MANCHESTER 1824
The University of Manchester

Neural-symbolic Knowledge Representation with Ontology and Knowledge Graph Embeddings

Jiaoyan Chen

Lecturer in Data Intensive Systems (from 11/2022)

jiaoyan.chen@manchester.ac.uk

What is an ontology?

Formal, **explicit** and **shared** knowledge representation of a domain (e.g., concepts, instances and relationships)

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

A toy ontology on a family

Why do we use RDF, RDFS and OWL?

RDF: Resource Description Framework (<Subject, Predicate, Object>)

RDFS: RDF Schema

OWL: Web Ontology Language

Reason #1: Provide widely used vocabularies for defining all kinds of knowledge of ontologies (and Knowledge Graphs that will be mentioned later)

Why do we use RDF, RDFS, OWL?

Reason #2: OWL supports Description Logics

Examples of the toy family ontology:

$$Father \sqsubseteq Parent$$
$$Father \sqsubseteq Parent \sqcap Male$$
$$Child \sqsubseteq \exists hasParent.Father$$

Real-world example: ‘food material’ ≡ ‘environmental material’ *and* (‘has role’ *some* ‘food’)

i.e., $food\ material \equiv environmental\ material \sqcap \exists hasRole.food$

Why do we use RDF, RDFS and OWL?

Reason #3: already have been widely deployed

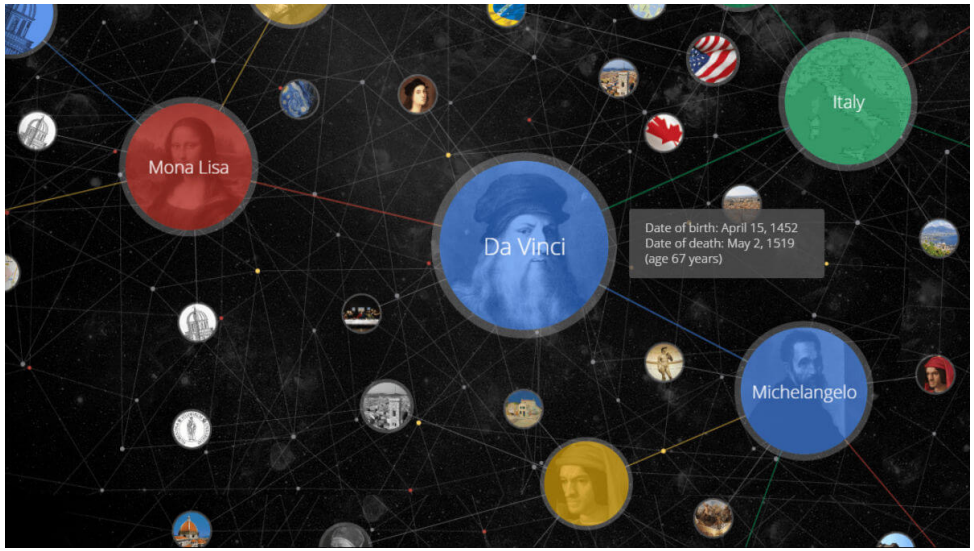
E.g., in life sciences:

SNOMED Clinical Terms,
The Gene Ontology (GO)
The Food Ontology (FoodOn)
Human Disease Ontology (DOID)
The Orphanet Rare Disease ontology (ORDO)
…

Chen, J., et al. "Knowledge Graphs for the Life Sciences: Recent Developments, Challenges and Opportunities." *Transactions on Graph Data and Knowledge (TGDK)* (2023).

What is Knowledge Graph?

