MANCHESTER 1824
The University of Manchester

Exploring Large Language Models in Ontology Reasoning and Construction

Jiaoyan Chen
Department of Computer Science
The University of Manchester

Plate-Forme Intelligence Artificielle (PFIA) 2026
37es journées de l'**I**ngénierie des **C**onnaissances (IC),
1-3 July 2026, Arras, France

Part I: Knowledge Representation Learning

Part II: Ontology Reasoning with LLM Generation

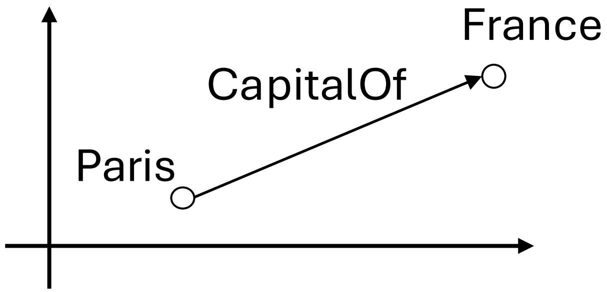
Embedding Symbolic Knowledge

- **Vector** or **parameter**-based representation of symbolic knowledge
  - A kind of neural-symbolic integration

- Low-dimension Word Embedding e.g., Word2Vec (Google 2013)
  - $\mathrm{V_{king}} - \mathrm{V_{queen}} \approx \mathrm{V_{man}} - \mathrm{V_{woman}}$ (distributional hypothesis)
  - $\cos(\mathrm{V_{cat}}, \mathrm{V_{kitten}}) > \cos(\mathrm{V_{cat}}, \mathrm{V_{dog}})$ (similarity in usage; cooccurrence)

Knowledge Graph Embedding

- Represent KG **entities** and **relations** by **low-dimension vectors** with their relationships concerned
  - Motivated by word embedding; mostly aim at facts of RDF triple e.g., <Paris, CapitalOf, France>
    - Geometric modeling: TransE, TransR, TransH, ..



TransE: modeling relation as a translation mapping

- Sequence learning: RDF2Vec, ...
- Graph propagation: R-GCN, ...

Ontology (Description Logic) Embedding

- How to represent more complex ontologies of Description Logic (DL) in Euclidean space?

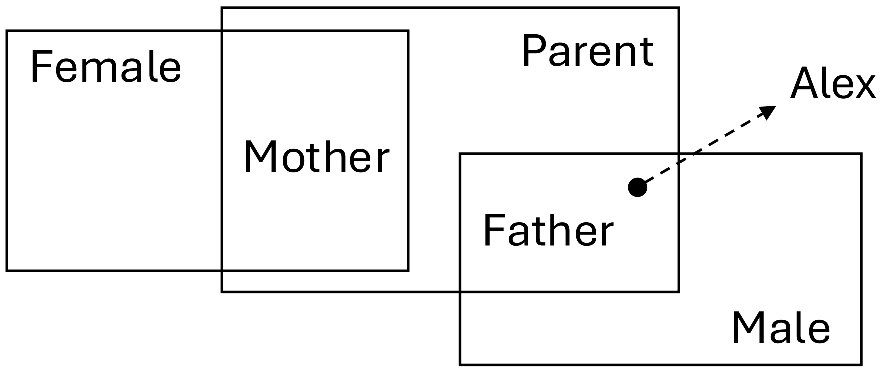
$$\mathcal{T} = \{\text{Father} \sqsubseteq \text{Parent} \sqcap \text{Male},\ \text{Mother} \sqsubseteq \text{Parent} \sqcap \text{Female},$$
$$\text{Child} \sqsubseteq \exists \text{hasParent.Father},\ \text{Child} \sqsubseteq \exists \text{hasParent.Mother},$$
$$\text{hasParent} \sqsubseteq \text{relatedTo}\}$$
$$\mathcal{A} = \{\text{Father}(\text{Alex}), \text{Child}(\text{Bob}), \text{hasParent}(\text{Bob}, \text{Alex})\}$$

A toy famil ontology in DL $\mathcal{EL}^{++}$ which allows complex concept construction:

