

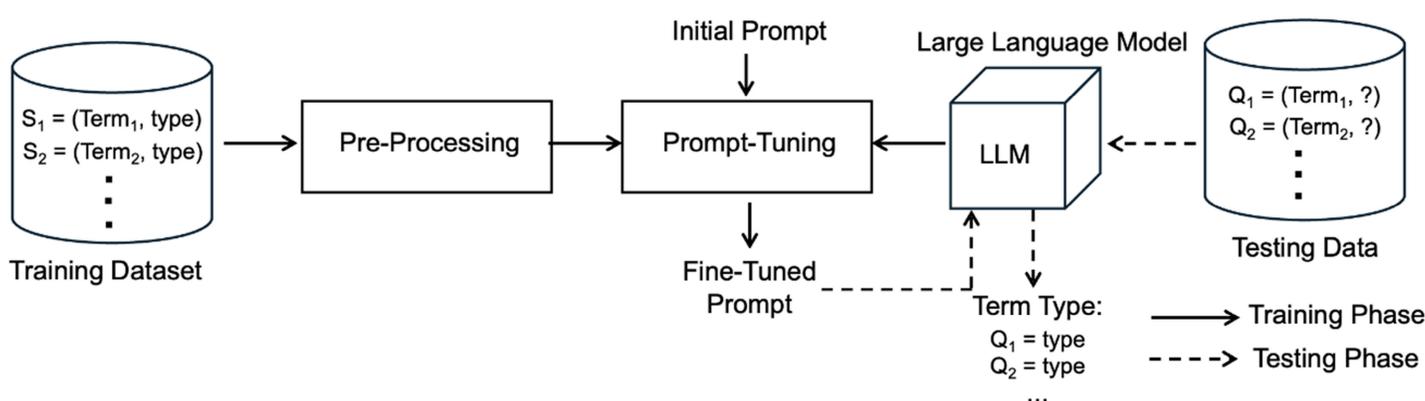
## 1 Abstract

The LLMs4OL Challenge @ ISWC 2024 aims to explore the intersection of Large Language Models (LLMs) and Ontology Learning (OL) through three main tasks: 1) Term Typing, 2) Taxonomy Discovery and 3) Non-Taxonomic Relation Extraction. In this project, we present our system's design for the term typing task. Our approach utilizes automatic prompt generation using soft prompts to enhance term typing accuracy and efficiency. We conducted experiments on several datasets, including WordNet, UMLS, GeoNames, NCI, MEDCIN, and SNOMEDCT US. Our approach outperformed the baselines on most datasets, except for GeoNames, where it faced challenges due to the complexity and specificity of this domain, resulting in substantially lower scores. Additionally, we report the overall results of our approach in this challenge, which highlight its promise while also indicating areas for further improvement.

## 2 Introduction

Currently, most data on the World Wide Web is in a format that humans can read and understand. However, computers require significant processing to interpret this data. To address this, the Semantic Web has been introduced as an extension of the World Wide Web, aiming to make internet data more interpretable and efficiently linked. This is achieved using Ontology, which models concepts within a specific domain. Traditionally, building Ontology is complex, time-consuming, and requires deep domain expertise. To simplify this process, Ontology Learning has been developed to automatically extract and structure information from unstructured data. With the rapid advancement of Large Language Models (LLMs), which have a deep understanding of language, it is crucial to explore and apply LLMs to support automatic Ontology creation.

## 3 Methodology



## 4 Technology



## 5 Result

Our approach outperforms the baseline in MAP@1 scores across WordNet, NCI, MEDCIN, and SNOMEDCTUS, improving term retrieval precision. However, GeoNames remains challenging due to place name ambiguity. Term typing results show strong performance on WordNet, NCI, MEDCIN, and SNOMEDCTUS, though recall is higher than precision, likely due to class imbalance.

Performance on GeoNames and Gene Ontology is lower, with challenges arising from name ambiguity and domain complexity. Larger datasets like Biological Process (792 classes) and GeoNames (680 classes) are harder to predict than smaller ones like WordNet (4 classes) or NCI (125 classes).

Table 1. MAP@1 Scores for Our Approach Compared to the Baseline Across Datasets

	WordNet	GeoNames	NCI	MEDCIN	SNOMEDCT_US
Baseline	0.9170	<b>0.4330</b>	0.3280	0.5180	0.4340
Our Approach	<b>0.9368</b>	0.3863	<b>0.6009</b>	<b>0.7397</b>	<b>0.6707</b>

Table 2. The result on term typing task

Dataset	F1	Precision	Recall
WordNet	0.9392	0.9389	0.9395
GeoNames	0.1489	0.1461	0.1519
NCI	0.5370	0.4450	0.6769
MEDCIN	0.5328	0.4183	0.7336
SNOMEDCT_US	0.5275	0.4266	0.6910
Cellular Component	0.1877	0.1653	0.2171
Biological Process	0.1025	0.0964	0.1095
Molecular Function	0.1270	0.1278	0.1261

## 6 Conclusion

This study explored soft prompt tuning for term typing in the LLMs4OL Challenge @ ISWC 2024. Strong performance was observed on WordNet and UMLS sub-datasets, but results on GeoNames and Gene Ontology were weaker due to class imbalance and domain complexity. Future improvements could involve adding contextual information and exploring techniques like Retrieval-Augmented Generation (RAG) to enhance prediction accuracy and knowledge access.

### References

- Babaei Giglou, H., D'Souza, J., & Auer, S. (2023, October). LLMs4OL: Large language models for ontology learning. In International Semantic Web Conference (pp. 408-427). Cham: Springer Nature Switzerland.
- T. Phuttaamart, N. Kertkeidkachorn, and A. Trongratsameethong, "The Ghost at LLMs4OL 2024 Task A: Prompt-Tuning-Based Large Language Models for Term Typing", Open Conf Proc, vol. 4, pp. 85-91, Oct. 2024.