Nacira Abbas, Jérôme David, Amedeo Napoli. **Linkex**: A Tool for Link Key Discovery Based on Pattern Structures. *2019*. ffhal-02168775f.
16. Abbas, Nacira, Jérôme David, and Amedeo Napoli. "Discovery of link keys in RDF data based on pattern structures: Preliminary steps." *CLA 2020-The 15th International Conference on Concept Lattices and Their Applications*. 2020.
17. W. Zeng, X. Zhao, Z. Tan, J. Tang and X. Cheng, "Matching Knowledge Graphs in Entity Embedding Spaces: An Experimental Study" in *IEEE Transactions on Knowledge & Data Engineering*, vol. 35, no. 12, 2023.
18. Pedro Giesteira Cotovio, Ernesto Jimenez-Ruiz, and Catia Pesquita. "What can knowledge graph alignment gain with **Neuro-Symbolic** learning approaches?" *In: arXiv preprint arXiv:2310.07417 (2023)* (cit. on p. 21)

References

19. Chen, Shengyuan, et al. "**Neuro-symbolic** entity alignment via variational inference." *arXiv preprint arXiv:2410.04153* (2024).
20. Liu, B., Scells, H., Hua, W., Zuccon, G., Zhao, G., & Zhang, X. (2022). Guiding neural entity alignment with **compatibility**. *In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 491–504). Association for Computational Linguistics.*
21. Qi, Z., Zhang, Z., Chen, J., Chen, X., Xiang, Y., Zhang, N., & Zheng, Y. "Unsupervised knowledge graph alignment by **probabilistic reasoning** and semantic embedding." *In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (pp. 2019–2025). International Joint Conferences on Artificial Intelligence Organization.*
22. Chloé Khadija Jradeh, Ensiyeh Raoufi, Jérôme David, Pierre Larmande, François Scharffe, Konstantin Todorov, and Cassia Trojahn. 2025. **Graph Embeddings Meet Link Keys** Discovery for Entity Matching. *In Proceedings of the ACM on Web Conference 2025 (WWW '25). Association for Computing Machinery*, New York, NY, USA, 3344–3353.
23. Suchanek, F. M., Abiteboul, S., & Senellart, P. (2011). **PARIS**: Probabilistic alignment of relations, instances, and schema. *Proceedings of the VLDB Endowment*, 5(3), 157–168.
24. B.G. Happi Happi, G. Fokou Pelap, D. Symeonidou and P. Larmande, **DLinker** Results for OAEI 2022, in: 17th International Workshop on Ontology Matching, *OM 2022, CEUR Workshop Proceedings, Vol. 3324,* CEUR-WS, 2022, pp. 166–173.
25. Zequn Sun, Jiacheng Huang, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. “**Transedge**: Translating relation-contextualized embeddings for knowledge graphs”. In: *The Semantic Web–ISWC 2019*.
26. Sun, Zequn, Wei Hu, and Chengkai Li. "Cross-lingual entity alignment via joint attribute-preserving embedding.*" International semantic web conference. Cham: Springer International Publishing,* 2017.
27. Raoufi, E., Happi, B. G. H., Larmande, P., Scharffe, F., & Todorov, K. **An analysis of the performance** of representation learning methods for entity alignment: Benchmark vs. real-world data.
28. Sun, Zequn, et al. "**Bootstrapping** entity alignment with knowledge graph embedding." *Ijcai. Vol. 18. No. 2018.*
29. Zhang, Rui, et al. "A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning." *The VLDB Journal 31.5 (2022)*: 1143-1168.
30. N. Fanourakis, V. Efthymiou, V. Christophides, D. Kotzinos, E. Pitoura and K. Stefanidis, “Structural Bias in Knowledge Graphs for the Entity Alignment Task”, in: *The Semantic Web: 20th International Conference, ESWC 2023.*
31. Kexuan Xin, Zequn Sun, Wen Hua, Wei Hu, and Xiaofang Zhou. Informed multi-context entity alignment. *In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022.*
32. Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. LightEA: A Scalable, Robust, and Interpretable Entity Alignment Framework via Three-view Label Propagation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP. 2022.*
33. Danai Symeonidou, Vincent Armant, Nathalie Pernelle, and Fatiha Saïs. SAKey: Scalable Almost Key Discovery in RDF Data. *In The Semantic Web – ISWC. 2014.*