$$\bot \mid \top \mid A \mid C \sqcap D \mid \exists r.C \mid \{a\}$$

- Embedding: Region-based
  - Individual – Point
  - Concept – Region

    E.g., n-dimensional Ball and Box

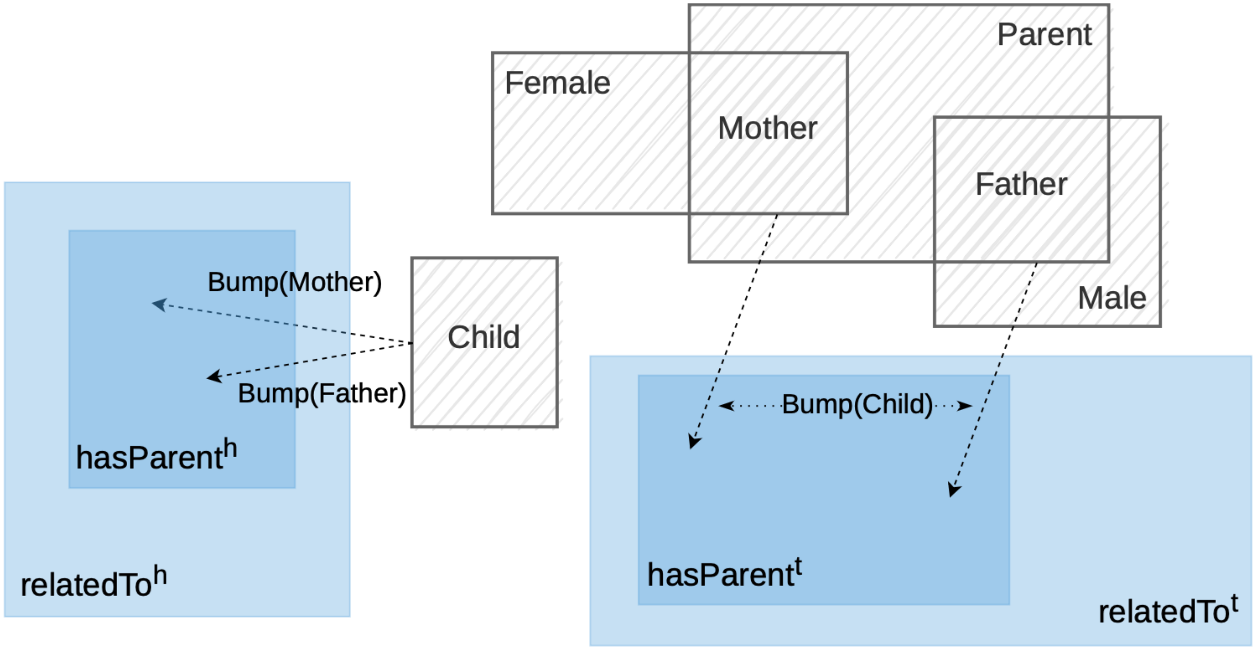


Subsumption, Conjunction, Membership

Ontology (Description Logic) Embedding

- Example: Box²EL (2024)
  - Individual / Concept: Point / Box
  - Relation: two boxes (corresponding to **h**ead & **t**ail)
  - Existential quantification: one translational **bump**ing vector for concept;
    - Modeling directional iteration between concepts

    e.g.,
    $Child \sqsubseteq \exists hasParent.Father$
    $Child \sqsubseteq \exists hasParent.Mother$



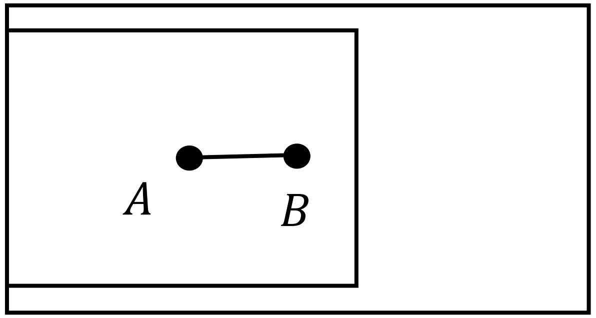
Representation of the family ontology in Box²EL

Jackermeier, Mathias, Jiaoyan Chen, and Ian Horrocks. "Dual box embeddings for the description logic EL++." *Proceedings of the ACM Web Conference 2024*. 2024.

Ontology (Description Logic) Embedding

- Box[2]EL training (embedding optimization):
  - Score/loss of **concept subsumption** (inclusion of two boxes):

$$\mathcal{L}_{\subseteq}(A,B) = \begin{cases} \|\max\{\mathbf{0},\ \boldsymbol{d}(A,B) + 2\boldsymbol{o}(A) - \gamma\}\| & \text{if } B \neq \emptyset \\ \max\{0, \boldsymbol{o}(A)_1 + 1\} & \text{otherwise,} \end{cases}$$



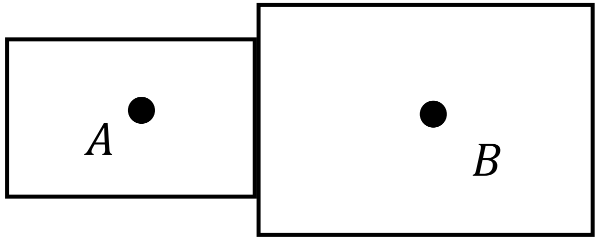
The horizontal loss is 0

Box A moves left, the horizontal loss > 0;

Box A moves right, the horizontal loss keeps 0

- Element-wise distance of two boxes:

$$\boldsymbol{d}(A,B) = |\boldsymbol{c}(A) - \boldsymbol{c}(B)| - \boldsymbol{o}(A) - \boldsymbol{o}(B)$$



The horizontal distance is 0

Ontology (Description Logic) Embedding

- Box²EL training: loess/scores for axioms of each normal form (NF)
  - NF1: $C \sqsubseteq D$ $\quad \mathcal{L}_1(C, D) = \mathcal{L}_{\subseteq}(\mathrm{Box}(C), \mathrm{Box}(D))$
  - NF2: $C \sqcap D \sqsubseteq E$ $\quad \mathcal{L}_2(C, D, E) = \mathcal{L}_{\subseteq}\Big(\mathrm{Box}(C) \cap \mathrm{Box}(D), \mathrm{Box}(E)\Big)$
  - NF3: $C \sqsubseteq \exists r.D$ $$\mathcal{L}_3(C, r, D) = \frac{1}{2}\Big(\mathcal{L}_{\subseteq}(\mathrm{Box}(C) + \mathrm{Bump}(D), \mathrm{Head}(\boldsymbol{r})) + \mathcal{L}_{\subseteq}(\mathrm{Box}(D) + \mathrm{Bump}(C), \mathrm{Tail}(\boldsymbol{r}))\Big).$$
  - NF4: $\exists r.C \sqsubseteq D$ $\quad \mathcal{L}_4(r, C, D) = \mathcal{L}_{\subseteq}(\mathrm{Head}(\boldsymbol{r}) - \mathrm{Bump}(C), \mathrm{Box}(D))$
  - NF5: $C \sqcap D \sqsubseteq \perp$ $\quad \mathcal{L}_5(C, D) = \|\max\{\mathbf{0}, -(\boldsymbol{d}(\mathrm{Box}(C), \mathrm{Box}(D)) + \gamma)\}\|$

