Journal of Electronic Research and Application, 2025, Volume 9, Issue 4
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

# Research on the Construction of an Accounting Knowledge Graph Based on Large Language Model

**Yunfeng Wang***

Joint Logistics Support Force Engineering University, Chongqing 400000, China

**Author to whom correspondence should be addressed.*



**Abstract:** The article is based on language model, through the cue word engineering and agent thinking method, automatic knowledge extraction, with China accounting standards support to complete the corresponding knowledge map construction. Through the way of extracting the accounting entities and their connections in the pattern layer, the data layer is provided for the fine-tuning and optimization of the large model. Studies found that, through the reasonable application of language model, knowledge can be realized in massive financial data neural five effective extracted tuples, and complete accounting knowledge map construction.

**Keywords:** Accounting; Large language model; Knowledge graph; Knowledge extraction; Knowledge optimization

**Online publication:** August 7, 2025

## 1. Introduction

In the process of constructing the knowledge graph of accounting, the extraction of entities and relations from unstructured data is a key part to make it systematic enough. In this data processing, the application advantage of large language model is very significant. Large language model is the main intelligent model for performing key tasks such as text understanding, text generation and text reasoning in natural language processing. Through the reasonable application of this intelligent model, it can provide a new technical means for researchers to build corresponding knowledge graph, and assist them to realize the automatic extraction of entities, relations, and attributes in unstructured text. In order to ensure the efficiency and accuracy of knowledge map query. Based on this, researchers can use the current typical GPT-4o mini large language model as a support model to extract unstructured information. The text data in the accounting standards for Enterprises of China is used as the input of the dataset, and the entity and relationship design and setting are completed manually. The text data is used as the output qualification of the dataset in the large language model. After setting the corresponding model parameters, the information can be automatically extracted [1–3].

## 2. Research background

According to the State Council in 2024, the government work report on requirements, with the development of the digital economy in artificial intelligence + should be targeted at the main technology in the digital economy era. Based on this, the construction of an accounting knowledge graph supported by artificial intelligence has also become the focus of the modern economic field and related researchers.

In the study, the researchers used through fiscal website query and obtain the 42 Chinese accounting standards, according to the items, so as to implement segmentation extraction operation, using regular expression to its article extract 1257 accounting standards, will have no message to extract entries are removed, the rest of the amount of data is 1215. Considering the accounting subjects and accounting standards in financial reporting is numerous, so in extraction, researchers can select end entities, extract nouns, or financial accounting subjects. And will be one of the "department," "date," "chapters," and "entry," such as in the item to be processed only standard content, again through the GPT-4o mini model imposed extraction operation, and realize the rapid generation of a five yuan group. In this paper, combined with the background, the language model under the support of the financial accounting knowledge map building method was studied.

## 3. Construction method of accounting knowledge graph based on large language model

According to the background and ideas of the basic construction of the accounting knowledge graph supported by the large language model, the researchers determined the following steps in the specific research: the first is the construction of pattern layer; The second is the construction of data layer; The third is knowledge extraction; The fourth is knowledge optimization. The following are the main building methods to analyze the knowledge map.

### 3.1. Pattern layer construction

Model layer refers to the concept in the sense of a knowledge map, a concrete building, the researchers usually need to specify its entity, relation types, and attributes to specification, to form a normative framework, equivalent to a map of the skeleton. For the accounting mode of knowledge map layer, to describe the design more clearly, the researchers define the type of entity and relationship of graph theory, knowledge map, in the form of directed graph (entity) on behalf of the node set, each entity can be here on behalf of the accounting entity or accounts or accounting standards; $G=(V,R)$ $V$ $V_i \in V$ $R$ Collection on behalf of the side of the relationship between (entity), each of them an edge can represent a triad, here on behalf of the end of entity relationship, on behalf of relationship types [4,5]. $R_i=R$ $R_i–(R_j,R_k)R_j$ $R_k$ Based on this, the five yuan group building form in the model layer is as follows:

$$S=\{[Vi, type(V_i),Rj,Vk, type(Vk)]\} \quad (1)$$

Where, represents the head entity; $Vi$ $Rj$ represents an entity relationship; $Vk$ represents the tail entity; *type* represents the entity attribute or type.