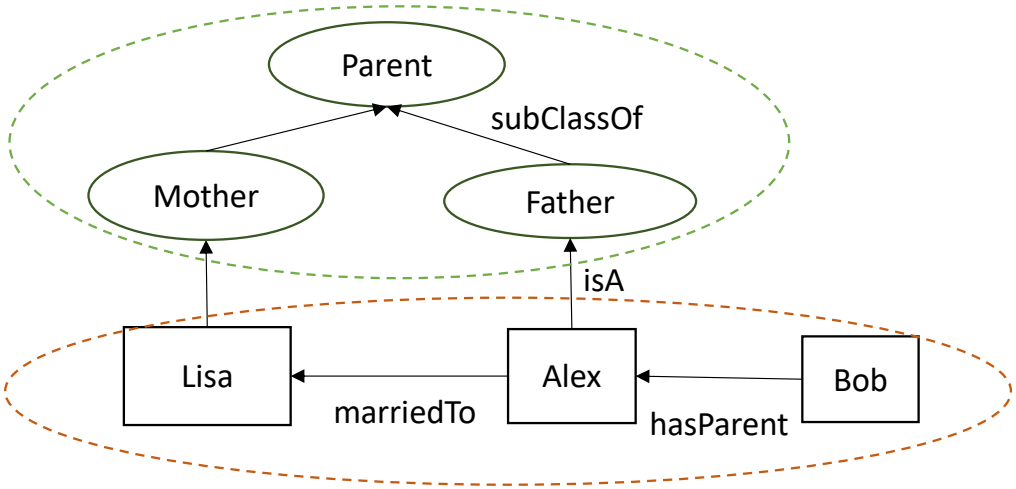
- “Knowledge Graph” was proposed by Google in 2012, referring to its services to enhance its search engine’s results with knowledge gathered from a variety of sources



- Knowledge ≈ Instances + Facts, represented as RDF triples e.g., <Box, hasParent, Alex>
- Linked and graph structured data

Knowledge Graph vs Ontology

- *Database perspective*: ontology as the schema of Knowledge Graph



- *Knowledge Representation Perspective*: an OWL ontology can be regarded as the composition of a TBox (terminologies) and an ABox (assertions); a Knowledge Graph can be understood as (a part) of the **ABox**

Many researchers prefer to not clearly distinguish these two terms

(Ontology sometimes is regarded as logic equipped Knowledge Graph)

Ontology and Knowledge Graph Embedding

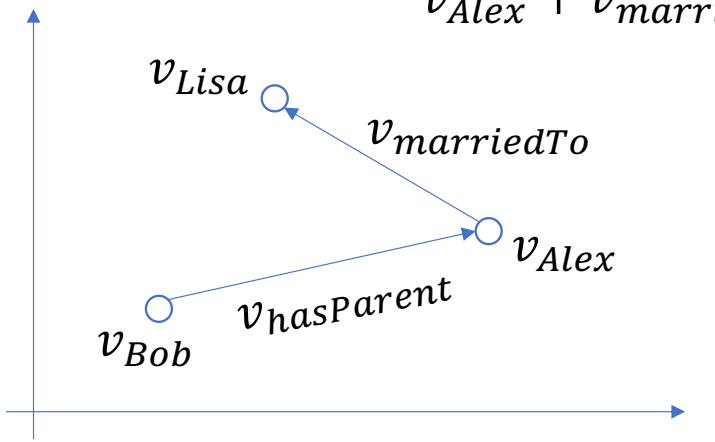
- To represent symbols (e.g., entities and relations) in a vector space with their relationships concerned, mainly for being consumed by statistical analysis and machine learning

$$v_{Bob} + v_{hasParent} \approx v_{Alex}$$
$$v_{Alex} + v_{marriedTo} \approx v_{Lisa}$$
$$v_i \in R^n$$

Example: **TransE**

<Bob, hasParent, Alex>
<Alex, marriedTo, Lisa>
…

Learning algorithms



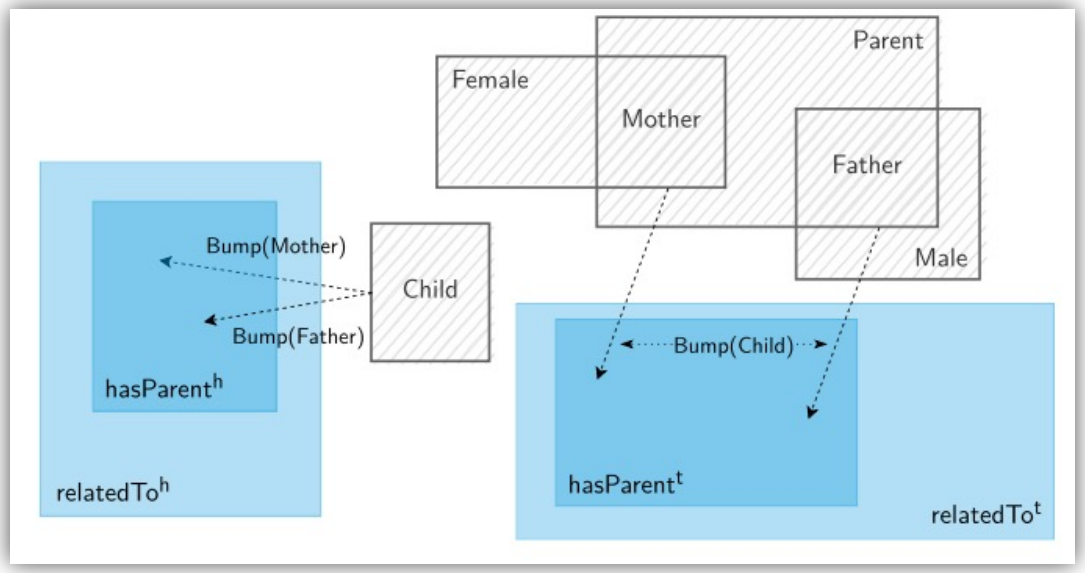
Bordes, A., et al. "Translating embeddings for modeling multi-relational data." *Advances in neural information processing systems* 26 (2013).

