

Construction of a Traditional Chinese Medicine Knowledge Graph Based on Multi-Agent Systems

Ziqiang Wang ¹, Yuebin Su ¹ and Jing Song ²

¹ School of Mathematics and Statistics, Sichuan University of Science & Engineering, Zigong Sichuan, 643000, China
² Southwest Jiaotong University, Chengdu Sichuan, 610031, China

Abstract: With the rapid advancement of artificial intelligence technology, the digitization of traditional Chinese medicine knowledge systems remains an unresolved challenge, hindered by governance bottlenecks stemming from data complexity and fragmented knowledge. Traditional manual curation struggles to balance scale and precision, creating an urgent need for automated, high-accuracy knowledge graph construction solutions. This paper proposes a multi-agent-based method for constructing a knowledge graph of traditional Chinese medicine. By designing the multi-agent framework KGMA, the knowledge graph construction task is decomposed into sub-tasks: entity extraction for each category and relationship construction. Each sub-task is handled by a specialized agent. The data used in this study is sourced from authoritative and comprehensive Chinese medicine encyclopedias, with detailed information on 10,268 types of Chinese medicinal materials collected via Python web scraping technology. Experimental results demonstrate that the construction of a traditional Chinese medicine knowledge graph based on the KGMA framework achieves high accuracy. Comparative analysis reveals that triples constructed using DeepSeek-V3 as the base model exhibit the highest correctness rate, reaching 99%. Ultimately, a traditional Chinese medicine knowledge graph comprising 55,089 entities and 192,688 relationships was successfully constructed. The study confirms that the KGMA framework significantly enhances the efficiency and accuracy of TCM knowledge graph construction. It provides robust technical support for addressing TCM data governance challenges, advancing the systematic integration of TCM knowledge, and facilitating its modern applications. This holds significant practical implications for promoting the innovative development of the TCM industry.

Keywords: Multi-Agent Systems; Traditional Chinese Medicine; Knowledge Graph.

1. Introduction

As a traditional medical system with millennia of history, Traditional Chinese Medicine (TCM) has made immense contributions to human health through its unique theoretical framework and rich clinical experience. However, with the rapid advancement of modern medicine, the complexity of TCM knowledge has become increasingly prominent, emerging as a critical bottleneck constraining its modernization, internationalization, and intelligent development. The TCM knowledge system exhibits multidimensional characteristics, such as the properties of herbal medicines including taste, nature, efficacy, main

indications, and meridian tropism. These attributes are interconnected through complex networked relationships. The synergistic interaction among "nature-taste-meridian tropism" determines a drug's efficacy and primary indications [1]. Figure 1 presents the theoretical framework of TCM, illustrating the core logical chain of herbal clinical application. The synergistic mechanism of nature-taste-meridian affinity is pivotal. Taking the herbal medicine *Coptis chinensis* (Huanglian) as an example: its nature is cold, addressing "excess heat" conditions by clearing heat and purging fire; its taste is bitter, targeting "internal accumulation of damp-heat and ascending fire-toxin" by drying dampness, detoxifying, and purging;

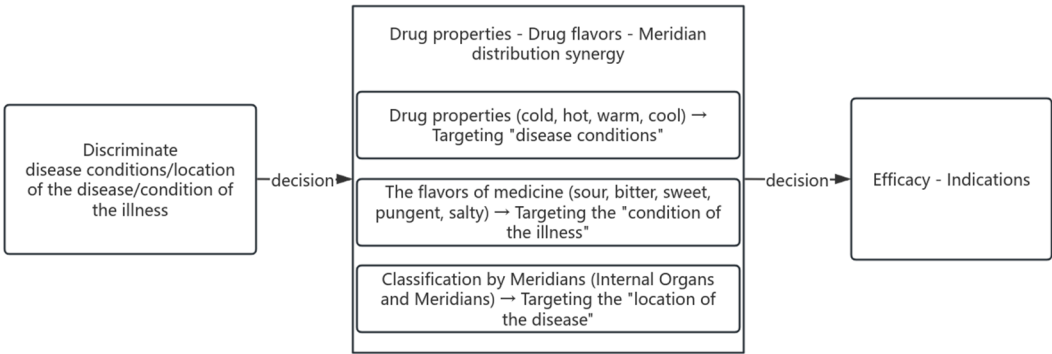


Fig 1. Theoretical Framework Diagram of Traditional Chinese Medicine

Its meridian tropism targets the heart, stomach, and large intestine meridians, addressing pathological locations of "heart-stomach fire excess and large intestine damp-heat." It thus clears heart fire, drains stomach fire, and eliminates large

intestine damp-heat. The synergy of these three elements endows Huanglian with the comprehensive efficacy of "clearing heat and drying dampness, draining fire and detoxifying," corresponding to its main indications for

syndromes such as "damp-heat diarrhea, stomach heat vomiting, and excessive heart fire." In clinical practice, taking the classic case of Coptis treating damp-heat diarrhea as an example, the clinical presentation involves high fever, abdominal pain, tenesmus, and bloody diarrhea. This is diagnosed as "large intestine damp-heat syndrome." The rationale for its use is as follows: - For heat pathology, the "cold" nature of Coptis is selected to clear heat; - For dampness pathology, the "bitter" taste of Coptis is chosen to dry dampness; - For the affected location in the large intestine, Coptis—which enters the Large Intestine Meridian—is selected to directly target the site of pathology.

