Data ScienceTech Institute

DSTI at LLMs4OL 2024 Task A: Intrinsic versus extrinsic knowledge for type classification

Applications on WordNet and GeoNames datasets

Hanna Abi Akl

ISWC Baltimore, USA

UNIVERSITÉ CÔTE D'AZUR

wimmics Inria INVENTEURS DU MONDE NUMÉRIQUE

LLMs4OL@ISWC 2024

Content

- Task Introduction
- Dataset
- Domain Semantic Primitives
- Domain Semantic Towers
- Model
- Experimental Setup
- Results
- Conclusion

Task Introduction

- Task A (Term Typing): classification of lexical terms into categories (types)
- Formally:

$$[f_{inst}^{TaskA}(L) := [S?].\ ([L],\ [T]]$$

- **S** = optional context sentence; **L** = lexical term; **T** = concept term type
- *RQ1: How does external source knowledge fair against LLM intrinsic knowledge on lexical term typing?*
- *RQ2: How does external knowledge affect semantic grounding of LLMs?*

Dataset

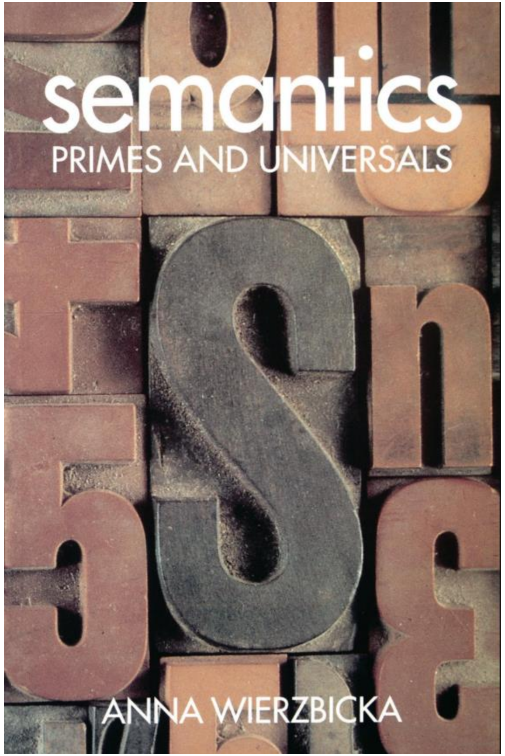
- Focus on WordNet and GeoNames datasets
- WordNet: **40,559** train terms and **9,470** test terms
- GeoNames: **8,078,865** train terms and **702,510** test terms
- WordNet types: *noun, verb, adjective, adverb*
- GeoNames: 660 geographical location types (e.g., lake, peak)

Lexical Term L	Sentence Containing L (Optional)	Type
question	there was a question about my training	noun
lodge	Where are you lodging in Paris?	verb
genus equisetum		noun

Lexical Term L	Type
Pic de Font Blanca	peak
Roc Mele	mountain
Estany de les Abelletes	lake

Domain Semantic Primitives

- Based on Wierzbicka's work on semantic primes and universals
- Domain-based instead of language-based
- Define semantic set *ST* for each domain as a list of *minimal semantic properties*
- Semantic properties fetched from Wikidata for each lexical term