Embedding OWL Ontologies

$\mathcal{T}$ = {Father $\sqsubseteq$ Parent $\sqcap$ Male, Mother $\sqsubseteq$ Parent $\sqcap$ Female,
Child $\sqsubseteq \exists$hasParent.Father, Child $\sqsubseteq \exists$hasParent.Mother,
hasParent $\sqsubseteq$ relatedTo}

$\mathcal{A}$ = {Father(Alex), Child(Bob), hasParent(Bob, Alex)}

Learning Algorithms



The family ontology (OWL 2 EL). How to embed such an OWL ontology?

Box^{2}EL for OWL ontologies of Description Logic $\mathcal{EL}^{++}$

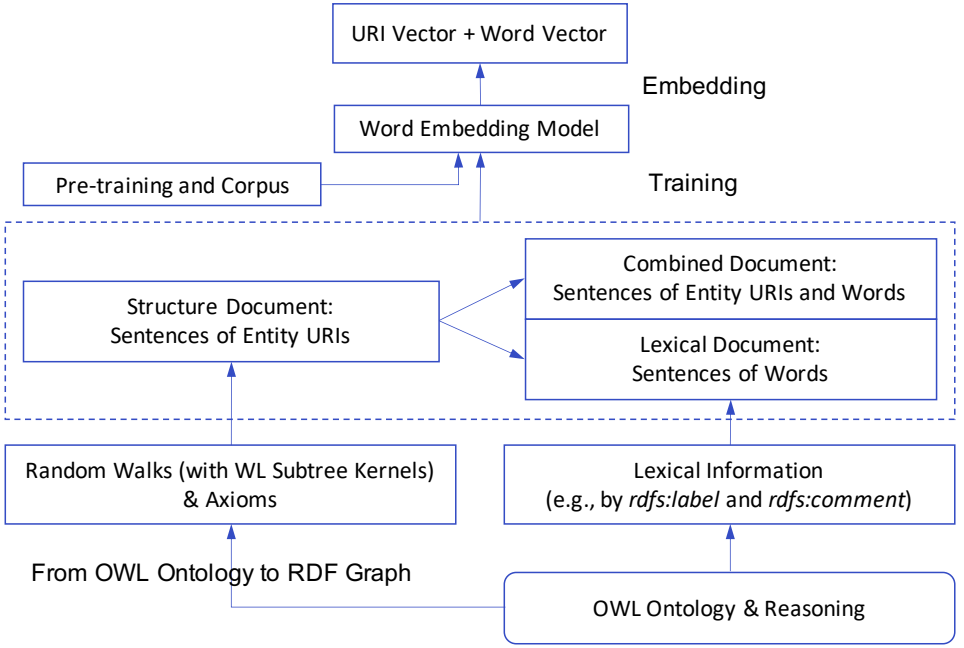
Concept (Instance): Box (Point)
Relation: Head Box & Tail Box

$C \sqsubseteq \exists r.D$: Box(C) $\otimes$ Bump(D) $\subseteq$ Head(r)
Box(D) $\otimes$ Bump(C) $\subseteq$ Tail(r)

Jackermeier, M., Chen, J., Horrocks, I.,"Dual Box Embeddings for the Description Logics EL++." The Web Conference 2024.