Although existing research attempts to format and store Chinese herbal medicine data in traditional databases to advance modernization, the complexity of raw data makes effective modeling challenging. Dimensioning this unstructured herbal knowledge is crucial.

In recent years, Large Language Model (LLM) technology has achieved breakthrough progress. Leveraging the hierarchical neural network structure of the Transformer architecture, this technology employs self-attention mechanisms to perform deep semantic modeling on massive unlabeled corpora. It not only effectively captures long-range dependencies between words but also develops distributed representation capabilities for grammatical rules, semantic logic, and pragmatic features. This probabilistic language understanding paradigm enables LLMs to outperform traditional methods in natural language processing tasks such as machine translation and text generation. Concurrently, the field of intelligent agent research is undergoing a paradigm shift. By integrating LLMs as the cognitive core of agents and deeply fusing semantic comprehension, knowledge reasoning, and decision-making capabilities, intelligent agent systems have achieved significant enhancements in autonomous decision-making, dynamic collaboration, and environmental adaptation. Against this backdrop, constructing an integrated paradigm for LLMs and multi-agent systems has emerged as a frontier topic in artificial intelligence.

To address these challenges, this paper proposes a novel multi-agent framework named KGMAL. Each agent employs an LLM as its foundational model to execute distinct tasks, ultimately processing traditional Chinese medicine (TCM) data into a TCM knowledge graph. Data is sourced from the TCM Compendium, an authoritative and comprehensive TCM knowledge repository. Utilizing Python web scraping technology, extensive herbal medicine data was extracted from this website, including key attributes such as herb names, aliases, properties, meridian tropism, primary indications, and origins. This paper makes the following contributions:

- 1) Employing multi-agent systems for complex text entity extraction, where each entity extraction agent focuses on distinct attributes of herbal medicines. This approach significantly reduces extraction errors and incorporates confidence scoring to filter entities, further enhancing accuracy.

- 2) Fixed relationships are established for entities of different attributes to eliminate ambiguity. A triplet clarity scoring agent is introduced to ensure the accuracy of triplets.

- 3) The TCM knowledge graph construction framework automates the entire process. It takes complex herbal medicine data as input, processes it through the KGMAL framework, and automatically generates a TCM knowledge graph. This graph is then stored in Neo4j for visualization.

- 4) A TCM knowledge graph dataset was constructed,

comprising 55,089 entities and 192,688 triples.

2. Related Work

2.1. Construction of Traditional Chinese Medicine Knowledge Graphs

The Traditional Chinese Medicine knowledge graph organizes and represents knowledge in the TCM field in a structured format. It clearly displays TCM concepts, entities, and their interrelationships, providing crucial support for the informatization and intelligent application of TCM.

TCM knowledge is primarily stored in textual form within ancient texts, literature, and databases. Consequently, text mining techniques play a vital role in knowledge graph construction. Employs natural language processing techniques, combined with specialized dictionaries and rules specific to the TCM domain, to achieve automatic recognition and extraction of entities such as herbal medicine names, properties, and efficacy.

Accurate extraction of relationships between entities is a critical step in knowledge graph construction following entity extraction. Proposed a deep learning-based relation extraction method by training neural network models. Further optimization of algorithms and models is required to enhance relation extraction accuracy. However, this approach necessitates pre-training with data, imposing high demands on data quality and rendering it unsuitable for practical entity extraction tasks in TCM materials. In contrast, utilizing LLMs for data extraction eliminates the need for cumbersome data annotation, enabling efficient and accurate completion of entity extraction tasks.