Ontology (Description Logic) Embedding

- Box²EL training: Loess for axioms of each normal form (NF)

  - NF6: $r \sqsubseteq s$

$$\mathcal{L}_6(r, s) = \frac{1}{2}\Big(\mathcal{L}_{\subseteq}(\mathrm{Head}(r), \mathrm{Head}(s)) + \mathcal{L}_{\subseteq}(\mathrm{Tail}(r), \mathrm{Tail}(s))\Big)$$

  - NF7: $r_1 \circ r_2 \sqsubseteq s$

$$\mathcal{L}_7(r_1, r_2, s) = \frac{1}{2}\Big(\mathcal{L}_{\subseteq}(\mathrm{Head}(r_1), \mathrm{Head}(s)) + \mathcal{L}_{\subseteq}(\mathrm{Tail}(r_2), \mathrm{Tail}(s))\Big)$$

ABox can be transformed into TBox:

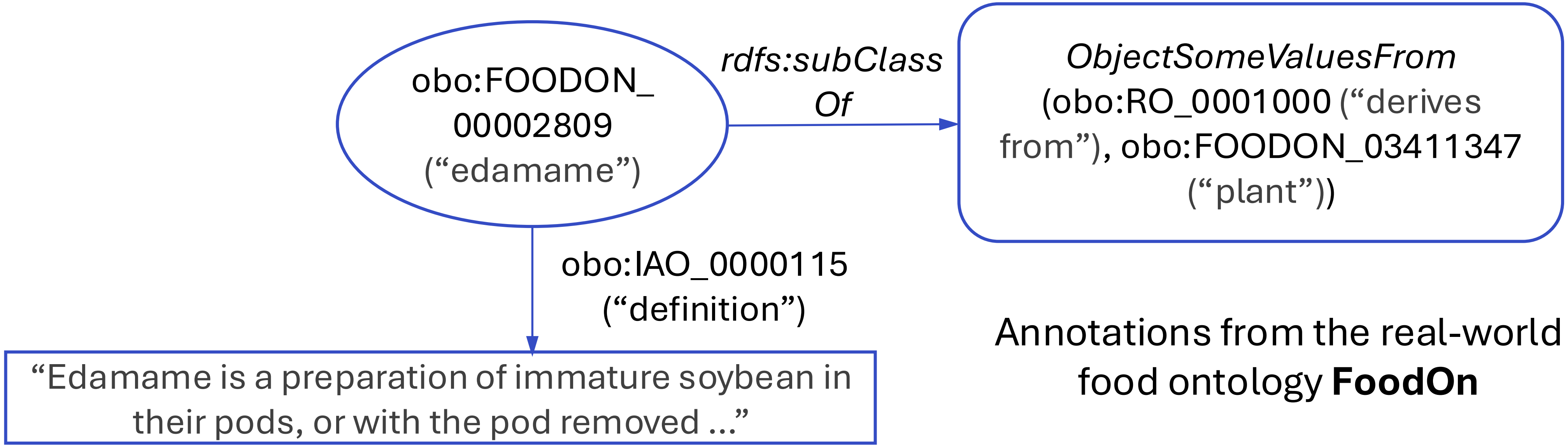
$$C(a) \rightsquigarrow \{a\} \sqsubseteq C$$
$$r(a, b) \rightsquigarrow \{a\} \sqsubseteq \exists r.\{b\}$$

Box^{2}EL Application & Evaluation

- Link prediction (Protein-protein interaction (PPI))
  - Gene Ontology + STRING Database of PPIs
  - Predict: $\{\mathrm{P}_1\} \sqsubseteq \exists \mathrm{interacts}.\{\mathrm{P}_2\}$ (P_1 and P_2 denotes two protein instances)
- Ontology subsumption prediction
  - Predict subsumption axioms of normal forms of $C \sqsubseteq D$, $C \sqcap D \sqsubseteq E$, $C \sqsubseteq \exists r.D$ and $\exists r.C \sqsubseteq D$
    - Rank the named concepts of being D, E, C, D in axioms of these forms, respectively
  - Ontologies of GALEN, Gene, Anatomy (20% and 10% masked for testing and validation, respectively)
- Ontology subsumption deductive inference (approximate)
- Outperforms four baselines: ELEm and EmEL++ (ball-based), and ELBE and BoxEL (box-based)

Text-aware Ontology Embedding

- OWL ontology includes more than formal semantics (e.g., labels, textual definitions)



Annotations from the real-world food ontology **FoodOn**

- How to jointly embed (and utilise) the informal text and the formally defined knowledge?

