OAEI-2022 Bio-ML Track: ML-Friendly Biomedical Datasets for Equivalence and Subsumption OM

Yuan He[1], Jiaoyan Chen[1], Hang Dong[1], Ernesto Jiménez−Ruiz[2,3], Ian Horrocks[1]
1. Department of Computer Science, University of Oxford; 2. City, University of London; 3. SIRIUS, University of Oslo

Introduction

Ontology Matching

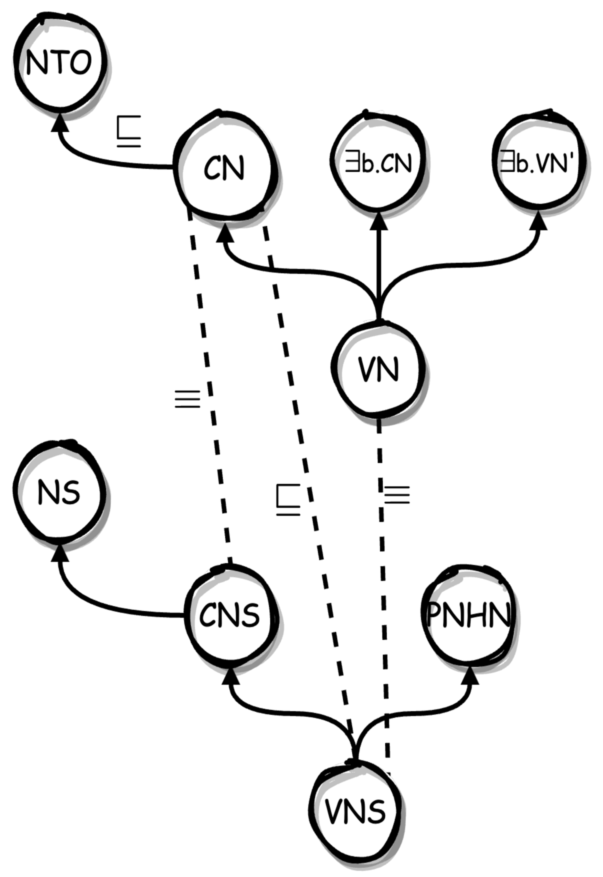
- Example of matching T-Box named concepts through equivalence and subsumption relationships

FMA Axioms

…
VestibulocochlearNerve ⊑ CanialNerve
VestibulocochlearNerve ⊑ ∃branch.CochlearNerve
VestibulocochlearNerve ⊑ ∃branch.VestibularNerve
CanialNerve ⊑ NeuralTreeOrgan
…

SNOMED Axioms

…
VestibulocochlearNerveStructure ⊑ CanialNerveStructure
VestibulocochlearNerveStructure ⊑ PeripheralNerveOfHeadAndNeck
CanialNerveStructure ⊑ NerveStructure
…