ISWC Baltimore, USA

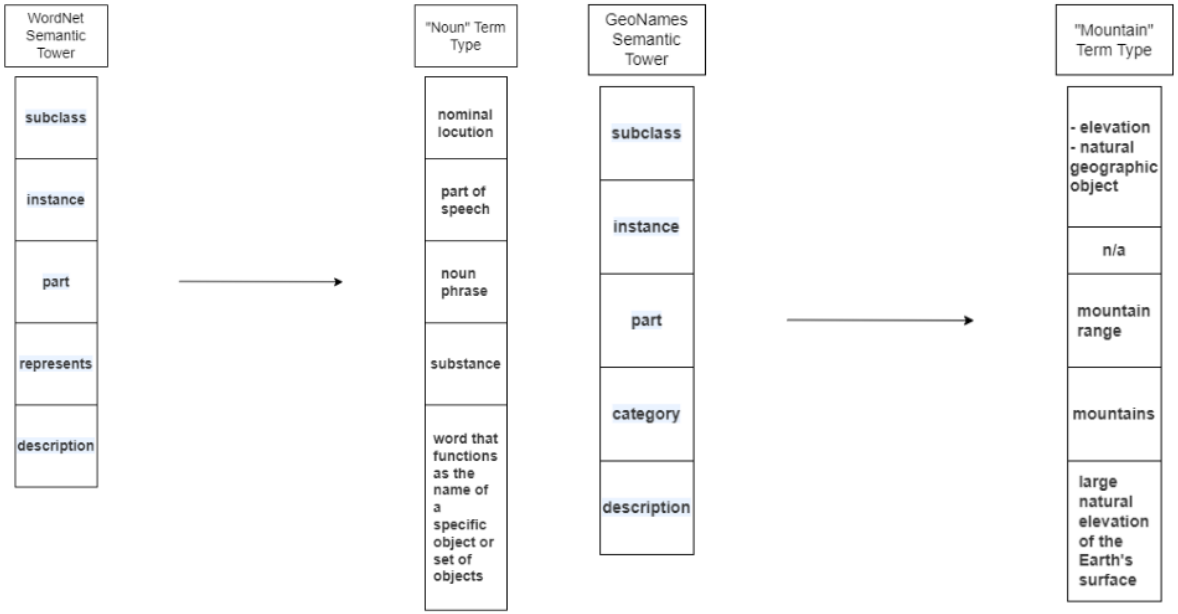
wimmics Inria INVENTEURS DU MONDE NUMÉRIQUE

UNIVERSITÉ CÔTE D'AZUR

DSTI

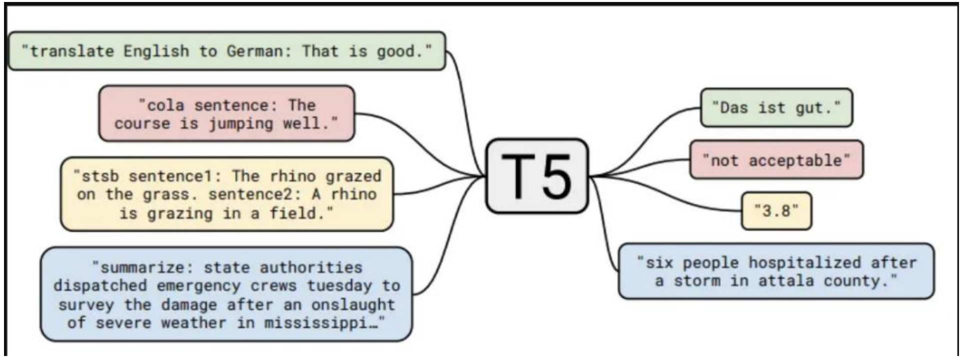
Domain Semantic Towers

- Preprocessing of semantic primitives (e.g., cleaning, tokenization, etc.)
- Transformation to vector embeddings using Google gte-large model
- Stored and indexed in vector store