Text-aware Ontology Embedding

- Non-contextual language model
  - **OWL2Vec*** (2021): Train a **Word2Vec** model from an OWL ontology
    - Corpus (sentences) extraction via seriation, walking on the graph, OWL to RDF projection, etc.
- Contextual encoder: Transformer-based pre-trained language models (PLM)
  - **Task-specific fine-tuning**

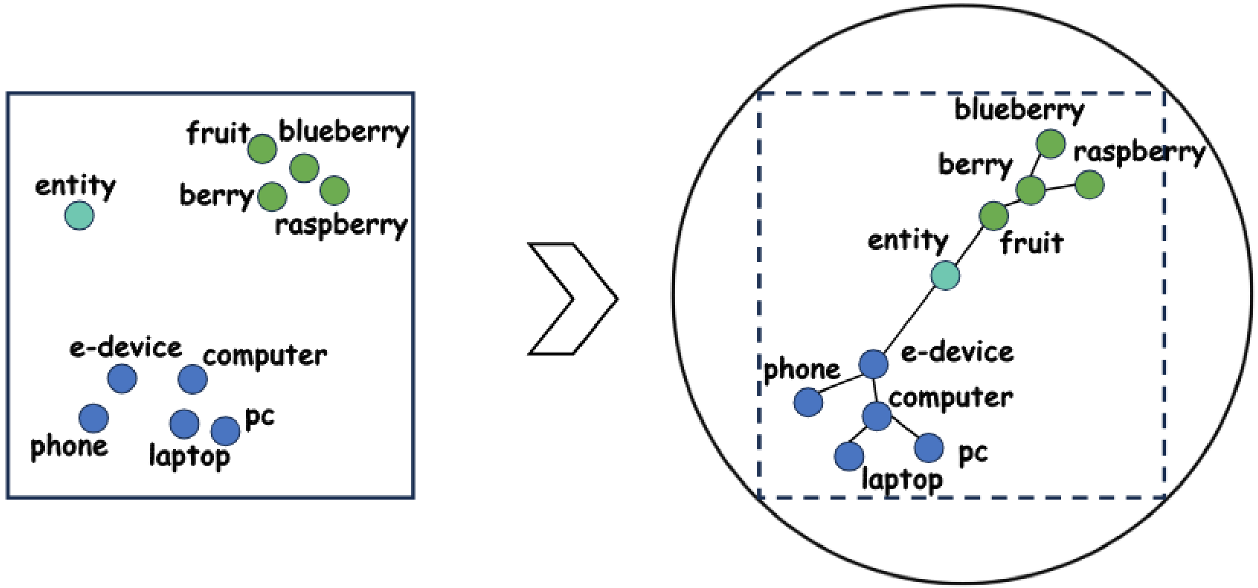
  E.g., **BERTMap** (2022): fine-tunes a PLM for ontology alignment (classify whether two concepts are equivalent)

These models do not preserve formal semantics
(cannot do inference with just the embeddings)

Chen, Jiaoyan, et al. "OWL2Vec*: embedding of OWL ontologies." *Machine Learning* 110.7 (2021): 1813-1845.
He, Yuan, et al. "BERTMap: a BERT-based ontology alignment system." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 5. 2022.

Text-aware Ontology Embedding

- Language model as hierarchy encoder (**HiT**)
  - **Re-train** a BERT alike PLM by an ontology
  - Force the PLM's concept embeddings to a hierarchy in a Poincaré ball (which is efficient for representing hierarchies)



Concept's text embedding in Euclidean Space by the last layer (tanh activation) of a **PLM**, which is in a d-dimensional hyper-cube

Concept's text embedding by an **ontology retrained PLM**, which is in a Poincaré Ball of radius $\sqrt{\mathrm{d}}$ that circumscribe the hyper cube

He, Yuan, et al. "Language models as hierarchy encoders." *Advances in Neural Information Processing Systems* 37 (2024): 14690-14711.

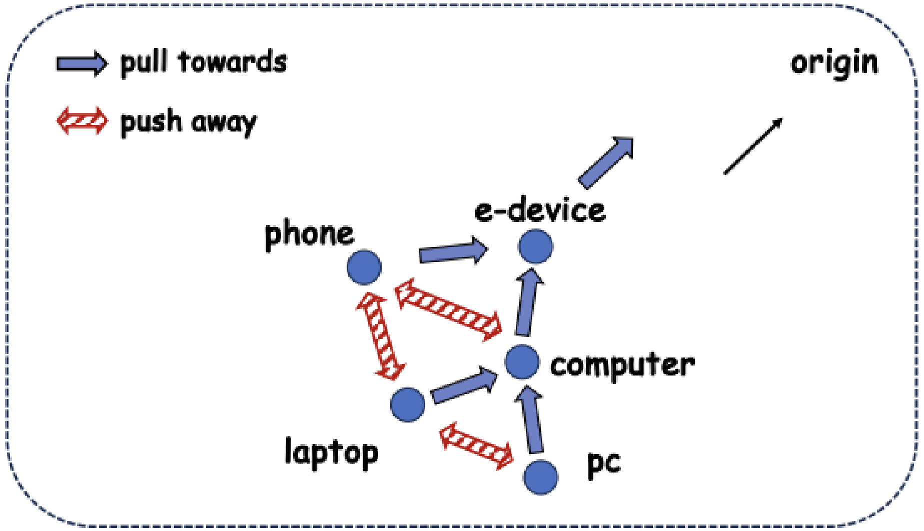
Text-aware Ontology Embedding

- Training of **HiT**
  - Hyperbolic clustering loss: clustering related concepts and distancing unrelated concepts (a concept e is close to its positive parent e^+ and distanced from its negative parent e^-)

$$\mathcal{L}_{cluster} = \sum_{(e,e^+,e^-)\in\mathcal{D}} \max(d_c(\mathbf{e},\mathbf{e}^+) - d_c(\mathbf{e},\mathbf{e}^-) + \alpha, 0)$$

  - Hyperbolic centripetal loss: make parent closer to origin

$$\mathcal{L}_{centri} = \sum_{(e,e^+,e^-)\in\mathcal{D}} \max(\|\mathbf{e}^+\|_c - \|\mathbf{e}\|_c + \beta, 0)$$