### 3.2. Build data layer

The data layer refers to the instantiation process implemented on the schema layer. In the process of expressing knowledge with real financial data, the representation of quintuples at the data level is as follows:

$$D=\{[Vi, type(V_i),R_j,V_k, type(V_k)]\} \quad (2)$$

Through this structure, the association and attribute between accounting items can be captured, and the

complex association can be clarified, so as to lay the foundation for the subsequent retrieval and classification of the knowledge graph.

Therefore, no specific enterprise data is introduced into the data layer of this study, and all the data are in the abstract form of accounting knowledge. Therefore, in the specific research, researchers can merge the data layer into the schema layer, and use the basic enterprise information as the data layer to add to the knowledge graph.

## 3.3. Knowledge extraction

The main goal of this step is to transform the non-/ semi-structured form of accounting standards into a structured form of information and store it in a tuple form. The following are the main steps of knowledge extraction and their methods.

The first is to set the entity and relationship types. The GPT-4o mini large language model is used as the base model to extract knowledge from the schema layer. In order to realize the multidimensional representation of information and make its semantics more interpretable, so as to better understand the context and promote the orderly implementation of downstream tasks, researchers transform $[h, type, r, o, type]$ [6–8] the extracted information in the form of quintuples (i.e.). Among them, on behalf of the entity; *h o* Represents the tail entity; *r* Represents the entity relationship; *type* Represents entity type. After reasonably determining the entity types according to the actual situation and requirements, researchers need to use the word2vec model combined with the K-Means algorithm to determine the types of different entity relationships in the way of classification and integration. Accounting text within the relationship has a lot of, so in this study, the researchers chose from 18 kinds of relationship extraction processing to extract the types of entity relationships to make a reasonable limit.

The second is to set up the prompt word project. Considering the industry characteristics and the very professional accounting information, the model design of the cue word structure should be very careful, so as to ensure exclusive concepts and relationships of accounting recognition, and accurate extraction. In order to ensure the output quality, the researchers designed the prompt word project according to the following steps: (1) Role overview, the main role should include experts in the field of finance and accounting and knowledge graph; (2) process control, the design is fine tip of words, in order to guide model of the whole process of the output distribution, and prompt the running order, provide convenience for data extraction of layered distribution; (3) rely on limited, exclusive types of entities in the field of accounting and the relationships between words make clear limits, the dependence on cue word to limit model output; (4) sample drivers, introducing what - shot sample model, make its output structure be clearly understood, here is what cyber-shot sample mainly five yuan for a number of groups in the form of the sample, which should contain the different concepts and their relations in the field of accounting, so that the model can more in-depth understanding of accounting field unique professional terms and their relationship; (5) format, requirements model with five yuan JOSN format output structured set of data information, clear structure that its generated five yuan group, and reflect in JSON [9–12]. In this way, the effective transformation of accounting text to a knowledge graph can be realized in the model, which provides support for automatic extraction and association construction in the application of accounting information.

The third is to set up the filtering and checking function. After the text generated by the tooltip text is completed, researchers need to go through the way of filtering and check function reasonable setting, screening, and testing the model output text content, in order to ensure the accuracy, consistency. To end with a filter function of five yuan group entity type and its relationship to check, look to whether types associated with the cue word set will not match the data filter, thus further enhancing the professionalism of the output data and its consistency.

To check again after screening the content of the logical function, check compliance, make the five yuan group consistent with reasonable accounting logic structure and relationship; Should be checked with function check the output formats at the same time, make its are in conformity with the JSON format standard, and make $5 set of intact [13]. For the specific financial standards involved, the check function can also be used to compare the generated content with the actual names and numbers of the relevant standards, so as to ensure that the quoted content is consistent with the financial standards.