2.2. Intelligent Agents

The development of large-scale language models traces back to Google's 2018 release of BERT (Bidirectional Encoder Representations from Transformers)[2]. BERT pioneered the Transformer architecture and the concept of pre-trained language models. By undergoing unsupervised pre-training on massive corpora, it learned general language representations and achieved remarkable results when fine-tuned for specific tasks. Since then, language models based on the Transformer architecture have proliferated. Released in November 2022, ChatGPT significantly enhanced conversational capabilities and generative quality through reinforcement learning and human feedback techniques, marking another milestone in natural language processing. Multi-agent systems were first applied to natural language processing in 1998 for entity extraction tasks [3]. Traditional intelligent tasks relied on planning languages and symbolic reasoning to accomplish objectives. While effective for structured, standardized tasks, these approaches faced severe limitations with non-standardized textual data and struggled to process uncertain information. The emergence of large language models has brought new possibilities to such intelligent tasks. Lu Y et al. employed multi-agent systems for entity and relationship extraction in genetics, demonstrating their effectiveness in knowledge graph construction[4]. Therefore, this paper utilizes large language models as the backbone for multi-agent systems, offering novel approaches and methodologies for constructing knowledge graphs in traditional Chinese medicine.

3. Data Sources

TCM, originating and developing in China, is a medical system characterized by holistic concepts, syndrome differentiation and treatment, and other features. For a single herbal medicine, TCM requires consideration of complex attributes such as its medicinal properties, taste, meridian tropism, efficacy, and primary indications.

Therefore, clearly displaying each attribute of a medicinal herb is crucial.

The data in this paper primarily originates from the Chinese Medicine Encyclopedia ([https://zhongyibaodian.com]), a professional knowledge repository for traditional Chinese medicine. It contains extensive information on classical texts, formulas, herbs, diseases, and treatment methods related to TCM. The data from the Chinese Medicine Encyclopedia is authoritative and comprehensive, ensuring accuracy and suitability as raw data for constructing a TCM knowledge graph. This paper extracts data on Chinese medicinal materials, with key fields including material name, aliases, properties and flavors, meridian tropism, main indications, and sources. Data collection was implemented using Python web scraping. The original dataset comprised 10,647 medicinal materials, which were deduplicated to yield 10,268 unique entries serving as the raw data for this study.

Among the crawled data fields, the herb name serves as the unique identifier. Its format is clear and highly standardized, requiring no additional processing for knowledge graph construction. However, five fields—alternative names, properties and flavors, meridian tropism, main indications, and sources—exhibit significant complexity. The properties and flavors field not only encompasses the medicinal nature and taste of herbs but also suffers from numerous inconsistencies in data format. While the Indications field

integrates two critical pieces of information: the herb's efficacy and the primary diseases it treats. This complexity makes traditional data processing methods ineffective and unable to meet the stringent requirements of knowledge graph construction. More challenging still, the rich entities inherent in the field of traditional Chinese medicine—such as the specific effects of herbs and their meridian pathways—are difficult to accurately extract using conventional methods.

With the rapid advancement of artificial intelligence, particularly the emergence of Multi-Agent Systems (MAS), new solutions have emerged for this challenge. Leveraging their strong autonomy, interactivity, and collaboration capabilities, MAS can design specialized agents tailored to each field's characteristics for precise entity extraction. These agents, powered by deep learning and natural language processing technologies, efficiently identify and extract key entities and relationships within TCM data, providing robust support for constructing TCM knowledge graphs. This innovative approach not only significantly enhances the efficiency and accuracy of TCM knowledge graph construction but also achieves breakthrough progress in entity extraction, laying a solid foundation for the digital and intelligent transformation of TCM knowledge.

4. Agent Principles

The KGMAL framework comprises six LLM-based agents responsible for extracting entities related to aliases, properties, meridian tropism, main indications, and sources, as well as evaluating triplet quality. Each agent performs specialized tasks through customized prompts and hyperparameters, ultimately collaborating to generate the TCM knowledge graph.