- Subsumption inference with HiT embeddings
  - Consider both contrastive and centripetal losses

$$s(e_1 \sqsubseteq e_2) = -(d_c(\mathbf{e}_1, \mathbf{e}_2) + \lambda(\|\mathbf{e}_2\|_c - \|\mathbf{e}_1\|_c))$$

Text-aware Ontology Embedding

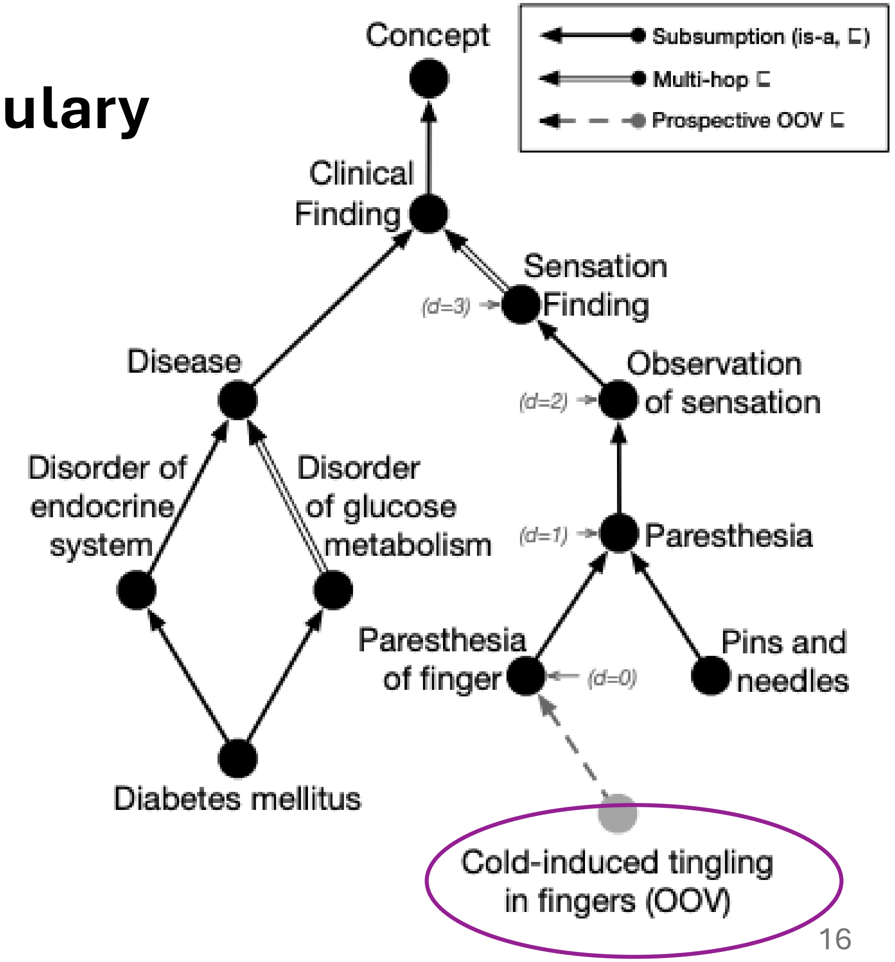
- LM as $\mathcal{EL}^{++}$ ontology encoder (**OnT**)
  - Extend HiT to complex concepts E.g., $\exists isParentOf.Person$
- Solutions: verbalization & relation by a rotation function for representing existential restriction

Yang, Hui, et al. "Language Models as Ontology Encoders." *The 24th International Semantic Web Conference (ISWC)*. 2025.

Discussion on Application

- HiT / OnT can be applied to subsumption prediction and inductive inference
  - Can predict subsumption for **out-of-vocabulary (OOV)** concepts using their labels
  - More transferable

- HiT / OnT for hierarchical retrieval
  - E.g., >80% queries to SNOMED CT are **OOV**, which require to return most close concepts with a subsumption relationship instead of equivalence



Dilworth, Jonathon, et al. "Hierarchical Retrieval with Out-Of-Vocabulary Queries: A Case Study on SNOMED CT." *arXiv preprint arXiv:2511.16698* (2025).

Part I: Knowledge Representation Learning

Part II: Ontology Reasoning with LLM Generation

LLMs for Ontology Reasoning

- Language models for encoding, e.g., HiT and OnT:
  - Short and usually binary answer for a **given candidate axiom** for prediction or deductive inference

- How about LLMs for generation (e.g., GPT, Llama)?
  - Subsumption (axiom) prediction / inference (e.g., prompting for “yes” or “no” answer) √
  - Proving – given a conclusion, can LLMs generate the proving procedure with the axioms, and answer **why**?

Yang, Hui, Jiaoyan Chen, and Uli Sattler. "Large Language Model for OWL Proofs." *Proceedings of the ACM Web Conference 2026*. 2026.

OWL Ontology Justification

Query: Why DomesticDog is subclass of CompanionAnimal?
(i.e., prove conclusion: DomesticDog $\sqsubseteq$ CompanionAnimal)

Relevant Web Ontologies

ontology.org/animal:
DomesticDog $\sqsubseteq$ Mammal
Mammal $\sqsubseteq$ Animal
. . .

petcare.com/taxonomy:
CompanionAnimal $\equiv$ Animal $\sqcap$ $\exists$hasOwner.Human
Fish $\sqsubseteq$ $\exists$livesIn. AquaticEnvironment
. . .

biology.edu/species:
DomesticDog $\equiv$ $\exists$hasOwner. Human $\sqcap$ Dog
Bird $\sqsubseteq$ $\exists$canPerform. Flying
. . .

