https://doi.org/10.51408/issi2025_200

Can Large Language Models Accurately Discriminate Subject Term Hierarchical Relationship?

Yuanxun Li[1], Hongyu Wang[2], Kaiwen Shi[3], Xiaoguang Wang[4]

[1]338018@whut.edu.cn, [2]hongyuwang@whut.edu.cn
School of Management, Wuhan University of Technology, Wenzhi Street, Hongshan District, Wuhan (China)

[3]shikaiwen@whu.edu.cn, [4]wxguang@whu.edu.cn
School of Information Management, Wuhan University, Bayi Street, Wuchang District, Wuhan (China)

Introduction

In the fields of scientometrics and informetrics, accurately determining the hierarchical relationships between subject term is crucial for literature retrieval, domain ontology modeling, and knowledge graph construction. The series of Klink algorithms proposed by Osborne infer relationships between research keywords by integrating multiple data sources and using co-occurrence analysis (Osborne, F., & Motta, E.2012). However, these algorithms struggle with issues such as high computational complexity and low recall when dealing with large-scale data and complex semantic relationships. To address large-scale literature data, most studies adopt a "recall-discrimination" two-stage approach for determining hierarchical relationships.

With the rise of large language models (LLMs), Xu explored the application of LLMs in complex natural language reasoning tasks (Xu, F., Hao, Q., et al., 2025). Researchers have attempted to use text information to identify semantic relationships between words, while Wang utilized LLMs for named entity recognition and semantic relationship extraction (Wang, Z., Huiru Chen, et al., 2025).

The purpose of this study is to explore whether large language models can accurately determine the hierarchical relationship of subject terms, and to compare the performance of large language models in two different phases mentioned before, so as to derive a framework for the application of large language models in the hierarchical relationship determination task. The code is available on GitHub[1].

Methodology

To assess the accuracy of various discriminative strategies in identifying hierarchical relationships of topic words, the study proposes a two-phase framework consisting of a "recall" phase for generating candidate topic word pairs and a "discrimination" phase for evaluating hierarchical relationships. The study introduces two authoritative knowledge systems as gold standards: OpenAlex Concepts[2] and Computer Science Ontology[3] (CSO 3.4.1, containing 165,913 pairs of hierarchical relationships), as shown in Figure 1.

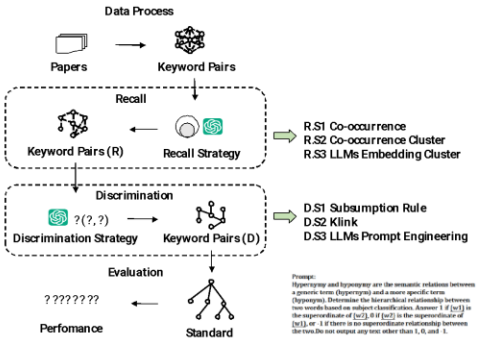

Figure 1. Framework for Discriminating Subject Term Hierarchical Relationships.

[1] https://github.com/Hipkevin/HierarchicalInfer
[2] https://api.openalex.org/concepts
[3] https://cso.kmi.open.ac.uk/downloads

Data Process

This study retrieved literature from the Web of Science Core Collection in the field of computer science from 2010 to 2023 (WC = "Computer Science"), totaling 932,210 articles. Author keywords were extracted from the dataset, and camel case nomenclature was applied to each keyword to ensure its semantic integrity (e.g., "information science" was transformed into "InformationScience"). The processed keywords served as potential subject term candidates, providing the foundation for the data used in the "recall" phase.

Subject Terms Recall

R.S1 Co-occurrence: Count the co-occurrence relationships in author keywords, and when the frequency of keyword co-occurrence pairs is greater than the retrieval year interval (14 years), the two keywords in the keyword co-occurrence pairs are used as candidate subject term pairs.

R.S2 Co-occurrence Cluster: Construct the co-occurrence frequency matrix of author keywords on the basis of R.S1, use this matrix to perform K-Means clustering, select the K value corresponding to the change point of the sum of the squared errors (SSE) curve as the number of clusters according to the principle of the elbow method, and then arrange and combine the keywords in each cluster according to the C_N^2 permutation and take them as candidate subject term pairs.