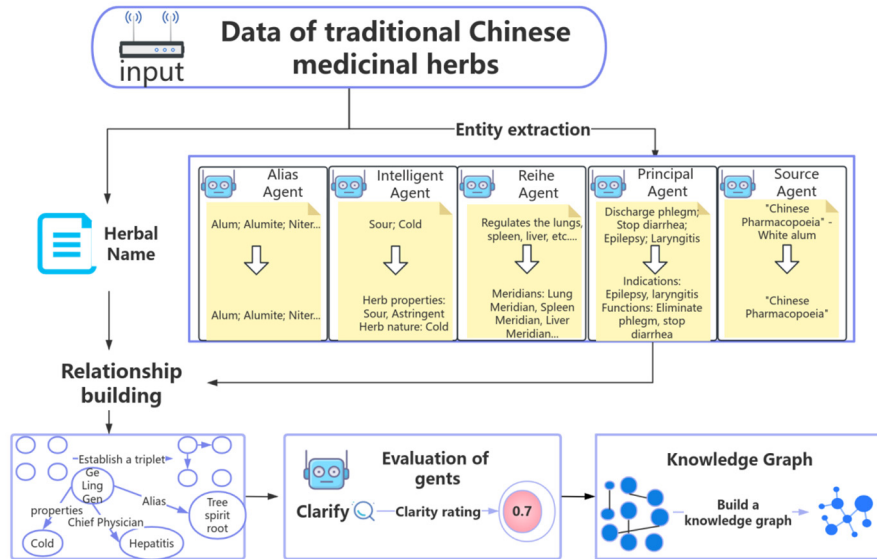


Fig 2. KGMAL frame diagram

This section introduces KGMAL, a multi-agent framework for constructing traditional Chinese medicine knowledge graphs, as illustrated above. It leverages specialized LLMs to perform end-to-end knowledge graph construction. Our approach decomposes the entire task into modular subtasks, from data extraction to final knowledge graph construction. For KGMAL, this chapter first proposes a problem formulation, then details the design and mathematical foundations of each agent within the workflow.

4.1. Problem Formulation

Let the final traditional Chinese medicine knowledge graph be defined as $G=(V,E)$ where V is the set of entities (e.g., aliases, medicinal properties, tastes, meridian tropism, etc.), and E is the set of edges representing relationships. Each relationship is defined as a triplet $t=(e_h, r, e_t)$, where $e_h, e_t \in V$, and r denotes the relationship type (e.g., alias, main indication, meridian tropism). The data crawled using Python web

scrapers is denoted as $P = p_1, p_2, \dots, p_n$. The objective is to automatically extract triplets from each data point p_i and ultimately construct the Traditional Chinese Medicine knowledge graph. $G=(V,E)=\text{Extract}(p_i)$, where $\text{Extract}(p_i)$ denotes the set of triplets extracted from input data p_i , for $p_i = (p_y, p_b, p_x, p_g, p_z, p_l)$. Here, p_y represents the herbal name, p_b denotes the alias, p_x indicates the nature and taste, p_g signifies the meridian tropism, p_z reflects the main indications, and p_l specifies the source of the herbal material.

4.2. Alias Agent

Based on the agent's entity extraction of herbal medicine aliases, it is defined as

$$R_b = \text{LLM}_{\text{alias}}(p_b, T_b) \quad (1)$$

Among these, $\text{LLM}_{\text{alias}}$ is an LLM equipped with the dedicated prompt T_b for extracting entities from herbal medicine aliases. It extracts entities from Chinese herbal medicine aliases and evaluates the confidence level of the task, using the confidence score to measure the completion degree of the extraction task. The output parsing function $\varphi(R_b)$ is

$$\varphi(R_b) = \begin{cases} V_b = \{v_1, v_2, \dots, v_n\} & \text{Alias collection} \\ C_b \in [0,1] & \text{Confidence score} \end{cases} \quad (2)$$

defined to extract structured results:

$$\varphi(R_b) = \begin{cases} V_b = \{v_1, v_2, \dots, v_n\} & \text{Alias collection} \\ C_b \in [0,1] & \text{Confidence score} \end{cases} \quad (2)$$

4.3. Properties and Flavors Agent

The agent performs entity extraction for the medicinal properties and tastes of herbs, defined as:

$$R_x = \text{LLM}_{xw}(p_x, T_x) \quad (3)$$

where LLM_{xw} is an LLM equipped with a dedicated prompt T_x for extracting entities related to medicinal properties and tastes. Medicinal properties are categorized into five types: cold, hot, warm, cool, and neutral. Tastes are classified into seven types: sweet, bitter, pungent, sour, salty, astringent, and bland. Within these constraints, it extracts entities for medicinal properties and tastes from p_x and evaluates the task's confidence level. The confidence score measures the completion degree of the extraction task. The output parsing function $\varphi(R_x)$ is defined to extract structured results:

$$\varphi(R_x) = \begin{cases} V_x = \{v_1, v_2, \dots, v_n\} & \text{Pharmaceutical properties collection} \\ V_w = \{v_1, v_2, \dots, v_n\} & \text{Collection of medicinal ingredients} \\ C_x \in [0,1] & \text{Confidence score} \end{cases} \quad (4)$$

4.4. Meridian-Targeting Agent

Perform entity extraction for meridian tropism based on the agent, defined as

$$R_g = \text{LLM}_{\text{Meridians}}(p_g, T_g) \quad (5)$$

where $\text{LLM}_{\text{Meridians}}$ is an LLM equipped with a dedicated prompt T_g for extracting meridian-affinity entities. The prompt includes the following meridians: Stomach Meridian,

Spleen Meridian, Large Intestine Meridian, Kidney Meridian, Lung Meridian, Heart Meridian, Liver Meridian, Bladder Meridian, Gallbladder Meridian, Triple Energizer Meridian, Small Intestine Meridian, and Pericardium Meridian. This LLM extracts entity-level meridian affiliations from p_g while evaluating task confidence. The confidence score measures extraction accuracy. The output parsing function $\varphi(R_g)$ extracts structured results:

$$\varphi(R_g) = \begin{cases} V_g = \{v_1, v_2, \dots, v_n\} & \text{Classification of Categories} \\ C_g \in [0,1] & \text{Confidence score} \end{cases} \quad (6)$$