Task 1: Extraction

Extract relevant ontological axioms:

- DomesticDog $\sqsubseteq$ Mammal (from ontology.org)
- CompanionAnimal $\equiv$ Animal $\sqcap$ $\exists$hasOwner.Human (petcare.com)
- DomesticDog $\equiv$ $\exists$hasOwner. Human $\sqcap$ Dog (from biology.edu)

Task 2: Simplification

Simplify axioms of complex ontological expressions:

- CompanionAnimal $\equiv$ Animal $\sqcap$ $\exists$hasOwner.Human
  $\rightarrow$ Animal $\sqcap$ $\exists$hasOwner.Human $\sqsubseteq$ CompanionAnimal
- DomesticDog $\equiv$ $\exists$hasOwner. Human $\sqcap$ Dog
  $\rightarrow$ DomesticDog $\sqsubseteq$ $\exists$hasOwner.Human

Task 3: Explanation

Describe reasoning process:

Step 1: DomesticDog $\sqsubseteq$ Animal $\sqcap$ $\exists$hasOwner.Human
BECAUSE DomesticDog $\sqsubseteq$ Mammal $\sqsubseteq$ Animal **AND**
DomesticDog $\sqsubseteq$ $\exists$hasOwner.Human

Step 2: DomesticDog $\sqsubseteq$ CompanionAnimal
BECAUSE Animal $\sqcap$ $\exists$hasOwner.Human $\sqsubseteq$ CompanionAnimal **AND** Step 1

Question: Can LLMs address Tasks 1-3?

End-to-end Example

LLM Output →

```
**AXIOMS_USED**: 1,2,3,4,5
**SIMPLIFY**:
[1] A ≡ ∃r.B  ⟶ ∃r.B ⊑ A
[2] C ⊑ B ⊓ H ⊓ I  ⟶ C ⊑ B
[3] D ≡ ∃r.C ⊓ G ⟶ D ⊑ ∃r.C
[4] E ≡ A ⊓ F ⟶ A ⊓ F ⊑ E
[5] D ⊑ F ⊓ J ⟶ D ⊑ F
**DERIVE**:
STEP1: [1,2,3] ⊢ D ⊑ A
EXPLANATION: D ⊑ ∃r.C ∧ C ⊑ B ⇒ D ⊑ ∃r.B ∧ ∃r.B ⊑ A
⇒ D ⊑ A
STEP2: [STEP1,4,5] ⊢ D ⊑ E
EXPLANATION: D ⊑ A ∧ D ⊑ F ⇒ D ⊑ A ⊓ F ∧ A ⊓ F ⊑
E ⇒ D ⊑ E
```

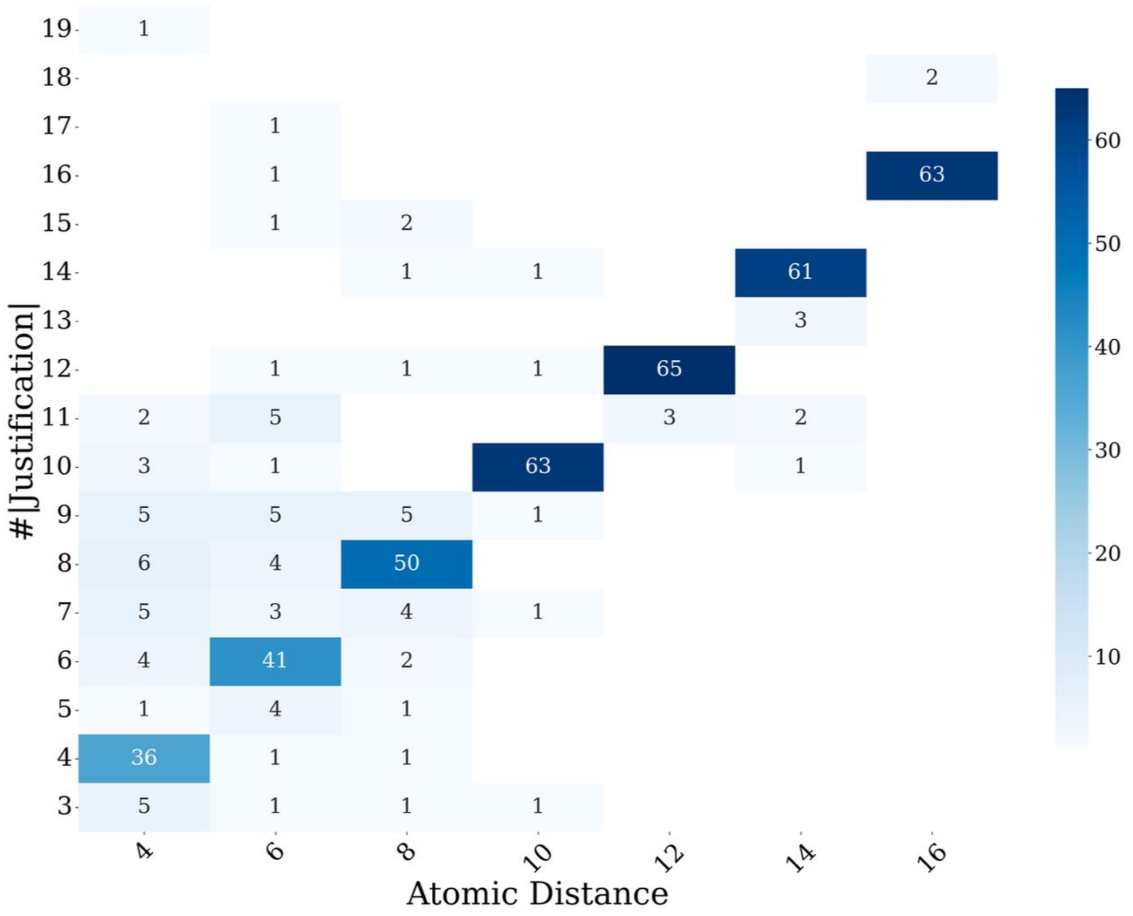
Input to LLM ↓

```
Given the following axioms:
[1] A ≡ ∃r.B
[2] C ⊑ B ⊓ H ⊓ I
[3] D ≡∃r.C ⊓ G
[4] E ≡ A ⊓ F
[5] D ⊑ F ⊓ J
[6] D ⊑ K
The desired explanation for deriving the conclusion D
⊑ E is as follows:
```