A General Ontology Embedding Tool **OWL2Vec***

- A **pipeline**
  - OWL axioms to RDF Graph
  - Corpora (sequences) extracted with literals (e.g., names and descriptions) and entities
  - Train Word2Vec embeddings over the corpora
- Embed **correlations** between entities and textual literals



Chen J., et al. "OWL2Vec*: Embedding of OWL ontologies." *Machine Learning* 110.7 (2021): 1813-1845.

Applications of the Embeddings

- Link Prediction / Recommendation
  - E.g., protein function prediction, ecotoxicological effect prediction
- Knowledge Engineering
  - E.g., entity alignment, completion
- Knowledge Retrieval
  - E.g., question answering
- Augmenting Machine Learning
  - E.g., injecting external knowledge of classes for zero-shot learning

Myklebust, Erik B., et al. "Prediction of adverse biological effects of chemicals using knowledge graph embeddings." *Semantic Web* 13.3 (2022): 299-338.

Chen, J, et al. "Zero-Shot and Few-Shot Learning With Knowledge Graphs: A Comprehensive Survey." Proceedings of the IEEE (2023).

Challenges and Opportunities from Large Language Models

- LLMs for knowledge graph, ontology and knowledge engineering
- Knowledge graph & ontology for LLMs

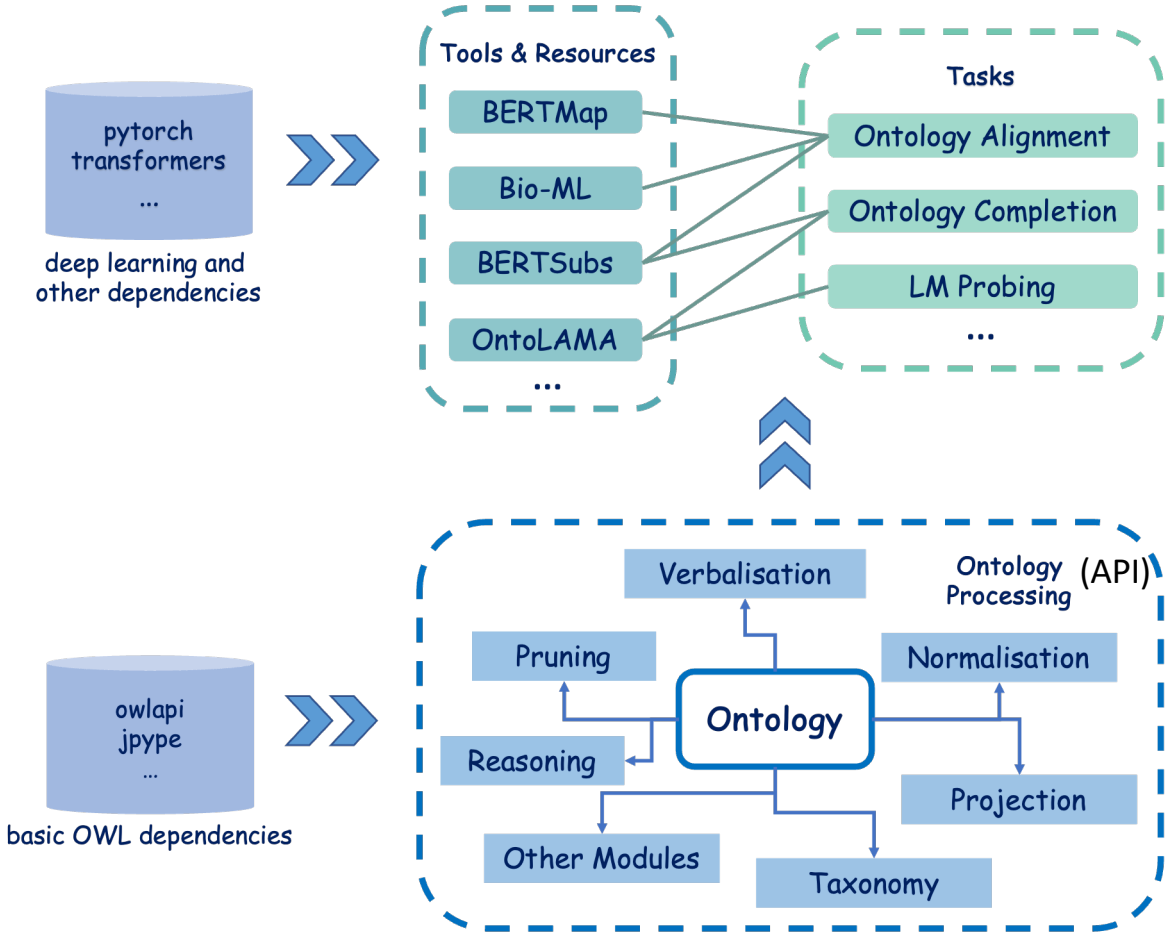
Pan, J., et al. "Large Language Models and Knowledge Graphs: Opportunities and Challenges." *Transactions on Graph Data and Knowledge* (2023).