4.5. Primary Indications Agent

The agent performs entity extraction for the therapeutic indications and efficacy of herbal medicines, defined as

$$R_z = \text{LLM}_{\text{treatment}}(p_z, T_z) \quad (7)$$

where $\text{LLM}_{\text{treatment}}$ is an LLM equipped with the dedicated prompt T_z for extracting entities related to the main indications and efficacy of herbal medicines. It extracts entities describing the main indications and efficacy of Chinese herbal medicines and evaluates the confidence level of this task. The confidence score measures the completion level of this extraction task. Define the output parsing function $\varphi(R_z)$ to extract structured results:

$$\varphi(R_z) = \begin{cases} V_z = \{v_1, v_2, \dots, v_n\} & \text{Main treatment plan} \\ V_e = \{v_1, v_2, \dots, v_n\} & \text{Efficacy set} \\ C_z \in [0,1] & \text{Confidence score} \end{cases} \quad (8)$$

4.6. Source Agent

Perform entity extraction for the source of medicinal materials using an agent, defined as:

$$R_l = \text{LLM}_{\text{from}}(p_l, T_l) \quad (9)$$

Where LLM_{from} is an LLM with a dedicated prompt T_l for extracting source entities of herbal medicines. It extracts source entities of traditional Chinese medicinal materials and evaluates the confidence level of this task. The confidence score measures the completion level of this extraction task. Define the output parsing function $\varphi(R_l)$ to extract structured results:

$$\varphi(R_l) = \begin{cases} V_l = \{v_1, v_2, \dots, v_n\} & \text{Source Collection} \\ C_l \in [0,1] & \text{Confidence score} \end{cases} \quad (10)$$

4.7. Evaluating Agents

In a multi-agent system, each agent is responsible for extracting specific entity types, such as aliases, properties, tastes, meridian tropism, and primary indications of Chinese medicinal materials. Predefined, fixed relationship patterns exist between these entity types. For example, the relationship between a medicinal material and its properties is explicitly defined as "a certain medicinal material possesses a certain property," represented as a triplet in the knowledge graph. However, although agents can extract these entities and construct preliminary triples, whether these triple expressions are sufficiently clear and accurate, and whether they authentically reflect the logical relationships within the traditional Chinese medicine knowledge system, still requires further verification and confirmation. Let data = (V, C), where

V is the entity set processed by the agent, i.e., $V=\{V_i|i=b,x,w,g,z,e,l\}$, and C is its corresponding confidence score, i.e., $C=\{C_i|i=b,x,g,z,l\}$. Then the relation processing function is defined as

$$\text{relation_deal}(p_y, \text{data}) = \begin{cases} t=(p_y, r, V) & , C \geq \theta \\ \text{None} & , C < \theta \end{cases} \quad (11)$$

This approach yields $\text{triplest}=(p_y, r, V)$ that constitute the knowledge graph. To ensure triplet accuracy, the confidence score of V must exceed θ , where θ requires experimental validation to determine an appropriate threshold. The clarity and correctness of triplet expressions necessitate further assessment. An agent evaluates generated triplets using a clarity score defined as

$$\text{Clarity} = \text{LLM}_{\text{Evaluator}}(t, T_e) \quad (12)$$

where $\text{LLM}_{\text{Evaluator}}$ is an LLM equipped with dedicated prompt T_e for extracting entities related to medicinal material origins. The clarity score Clarity is assigned to generated triples. A threshold α is set for clarity, and experimental testing evaluates the impact of different thresholds on generated triples. Triples scoring above α are ultimately selected as qualified triples for constructing the traditional Chinese medicine knowledge graph.

5. Experiments and Analysis

This section proposes a multi-agent-based knowledge graph construction framework (KGMAL). Unlike traditional NLP tasks relying on standardized datasets, this approach employs multi-agents to construct knowledge graphs from unstructured TCM datasets. Multiple evaluation methods are adopted to assess the framework's stability and the accuracy of the constructed TCM knowledge graph.

Regarding data, 1,000 randomly selected entries from the crawled dataset were designated as the test set. Manual entity extraction and triplet generation on this test set yielded 7,895 entities and 20,039 triplets.