- Motivations: knowledge & data integration, quality assurance

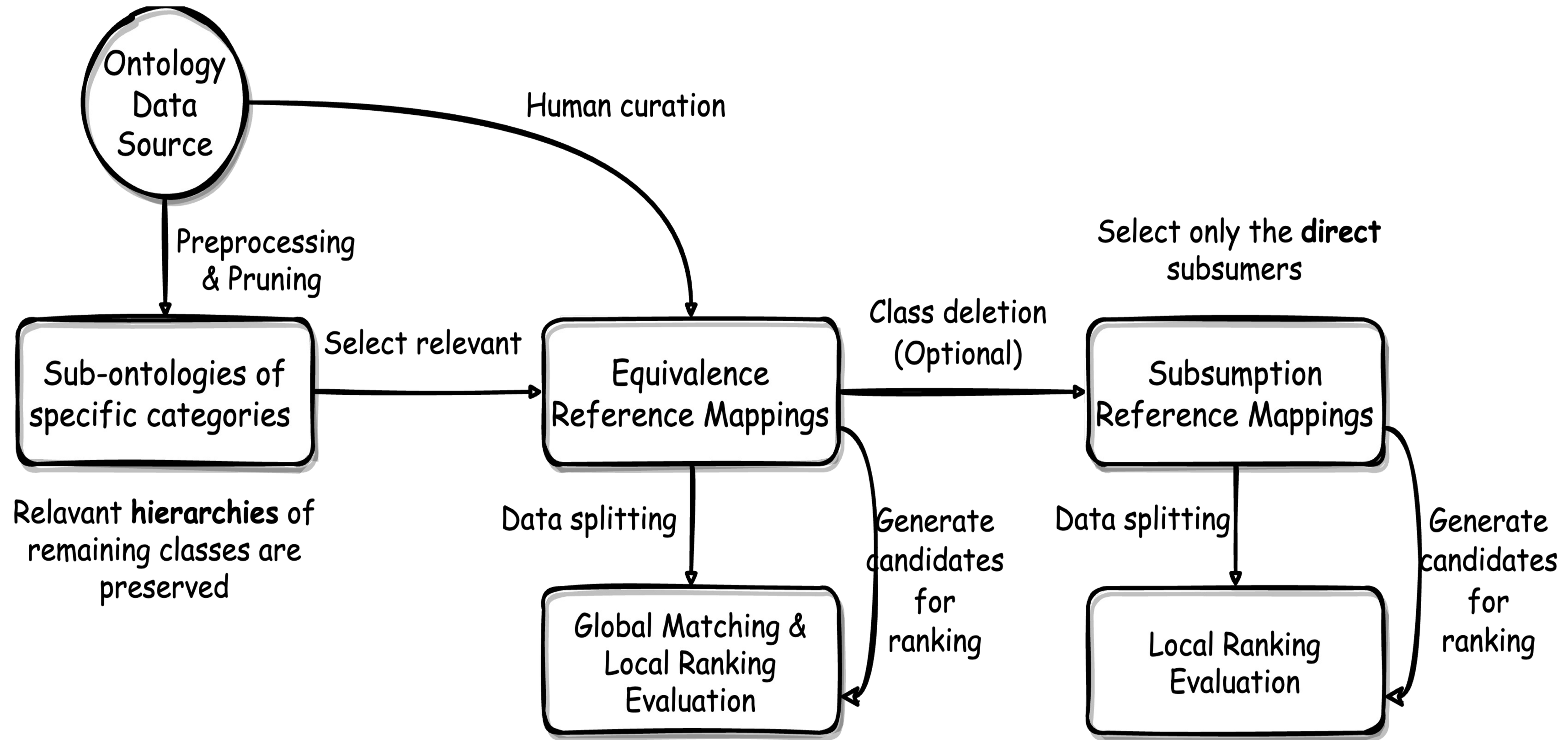
Challenges for OM systems

- **Variety** of naming scheme => synonyms; naming styles
- **Ambiguity** => similar naming but in different contexts
- **Scalability** => naïve traversal of OM takes O(n^2)
- **Inconsistency** => merging ontologies often lead to logical conflicts

Limitations of Existing OM Data

- Lack of **high-quality** gold standard mappings
- Lack of a **unified** evaluation framework
- Often **limited** to equivalence matching
- Lack of support for **machine learning-based** systems

Overall Workflow



Datasets and Settings

Source	Task	Category	#Classes	#RefMaps (equiv)	#Classes	#RefMaps (subs)
Mondo	OMIM-ORDO	Disease	9,642-8838	3,721	9,642-8,735	103
Mondo	NCIT-DOID	Disease	6,835-8,848	4,684	6,835-5,113	3,339
UMLS	SNOMED-FMA	Body	24,182-64,726	7,256	24,182-59,567	5,506
UMLS	SNOMED-NCIT	Pharm	16,045-15,250	5,803	16,045-12,462	4,225
UMLS	SNOMED-NCIT	Neoplas	11,271-13,956	3,804	11,271-13,790	213

Statistics for Equivalence Matching | Statistics for Subsumption Matching

- Matching: equivalence & subsumption
- Splitting: Unsupervised & semi-supervised
- Evaluation: global matching & local ranking

Participants & Results

Equivalence Matching

- ML-based systems including BERTMap (and -Lite), AMD, Matcha (and -DL)
- Traditional systems including LogMap (and -Lite), ATMatcher, LSMatcher
- ML systems generally perform better with Match-DL attaining best F1 on 4 out of 5 semi-supervised tasks, BERTMap (and lite) attains best F1 on 4 out of 5 unsupervised tasks, and best ranking scores of all tasks

Subsumption Matching

- ML-based systems including Word2Vec, OWL2Vec*, BERTSubs
  - BERTSubs performs the best on 2 out of 5 subsumption tasks, while OWL2Vec* performs the best on the remaining 3
- No participation of traditional systems

Full results: *https://www.cs.ox.ac.uk/isg/projects/ConCur/oaei/2022/#results*

Conclusion and Discussion

- Still too few participants using ML methods
- Only 3 participants on subsumption matching which is more challenging
- Some participants only submit the results
  - How to encourage both reproducibility and participation enthusiasm?
- A systematic benchmarking study on ML-based OM systems is required