An LM-based Ontology Engineering Library

DeepOnto https://github.com/KRR-Oxford/DeepOnto

- **Python interface** for more compact interaction with deep learning libraries (still call Java OWL API in the backend);
- **Ontology processing APIs** for fostering deep learning and NLP techniques in ontology engineering;
- **Ontology engineering tools and resources** implemented with our APIs, deep learning and (large) language models.

He, Y., et al. "DeepOnto: A Python package for ontology engineering with deep learning." *arXiv preprint arXiv:2307.03067* (2023).

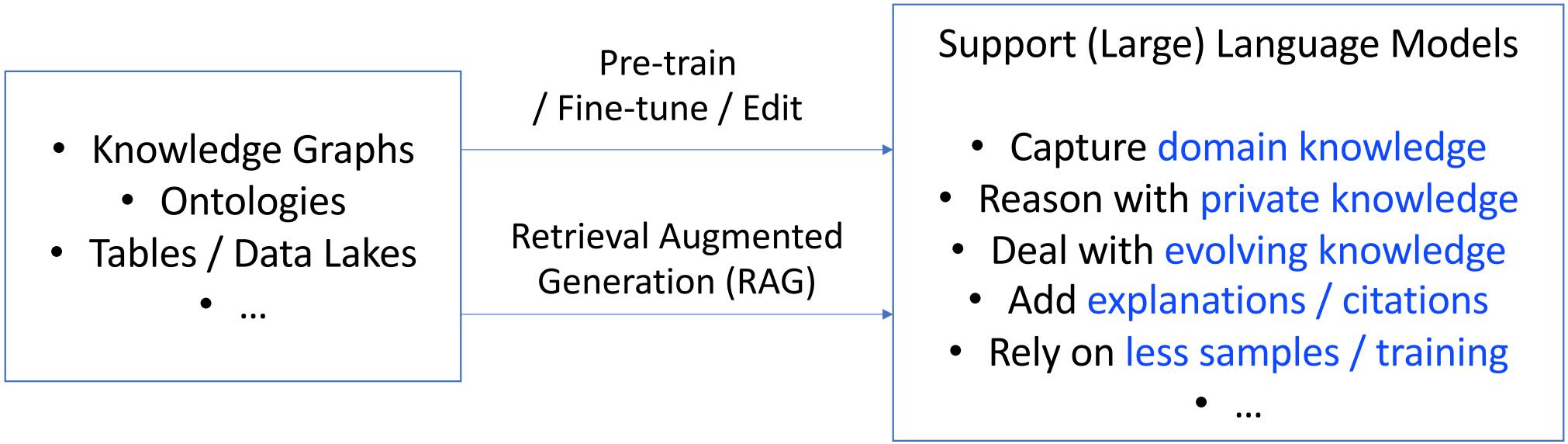


The Architecture of DeepOnto

Several tools implemented in DeepOnto

- **BERTMap: A BERT–Based Ontology Alignment System** by fine-tuning pre-trained language models (PLMs) by synonyms (AAAI 2022)
- **BERTSubs: ontology subsumption prediction** by prompts for encoding concept contexts and PLM fine-tuning (World Wide Web Journal 2023)
- **Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching** (ISWC 2022)
- **OntoLAMA: a Tool of Language Model Analysis** for Ontology Subsumption Inference (Findings of the ACL 2023)
- **ICON: taxonomy completion with missing common parents** (The Web Conference 2024)
- More in our TODO list; **External contributions are very welcomed**

How to augment LLMs?



Open Questions & Future Work

- DeepOnto for knowledge engineering: The research, development and impact acceleration
- What roles can knowledge representation/engineering play in the era of LLMs?
- Can knowledge embeddings support RAG and how?

Thanks for your attention