5.1. LLM Model

Selecting an appropriate LLM model is crucial for

achieving accurate and efficient results in this research. This paper screens models based on three core criteria. First is model accessibility: the model must offer one-click API access or official images on mainstream domestic computing platforms, enabling activation and sustained invocation without VPNs or overseas credit cards. This ensures stable, reproducible experimental workflows while reducing research costs and facilitating academic exchange and collaboration. Second is the model's Chinese language processing capability. Given the research's focus on tasks within the Chinese linguistic context, robust Chinese language processing is fundamental for the model to effectively analyze and generate relevant content. Finally, the model's API cost-effectiveness is considered. Reasonable API call costs ensure research can be conducted sustainably within budget constraints, avoiding limitations on model usage frequency and scope due to high expenses.

During model selection, options like GPT-4 were excluded. First, GPT-4 imposes geographic or account-level access restrictions in China's mainland, where access permissions are strictly controlled. Its complex application process and low approval rates significantly hinder research progress, making it unsuitable for large-scale, sustained research needs. Second, despite its multilingual capabilities, GPT-4 falls short in understanding Chinese linguistic conventions, grasping Chinese semantics, and processing specific Chinese cultural context compared to models specifically optimized for Chinese. This inadequacy prevents it from fully meeting the requirements for in-depth, precise analysis of Chinese text in this paper. Finally, GPT-4 costs approximately \$10 per million tokens for input and \$30 per million tokens for output, which is ten times the cost of moonshot-v1, making it prohibitively expensive.

Considering model accessibility, Chinese processing capability, and cost-effectiveness, moonshot-v1, GLM-4-Plus [5, 6], Qwen-Plus, and DeepSeek-V3[7] were selected as experimental models, as detailed in Table 1. Each agent in KGMAL used the same model for every experiment, ultimately analyzing the construction of traditional Chinese medicine knowledge graphs based on different models.

Table 1. Model Specifications

Model Name	Context Length	Acquisition Method	Chinese Token Efficiency	API Input Price	API Output Price
moonshot-v1	8k-128k	API	1.2x	1 yuan	10 yuan
GLM-4-Plus	128K	API	1.5x	1 yuan	8 yuan
Qwen-Plus	32k-128k	API	1.3x	0.8 yuan	2 yuan
DeepSeek-V3	64k	API	1.1x	0.5 yuan	8 yuan
GPT-4 (for comparison)	32k	API	0.9x	16 yuan	64 yuan

5.2. Evaluation Metrics

Despite the absence of standardized evaluation benchmarks, we employed a multidimensional assessment approach for experimental evaluation. Specifically, this included:

Core Metric:

M_{Con} : The average of all extracted entity confidence scores.

Average Clarity (M_{Cla}): The average clarity score, indicating the accuracy or directness of each triplet.

Graph Metrics:

Number of Triplets Δ_{Tri} : The quantity of qualified triples suitable for constructing the TCM knowledge graph following evaluation.

Number of Entities Δ_{Ent} : The number of qualified entities

for constructing the TCM knowledge graph after evaluation.

Quality Metrics:

For quality metrics, this paper employs the most widely used F1-score, precision (P_{Tri}), and recall (R_{Tri}) to evaluate the knowledge graph construction task.

$$P_{Tri} = \frac{TP}{TP+FP} \times 100\% \quad (13)$$

$$R_{Tri} = \frac{TP}{TP+FN} \times 100\% \quad (14)$$

$$F1\text{-score} = \frac{2 \times P_{Tri} \times R_{Tri}}{P_{Tri} + R_{Tri}} \times 100\% \quad (15)$$

TP denotes the number of correct triples extracted by the experiment, FP represents the number of incorrect triples extracted, and FN indicates the number of correct triples not extracted. For the obtained triples $t=(p_y, r, V)$, they are compared against the triples in the validation set to filter out

correct triples, yielding TP, FP, and FN. Subsequently, the F1-score (F1-score), precision (P_{Tri}), and recall (R_{Tri}). Therefore, precision (P_{Tri}) represents the proportion of correctly extracted triples to the total number of triples extracted in the experiment, while recall (R_{Tri}) represents the proportion of correctly extracted triples to the total number of triples in the dataset. The F1 score combines these two metrics and serves as a comprehensive indicator for evaluating the model in this experiment.

5.3. Results Analysis

The results of experiments conducted on the KGML framework using DeepSeek-V3 as the base model, along with a single agent and four large models, are shown in Table 2. After ablation experiments, the confidence and clarity thresholds were both set to 0.7. Comparing the knowledge graph statistics in Table 2 between the single agent and KGML frameworks reveals that the KGML framework demonstrates superior entity and triplet extraction capabilities. moonshot-v1 extracted the most entities and constructed the most triples, while GLM-4-plus produced the closest match in entity count to ground truth. This effectiveness stems from KGML's efficient extraction of complex semi-structured data. Furthermore, the agent design and entity extraction confidence filtering mechanism ensure both the stability of entity extraction tasks and the accuracy of extracted entities.