### 3.4. Knowledge optimization

With support from the language model, to further optimize the accounting knowledge map, the researchers can achieve this in actual effect as the foundation, through the following steps to implement optimization. (1) By means of duplicate data removal and steps to streamline the repeated information in the knowledge map and remove redundant information, make the knowledge of the rolls of entries unique, and ensure their coherence. (2) The relationship of the knowledge map and the real merge optimization makes its structure a more concise form, such as the synonymous relation or similarity relation to implement merge processing, combining the reality of accounting field needs further refinement, making all kinds of knowledge in the knowledge map more available. (3) The possible noise and false information for knowledge extraction to extract content, etc., to use an effective method for identification and cleaning, so that the accounting knowledge map is more accurate. In this process, in order to make its semantics more complete, researchers can also use the way of completing the context to enrich the content of each node, and further enhance [14,15] its semantic expressiveness under the premise of ensuring its usability. (4) by the model-driven enhancement method, optimize the structure of the knowledge base of the model and its contents automatically. To fine-tune again after processing of the language model as the foundation, to update the accounting knowledge expression, which makes the knowledge base optimized with more simpler structure, and further enhance the adaptability, to meet the demand of practical application.

## 4. Accounting knowledge map building based on language model results

In a big language model under the support of financial accounting knowledge map construction process, through the reasonable application of the above methods, the researchers completed the unstructured text to $5 sets of reasonable conversion, data structure and its in the form of a $5 sets of data stored in a structured database, for subsequent financial knowledge map visualization and data query support, etc. The basic format is five yuan group, its storage format for JSON. $[h, h-type, r, o, o-type]$ Through model testing, it is found that the readability and scalability of the accounting knowledge graph are very good, and the data storage efficiency is very high. **Table 1** shows the five-tuple data set (part) of the financial and accounting knowledge graph constructed based on the large language model.

After the construction of the quintuple dataset, the researchers also directly established database connections with the help of the existing visualization tools in the Neo4j database to meet the dynamic display requirements of financial knowledge graph data. In the specific application, we only need to execute the Cypher query command under the condition of the Neo4j database to query the relationship between each node in the knowledge graph. For example, when querying the entity of “intangible assets,” the user can input the statement into the Cypher database, and then the corresponding knowledge graph can be displayed in the form of a visualization, so that all kinds of complex relationships in the knowledge graph are more intuitive and easier to understand. In this way, accounting staff can quickly identify the relationships and potential connections among important accounting

entities, so as to provide informatization, intelligence, and visualization technical support for their accounting work. **Figure 1** shows the visual display effect (part) of the query results of "intangible assets" graph in the accounting knowledge graph constructed based on the large language model.

**Table 1.** The accounting knowledge map construction based on language model $5 groups of data sets (part)

| Serial number | *h* | *h-type* | *r* | *O* | *h-type* |
|---|---|---|---|---|---|
| 1 | Purchase cost | CA | Including | Related taxes and fees | IEA |
| 2 | Procurement costs | CA | Including | Insurance premium | IEA |
| 3 | Procurement costs | CA | Including | Transportation costs | IEA |
| 4 | Procurement costs | CA | including | Handling charges | IEA |
| 5 | Distribution of manufacturing expenses | CA | According to | Nature of manufacturing expenses | OA |
| 6 | Manufacturing expenses | CA | Belong | Overhead expenses | CA |
| 7 | Storage costs | CA | Not counted | Cost of inventory | CA |
| 8 | Borrowing costs | IEA | Count | Cost of inventory | CA |

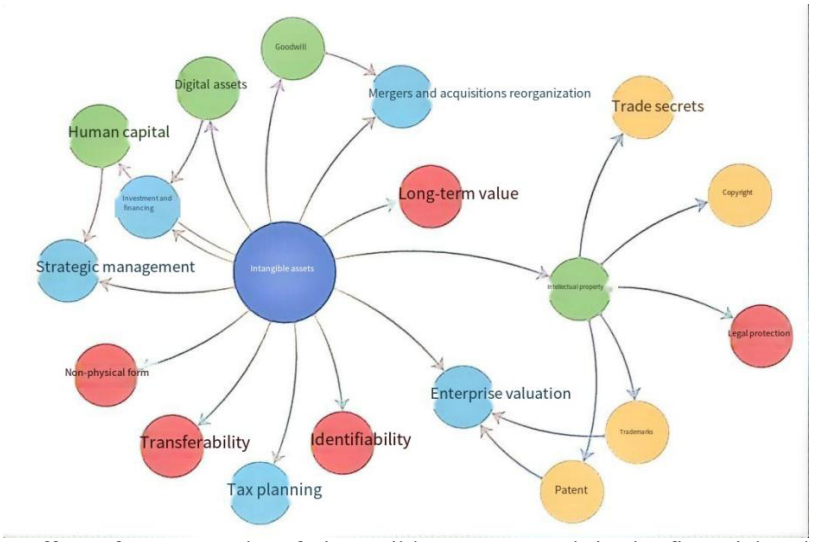


**Figure 1.** Visual display effect of query results of "intangible assets" graph in the financial and accounting knowledge graph constructed based on large language model (part)