Dataset Construction

- Sample conclusions (subsumption axioms of ranging ***atomic distances***), and their golden answers (justifications)

  For an atomic subsumption $\mathcal{O} \models \mathrm{A} \sqsubseteq \mathrm{B}$, its **atomic distance** is defined as the minimal number $\boldsymbol{k}$ such that there exists a chain:
  $\mathrm{A} \sqsubseteq \mathrm{A}_1, \mathrm{A}_1 \sqsubseteq \mathrm{A}_2, \ldots, \mathrm{A}_{k-1} \sqsubseteq \mathrm{B}$



Distribution of sampled conclusions:
SNOMED CT (350), FoodOn (70), GO-Plus (70)

MANCHESTER 1824
The University of Manchester

Tested Models and Performance

- Basic setting: formal logic input, relevant vs irrelevant axioms (1:1)

Model	Format-Correct	Jaccard Avg.
GPT-o4-mini	83.43 / 92.26	99.19 / 99.66
DeepSeek_R1_Llama-8B	58.86 / 60.86	68.18 / 66.21
DeepSeek_R1_Qwen-32B	97.43 / 86.86	42.23 / 35.63
DeepSeek_R1_Qwen-8B	15.71 / 16.57	52.63 / 49.34
Qwen2.5-32B	99.71 / 100.00	39.81 / 39.41
Qwen3-32B	73.71 / 29.71	95.87 / 82.28
Qwen3-8B	16.57 / 22.29	86.87 / 86.78
Magistral-Small-2506	33.71 / 29.14	84.77 / 81.55

Output Format & Task 1: Extraction

Results on SNOMED CT

With / without inference rules

Inference Rules
1. Subsumption: if $A \sqsubseteq B$ and $B \sqsubseteq C$, then $A \sqsubseteq C$
2. Conjunction: if $A \sqsubseteq B$ and $A \sqsubseteq C$, then $A \sqsubseteq B \sqcap C$
3. Existential: ...
4. Equivalence: ...
5. ...

Observations on GO-PLUS
and Foodon are similar

MANCHESTER 1824
The University of Manchester

Tested Models and Performance

- Basic setting: formal logic input, relevant vs irrelevant axioms (1:1)

	Simplification Acc.			Derivation Acc.		
	axiom-wise	overall	length-drop	step-wise	overall	#|Steps|
GPT-o4-mini	98.66 / 97.97	91.78 / 91.93	27.04 / 26.55	97.21 / 93.06	93.49 / 86.02	7.86 / 7.02
DeepSeek_R1_Llama-8B	78.57 / 76.14	34.58 / 29.78	23.85 / 24.90	49.39 / 51.48	14.02 / 14.22	4.59 / 4.20
DeepSeek_R1_Qwen-32B	96.55 / 93.25	8.00 / 6.02	33.29 / 40.21	53.39 / 58.77	7.71 / 7.16	4.89 / 4.95
DeepSeek_R1_Qwen-8B	77.16 / 72.96	40.48 / 29.41	15.82 / 23.91	63.85 / 73.44	8.33 / 7.84	8.17 / 6.90
Qwen2.5-32B	93.26 / 92.37	5.71 / 6.29	20.33 / 25.57	60.97 / 58.80	6.57 / 7.14	7.07 / 5.68
Qwen3-32B	97.97 / 97.12	69.47 / 51.33	30.24 / 29.70	94.15 / 88.03	72.14 / 61.95	7.31 / 6.50
Qwen3-8B	94.32 / 94.72	67.19 / 66.67	25.92 / 31.64	88.71 / 88.64	51.56 / 50.62	9.83 / 7.17
Magistral-Small-2506	87.62 / 89.29	38.56 / 36.96	35.36 / 36.49	78.55 / 88.90	32.68 / 34.78	6.22 / 6.07

Task 2: Simplification & Task 3: Explanation (Derivation)

Results on SNOMED CT

On GO-PLUS and Foodon, GPT-4o-mini and Qwen3 have better overall accuracy

Vulnerability to Noisy Axioms

- Increase the ratio of irrelevant axioms from 1:1 to 1:20
- The results (SNOMED CT, GPT-o4-mini, without rules) drops by
  - ~33% on Format-Correct
  - ~21% on Task 1 Extraction – Jaccard (~47% weighted on Format-Correct)
  - ~14% on Task 2 Simplification – overall accuracy (~42% weighted on Format-Correct)
  - ~30% on Task 3 Explanation – overall accuracy (~53% weighted on Format-Correct)

Natural Language

- Besides formal logic language, we also feed natural language using verbalization from DeepOnto