Within the KGML framework, the confidence score for

entity extraction by agents quantifies the task completion quality of individual agents, while the average confidence score reflects the overall framework performance. Experimental results show that moonshot-v1 achieved the highest average confidence score (0.9840), with all three other models exceeding 0.94, indicating the framework's high reliability in entity extraction tasks.

Regarding quality metrics, different large models exhibit varying accuracy in knowledge graph construction within the KGML framework. By comparing triples extracted from KGML with those from real knowledge graphs, we calculated triple accuracy, recall, and F1 scores to evaluate each model's knowledge graph construction capability. Results show DeepSeek-V3 excels in this domain, achieving a triplet accuracy of 0.9912, recall of 0.9889, and an F1 score of 0.9901. Compared to a single agent, these metrics improved by 0.4490, 0.7002, and 0.6177 respectively, demonstrating the effectiveness of the KGML framework. The F1 score improvement over a single agent reached 2.55–2.65 times. Additionally, the other three large models demonstrated outstanding accuracy in knowledge graph construction, all achieving high levels. This indicates that the KGML framework possesses high versatility, effectively supporting the application of multiple large models in knowledge graph construction tasks. It provides an efficient and reliable foundational framework for diverse models, contributing to the advancement and application of knowledge graph-related technologies.

Table 2. Experimental Results Table

Model	Core Metrics		Knowledge Graph Statistics		Quality Metrics		
	$M_{Con} \uparrow$	$M_{Cla} \uparrow$	$\Delta_{Ent} \uparrow$	$\Delta_{Tri} \uparrow$	$P_{Tri} \uparrow$	$R_{Tri} \uparrow$	$F1 \uparrow$
Single-Agent	NA	NA	4475	11034	0.5244	0.2887	0.3724
moonshot-v1	0.9840	0.9259	8020	19998	0.9542	0.9523	0.9533
QWen-plus	0.9480	0.9458	7653	19815	0.9613	0.9505	0.9559
GLM-4-plus	0.9476	0.9016	7828	19866	0.9805	0.9720	0.9762
DeepSeek-V3	0.9417	0.9496	7925	19992	0.9912	0.9889	0.9901

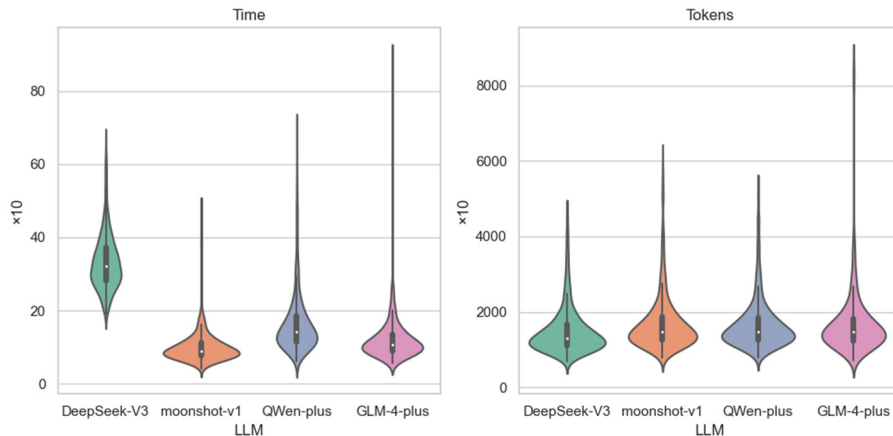


Fig 3. Time and Resource Consumption Diagram

Through detailed recording and analysis of experimental data, we conducted an in-depth exploration of the operational efficiency of different large models within this framework. Specifically, the number of tokens consumed and the corresponding runtime for each data processing step are intuitively presented in Figure 3. These data not only reveal differences in resource utilization among models but also reflect their relative strengths and weaknesses in processing speed.

moonshot-v1 demonstrated the highest time efficiency but

was not the most resource-efficient. While DeepSeek-V3 did not excel in time efficiency, it achieved the most resource-efficient performance. Given DeepSeek-V3's highest accuracy, this also indicates that longer processing times facilitate the agent's completion of extraction tasks.

As a vital medical knowledge resource, the accuracy of traditional Chinese medicine knowledge graphs is paramount. When constructing and applying TCM knowledge graphs, the reliability of data sources must be considered. This study utilizes data from the Chinese Medicine Encyclopedia as the

original dataset. Entity extraction and triplet construction are performed using the KGMAI framework. This process must also ensure correctness, meaning both accuracy metrics should be high. Therefore, DeepSeek-V3 is selected as the base model for KGMAI to facilitate subsequent knowledge graph construction.

