Term Typing using Large Language Models

DEBAENE Hugo, PRIN Paul

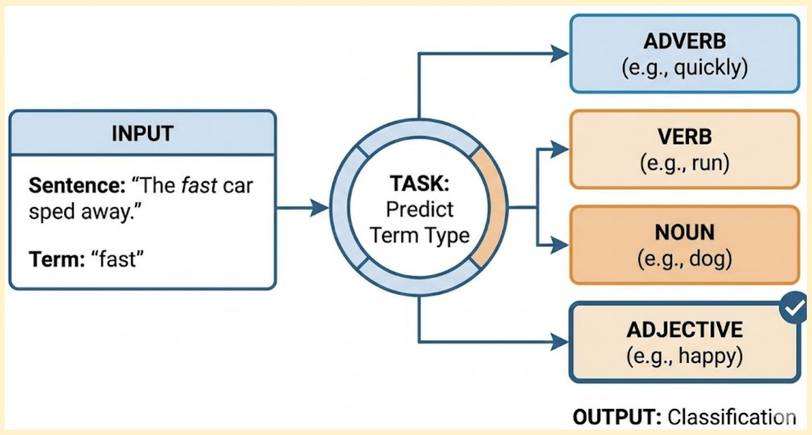
Télécom Paris, Institut Polytechnique de Paris, France

GitHub Code

LLMs4OL

1. Task

Term Typing consists in determining the type of a specific term in a given context. In this approach the types will be: *adjective, adverb, noun, verb* and the context will be a sentence or none.



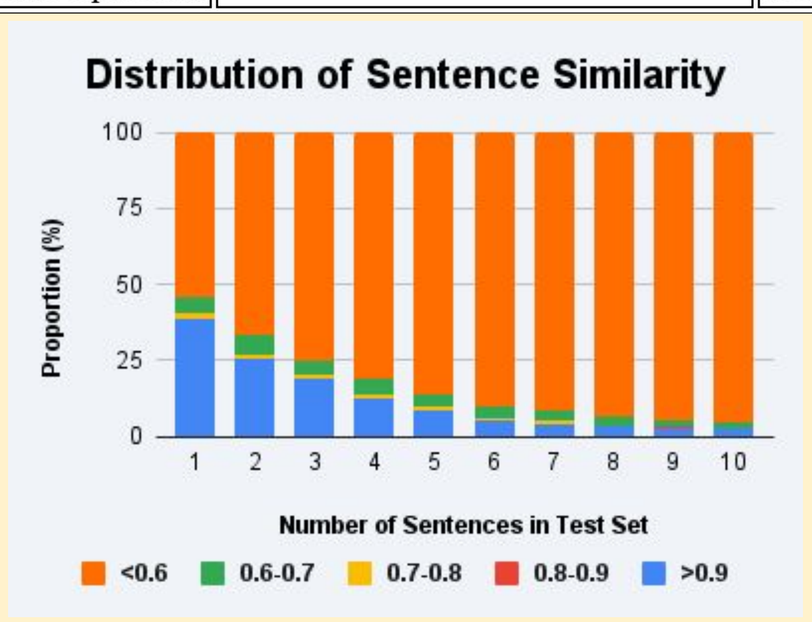
2. Dataset

- WordNet Characteristics:

Dataset	Train Size	Test Size	Number of Types
WordNet	40,559	9,470	4

- Dataset Content:

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



Above 3 sentences the retrieved context is "irrelevant"

3. Models used

Encoder-Decoder:
- Google Flan T5 Small
- Google Flan T5 Base
- Google Flan T5 Large