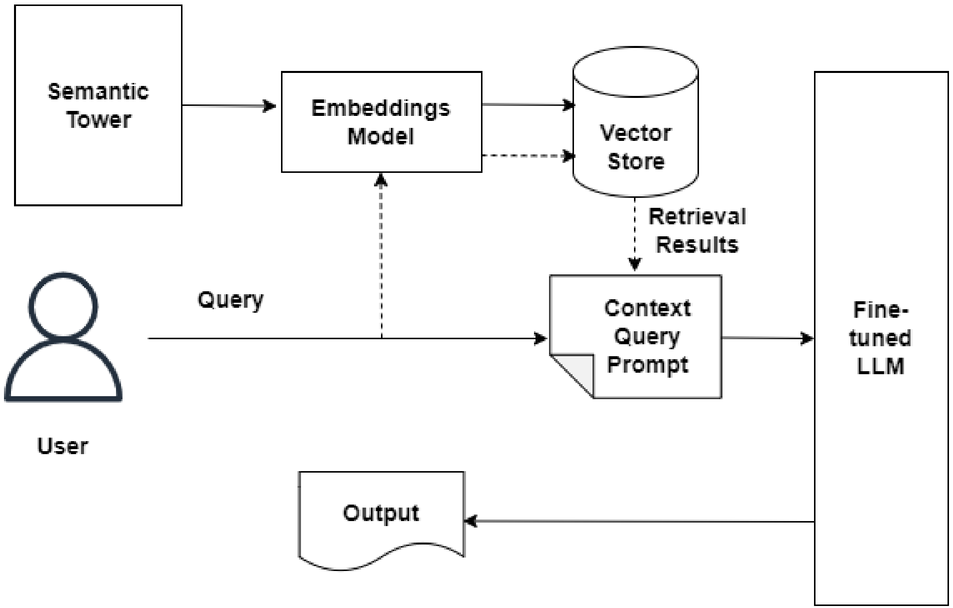
Model

- Flan-t5-small model fine-tuned on WordNet and GeoNames datasets
- Flan-t5-small-wordnet: 70% train, 30% validation
- Flan-t5-small-geonames: subset of very large dataset curated with 70% train, 30% validation
- Input: term **L** (+ context **S** when available) vectorized to vectors of size 1024 using gte-large



Experimental Setup

- 2 experiments per dataset (**WN1,WN2,GN1,GN2**)
- **WN1 & GN1:** prompting model on blind test set with instruction
- **WN2 & GN2:** RAG pipeline to find best type from vector store (semantic tower) and factor result in prompting instruction



Results (1)

- WN1 & GN1 better than WN2 & GN2
- Drop is consistent for both datasets
- WN: ST boosts detection of edge cases (e.g., *into the bargain* as *adverb*)
- GN: ST boosts plural type prediction (e.g., *peak* vs *peaks*) and nuances (e.g., *stream* vs *section of stream*)

Table 3. *Experimental results on the WordNet set.*

Experiment	F1
flan-t5-small-wordnet (WN1)	**0.9820**
flan-t5-small-wordnet + WordNet semantic tower (WN2)	0.8581

Table 4. *Experimental results on the GeoNames set.*

Experiment	F1
flan-t5-small-geonames (GN1)	**0.6820**
flan-t5-small-geonames + GeoNames semantic tower (GN2)	0.5636

Results (2)

- WN1: second place on few-shot testing
- WN1 & WN2: slight drop in performance of 1%
- Systems are sound and show no catastrophic drift
- GN1 & GN2 not submitted due to lack of resources for big few-shot set

Table 5. *Subtask A.1 (few-shot) WordNet term typing leaderboard.*

Teal Name	F1	Precision	Recall
TSOTSALearning	0.9938	0.9938	0.9938
DSTI (WN1)	**0.9716**	**0.9716**	**0.9716**
DaseLab	0.9697	0.9689	0.9704
RWTH-DBIS	0.9446	0.9446	0.9446
TheGhost	0.9392	0.9389	0.9395
Silp_nlp	0.9037	0.9037	0.9037
DSTI (WN2)	**0.8420**	**0.8420**	**0.8420**
Phoenixes	0.8158	0.7689	0.8687

ISWC Baltimore, USA

wimmics Inria INVENTEURS DU MONDE NUMÉRIQUE

UNIVERSITÉ CÔTE D'AZUR

DSTI

Conclusion

- *RQ1: How does external source knowledge fair against LLM intrinsic knowledge on lexical term typing?* **External knowledge sources offer an interesting trade-off between performance and semantic resonance**
- *RQ2: How does external knowledge affect semantic grounding of LLMs?* **External knowledge sources like Semantic Towers can be an important step in controlling fine-grained knowledge in LLM systems**
- Work is a springboard for more extensive research on knowledge representation for better semantic grounding

Thank You

Questions?

Feel free to contact at: hanna.abi-akl@dsti.institute