## 5. Conclusion

In conclusion, the large language model is a key natural language processing model in the field of modern artificial intelligence, and also a key supporting technology model for constructing relevant knowledge graphs in the field of modern economics. Especially in the process of accounting knowledge map construction, the researchers are more against large language models than doing enough. The existing accounting standards are introduced into the large language model as knowledge and constraints, and the pattern layer and data layer of the accounting knowledge graph can be constructed in the large language model with reasonable methods and measures. If necessary, the two can be combined into one pattern layer. Combined with the actual situation of accounting data, information, and so on, with support from the language model, at a reasonable level, to manipulate the model of knowledge extraction and optimization, etc. So can making a big language model in modern financial accounting

knowledge map construction plays a significant advantage, as its knowledge map subsequent application provides strong intelligence support.

## Disclosure statement

The author declares no conflict of interest.

## References


[1] Tao Y, 2024, Big Language Model Application in Higher Vocational Accounting Research and Practice. Journal of Jiangsu Vocational and Technical College of Architecture, 2024(4): 88–92.

[2] Huang J, Zhu H, Xu C, et al., 2025, Application of Large Language Model in Audit Judgment. Finance Monthly, 2025(6): 21–28.

[3] Dong J, Shangguan Y, Li S, 2025, Research on the Application of Big Language Model Intelligent Customer Service in the Financial Management Field of Y Group. International Business Finance and Accounting, 2025(3): 39–43.

[4] Zhong H, Chen S, Wang M, 2024, Research on the Ability of Accounting Vertical Domain Reasoning Based on Large Language Model. Journal of Finance and Accounting Monthly, 2024(11): 17–25.

[5] Du QX, 2024, ChatGPT Application Exploration in Accounting Work. Friends of Accounting, 2024(4): 130–137.

[6] Du QX, 2024, Application and Exploration of ChatGPT in Accounting Work: Based on Wen Xin Yi Yan, ChatGPT and ChatSonic Test. Friends of Accounting, 2024(4): 130–137.

[7] Wu W, Zhao YD, Zhao Y, et al., 2023, Application of GPT and Other Large Language Models in Accounting and Auditing. International Business and Accounting, 2023(22): 81–87.

[8] Jin Y, Li C 2023, The Impact and Response of ChatGPT on Accounting Personnel Function Transformation: An Example Based on Questionnaire. Monthly Journal of Finance and Accounting, 2023(18): 52–59.

[9] Ni Y, Wang X, 2023, Monetary Policy, Business Environment and Economic Growth: An Empirical Study from the Perspective of High-Quality Economic Development. International Business Accounting, 2023(16): 59–63.

[10] Tan Y, Zhang S, 2023, Application Prospects, Challenges and Countermeasures of ChatGPT in Accounting Field. Finance and Accounting Research, 2023(4): 48–52.

[11] Chen CH, Wu HH, Xiong ZY, et al., 2025, Research on Enterprise Financial Crisis Early Warning Based on DeepSeek Big Language Model. Journal of Xihua University (Philosophy and Social Sciences), 2025(3): 50–58.

[12] Wu Y, 2025, Big Model Enabling Financial Field. China Collective Economy, 2025(12): 193–196.

[13] Wang J, 2025, Intelligent Retrieval Method of Hospital Financial Data Driven by Large Language Model. Business Accounting, 2025(7): 25–29.

[14] Luo G, Wang B, 2025, Research Audit Boost Financial Revenue and Expenditure Audit Digital Intelligence Transformation. China Internal Audit, 2025(3): 67–70.

[15] Wen's Food Group Co. LTD, 2025, Multi-Scale and Cross-Domain Index Change Attribution Method and System Based on Knowledge Enhancement: CN202510029150.3, March 4, 2025.