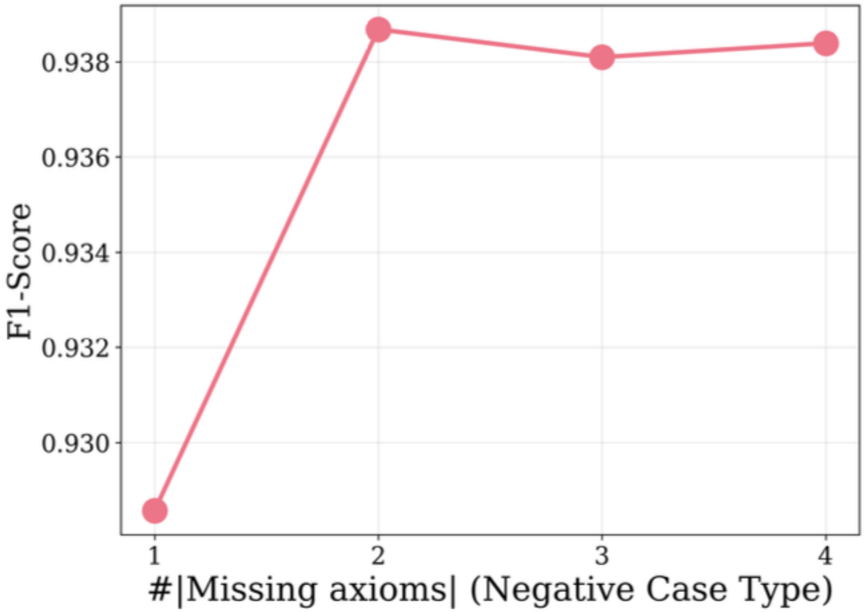
  E.g.,

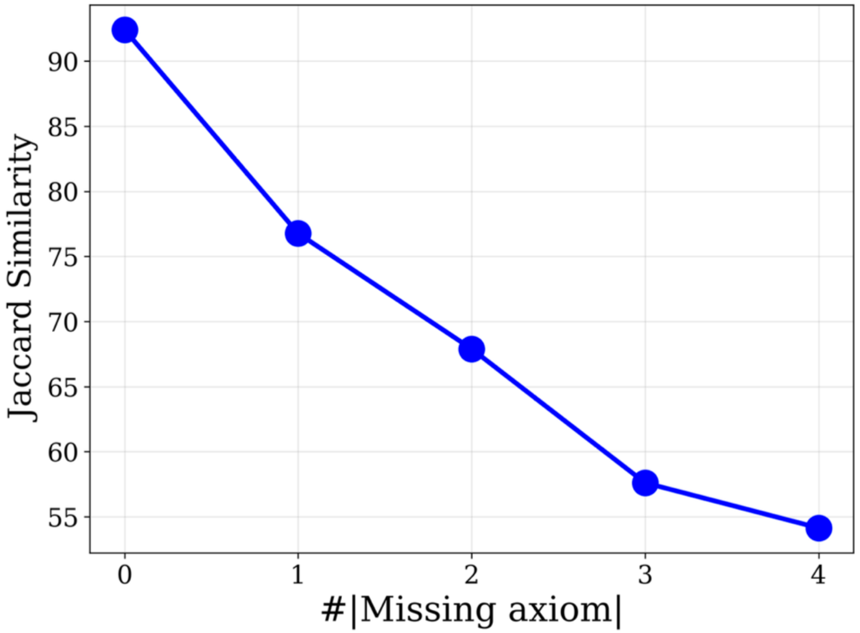
  - $\mathrm{DomesticDog} \sqsubseteq \mathrm{CompanionAnimal}$ (conclusion) → "Why is Domestic Dog a subclass of Companion Animal?"
  - $\exists \mathrm{hasOwner.Huamn}$ (justification axiom) → "something that has owner of some human"

- Results (SNOMED CT, GPT-o4-mini, without rules):
  - Only < 5% drops on Task 1 Extraction (< 8% weighted on Format-Correct)

Incompleteness

- What if axioms are missing in the premises?
  - LLMs detects such cases well (whether missing or not) (F1 ≈ 0.93)
  - Identifying the remaining useful axioms is hard
  - Task 1 Identification (Jaccard, format weighted, SNOMED CT, GPT-o4-mini) drops by ~38% when 4 axioms are missing





Discussion

- End-to-end OWL reasoning is challenging to LLMs
  - Considering the large scale of real-world ontologies (irrelevant axioms)
- Reasoning under incompleteness and uncertainty
  - Find all the relevant axioms, although not complete
  - Proving by completion with predicted axioms (abductive reasoning)
- Application to explainable Retrieval-Augmented Generation (like for question answering) with evidence reasoning

Summary

Part I: Knowledge Representation Learning

- Geometric Embedding (Box^{2}EL)
- Text-aware Embedding (OWL2Vec*, BERTMap, HiT & OnT)
- Applications (Link Prediction, Subsumption Prediction / Inference, Retrieval)

Part II: Ontology Reasoning with Generation

- Evaluating LLMs for proving – axiom extraction, simplification and explanation

- Acknowledgement
  - Other key contributors:
    - Yuan He (Oxford, now in Amazon), Hui Yang (Manchester), Mathias Jackermeier (Oxford), Jonathon Dilworth (Manchester), Ian Horrocks (Oxford)
  - Main funders (projects):
    - EPSRC (OntoEm & ConCur), Samsung Research UK (Personal KG)
- Other materials
  - DeepOnto (a library for ontology operation and LM-based curation; https://krr-oxford.github.io/DeepOnto/)
  - Ontology embedding survey (10.1109/TKDE.2025.3559023)
  - We will have a tutorial in ISWC'26 (Bari, Italy)
- Feel free to contact me
  - jiaoyan.chen@manchester.ac.uk
- Q&A