R.S3 LLMs Embedding Cluster: Also based on R.S1, the embedding vectors of the author keywords are obtained by using a LLM with a smaller number of parameters after distillation, and the candidate subject term pairs are obtained according to the clustering and permutation methods in R.S2.

Hierarchical Relationship Discrimination

D.S1 Subsumption Rule: For keywords x and y of a candidate subject term pair, $P(x|y)$ and $P(y|x)$ are computed, and x is the hypernymy of y when $P(x|y) = 1$ and $P(y|x) < 1$. Usually, the condition $P(x|y) = 1$ is relaxed to $P(x|y) > \alpha$, and α is chosen according to different domains and data sizes, usually 0.8.

D.S2 Klink: Semantic features are introduced on the basis of D.S1 to compute $L(x,y) = (P(x|y) - P(y|x)) * c(x,y) * (1 + N(x,y))$. $c(x,y)$ denotes the cosine similarity of keywords x and y in the co-occurrence matrix, and $N(x,y)$ denotes the string similarity of keywords x and y. In this study, the longest common subsequence distance (LCS) is used. x is the hypernymy of y for $L(x,y) > t$, and t is usually taken as 0.2.

D.S3 LLMs Prompt Engineering: The hierarchical relationship of each candidate subject term pair is discriminated by prompt engineering. The prompt template designed[4] in this study is as follows: `Hypernymy and hyponymy are the semantic relations between a generic term (hypernym) and a more specific term (hyponym). Determine the hierarchical relationship between two words based on subject classification. Answer 1 if {w1} is the superordinate of {w2}, 0 if {w2} is the superordinate of {w1}, or -1 if there is no superordinate relationship between the two.Do not output any text other than 1, 0, and -1`.

Result and Discussion

The accuracy of the recall and discrimination strategies in the experiments of this study on two datasets is shown in Table 1.

Table 1. The Accuracy (%) of the Recall and Discrimination Phases.

Strategy	OpenAlex	CSO
R.S1	**33.51**	**33.22**
R.S2	2.58	2.29
R.S3 (32b)	4.61	3.19
D.S1	4.05	3.45
D.S2	24.29	25.68
D.S3 (72b)	**51.42**	**42.49**
D.S3 (32b)	**49.39**	**39.34**

Without the recall strategy, the computational complexity is $\boldsymbol{O(N^2)}$, while with the recall strategy, the complexity of R.S1 is $\boldsymbol{O(N-M)}$ and R.S2 and R.S3 are $\boldsymbol{O(logN)}$. R.S1 based on co-occurrence frequency truncation performs best, while the LLM embedding-based clustering recall method outperforms the co-occurrence matrix clustering method.

[4]https://en.wikipedia.org/wiki/Hypernymy_and_hyponymy

In the discrimination strategy, D.S3 of qwen2.5 with 72b and 32b parameters correctly identifies all candidate subject terms recalled by R.S1, significantly outperforming the traditional co-occurrence analysis-based discrimination methods. Overall, both co-occurrence analysis and word embedding can detect hierarchical relationships from a semantic perspective. The co-occurrence relationships in R.S1 are broader, while the LLM word embedding provide more precise hierarchical information.

Conclusion

In this study, we propose a framework for the application of LLM in the subject term hierarchical relationship determination according to the two-stage approach of "recall-discrimination", and empirically demonstrate it on large-scale literature datasets in the computer science field. The results show that the large language models can accurately determine the hierarchical relationship between subject terms by relying on the zero-shot capability alone, and an efficient and accurate recall strategy is needed to realize the application framework on large-scale datasets. Follow-up studies can be carried out to optimize the clustering recall method.

Acknowledgments

This work was funded by the National Social Science Fund of China (21&ZD334) and National Natural Science Fund of China (No. 72404215).

References

Osborne, F., & Motta, E. (2012). Mining semantic relations between research areas. In The Semantic Web–ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I 11 (pp. 410-426). Springer Berlin Heidelberg.

Wang, Z., Chen, H., Xu, G., & Ren, M. (2025). A novel large-language-model-driven framework for named entity recognition. *Information Processing & Management*, *62*(3), 104054.

Huang, S., Huang, Y., Liu, Y., Luo, Z., & Lu, W. (2025). Are large language models qualified reviewers in originality evaluation? *Information Processing & Management*, 62(3), 103973.