Causal:
- Qwen3 4B Instruct

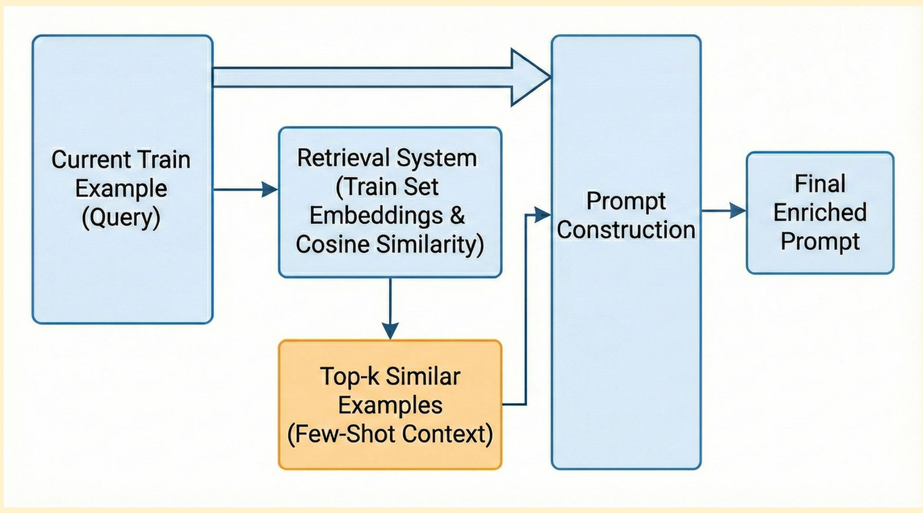
4. Original performances

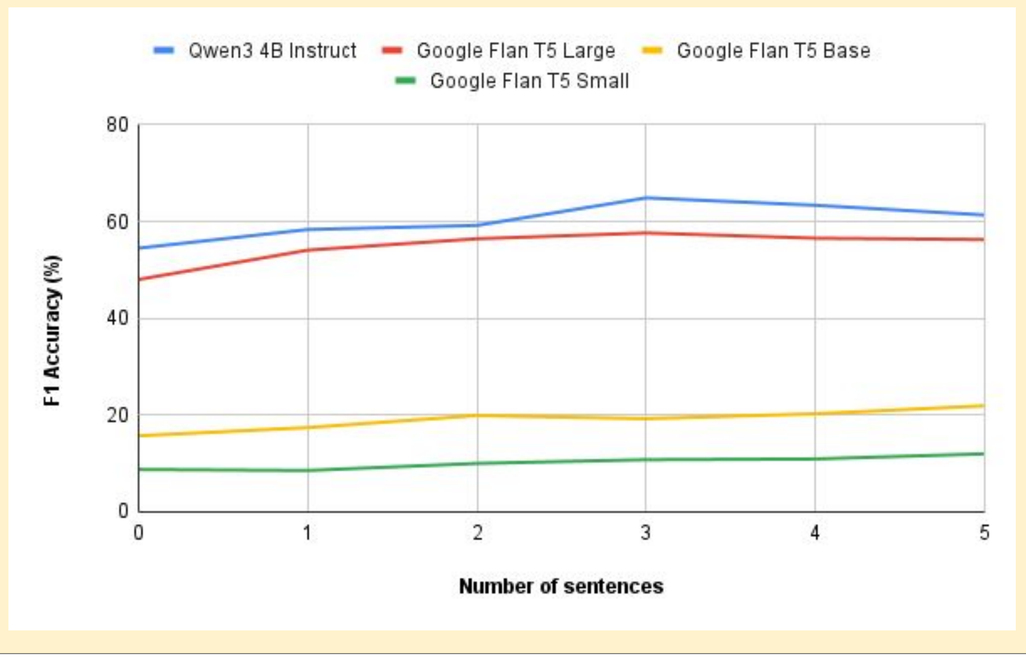
Prompt: Given the term … in the sentence …, what is the type of the term? Choose the term type from: adjective, noun, verb, adverb.

Model Name	Accuracy Score (%)
Qwen3 4B Instruct	54.50
Google Flan T5 Large	48.03
Google Flan T5 Base	15.73
Google Flan T5 Small	8.77

5. RAG

A *Few Shot Prompting / RAG* was done on the models adding to the prompts additional context retrieved by cosine similarity from the training set embedded with different context sizes.



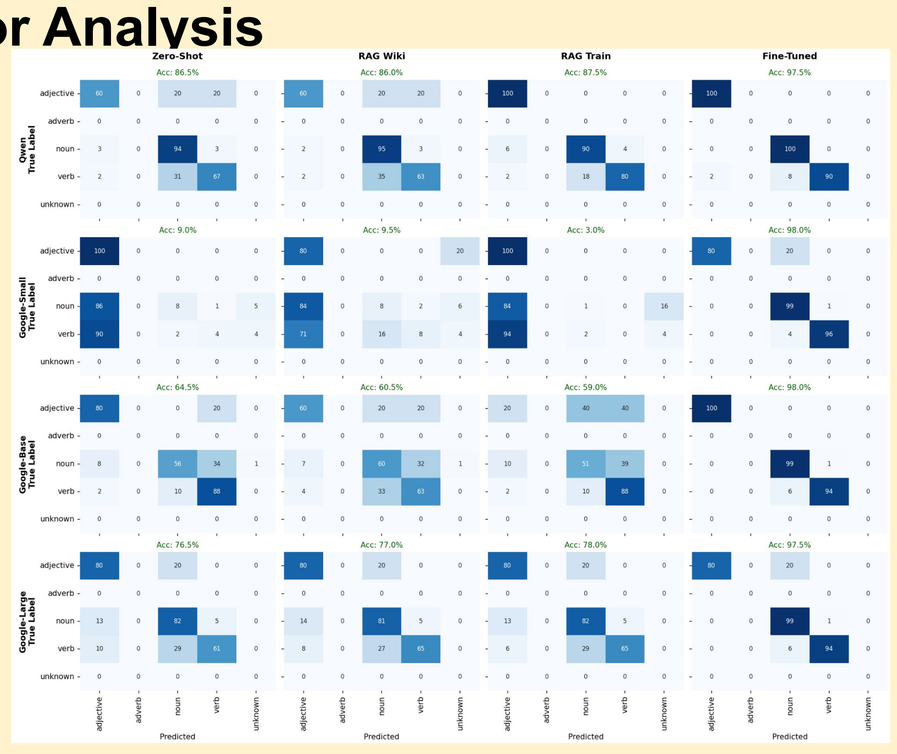


6. Fine Tuning

The models were fine tuned on 0.5% of their parameters (last layers) on the WordNet Training Set.

Model Name	Accuracy Score (%)
Qwen3 4B Instruct (Fine Tuned)	71.19
Google Flan T5 Large (Fine Tuned)	94.58
Google Flan T5 Base (Fine Tuned)	89.56
Google Flan T5 Small (Fine Tuned)	68.49

7. Error Analysis



8. Future Works

- Fine tune a bigger part of the models
- Fine tune ready for RAG
- Try on bigger models, state of the art large models

Fine tune the entire model
- other datasets