5.4. Ablation Study

To better demonstrate the accuracy of the KGMAI framework in constructing TCM knowledge graphs, this paper selected two highly representative large models—

DeepSeek-V3 and Moonshot-V1—as the foundational models for ablation experiments. DeepSeek-V3 achieved the highest entity extraction accuracy, while Moonshot-V1 delivered the fastest response speed. During the entity extraction phase of the KGMAI framework, agents operate in parallel without collaborative interaction or mutual influence. Leveraging this characteristic, ablation experiments primarily explored the specific impact of confidence and clarity mechanisms on the KGMAI framework by setting different confidence thresholds and clarity thresholds. Results are presented in Table 3.

Table 3. The results of the ablation experiment

Model	Threshold	$P_{Tri}\uparrow$	$R_{Tri}\uparrow$	F1 $\uparrow$
deepseek-v3	$\theta>0,\alpha>0$	0.9506	0.9496	0.9501
	$\theta>0.5,\alpha>0.5$	0.9677	0.9567	0.9622
	$\theta>0.7,\alpha>0.7$	0.9912	0.9889	0.9901
	$\theta>0.8,\alpha>0.8$	0.9917	0.9771	0.9844
	$\theta>0.9,\alpha>0.9$	0.9922	0.9686	0.9804
	$\theta>0.7,\alpha>0$	0.9897	0.9777	0.9837
	$\theta>0,\alpha>0.7$	0.9572	0.9504	0.9538
moonshot-v1	$\theta>0,\alpha>0$	0.9165	0.9056	0.9111
	$\theta>0.5,\alpha>0.5$	0.9198	0.9163	0.9181
	$\theta>0.7,\alpha>0.7$	0.9542	0.9523	0.9533
	$\theta>0.8,\alpha>0.8$	0.9589	0.9369	0.9479
	$\theta>0.9,\alpha>0.9$	0.9593	0.9334	0.9463
	$\theta>0.7,\alpha>0$	0.9362	0.9299	0.9331
	$\theta>0,\alpha>0.7$	0.9183	0.9101	0.9142

This paper conducted ablation experiments to test the combined effects of confidence and clarity thresholds, where θ and α represents the confidence and clarity thresholds, respectively. Threshold values were set at 0, 0.5, 0.7, 0.8, and 0.9. A total of seven combinations were designed. By comparing F1 scores, the optimal threshold combination was determined to be 0.7 (confidence) and 0.7 (clarity).

To evaluate the effectiveness of the confidence and clarity mechanisms, the optimal threshold of 0.7 was adopted as the experimental threshold. Setting the confidence threshold to 0.7 and the clarity threshold to 0 revealed that omitting the clarity mechanism caused a noticeable decline in triplet precision. However, this still yielded a significant improvement compared to setting both confidence and clarity to 0. Conversely, setting the confidence threshold to 0 and the clarity threshold to 0.7 yielded a similar trend, though the effect was less pronounced. This indicates that the confidence scoring mechanism is more effective than clarity for multi-agent tasks, but their combined use achieves the greatest improvement in multi-agent task correctness, demonstrating that both mechanisms are indispensable to the KGMAI framework.

6. Knowledge Graph Visualization

In experiments, KGMAI built with DeepSeek-V3 as the backbone model achieved an F1 score of 99.01%, extracting 7,925 entities and 19,992 triples from 1,000 data points. To ensure the accuracy of the TCM knowledge graph, this model was applied to construct a TCM knowledge graph from 10,268 herbal medicine records. The final output comprised 55,089 entities and 192,688 triples, demonstrating the KGMAI framework's effective extraction of latent knowledge from raw data. Results were stored in a Neo4j database for visualization. White fruit alternative names include ginkgo kernel, lingyan, Buddha's fingernail, Buddha's finger citrus, and duck foot seed. Its medicinal properties are neutral, with a taste profile of sweet, bitter, and astringent. It

is classified under the lung meridian and possesses three therapeutic effects: astringing the lungs to relieve asthma, stopping leukorrhea, and reducing frequent urination. It can treat symptoms such as excessive phlegm with wheezing and coughing, leukorrhea with turbid discharge, and enuresis with frequent urination. Ginkgo biloba is documented in 14 classical Chinese medical texts, including the Chinese Pharmacopoeia, the Great Dictionary of Chinese Medicinal Materials, and the Chinese Materia Medica.

Visualization of knowledge graphs not only presents specific information about a medicinal herb but also reveals direct relationships associated with any node. Furthermore, it can display node information related to indirect connections between nodes, as shown in Figure 4. The medicinal White fruit can treat enuresis and frequent urination. This visualization can display other Chinese medicinal herbs used to treat this condition, such as Deer antler powder, Chinese chive seeds, Dvodia root, Mantis egg case, and Chinese rosehip.

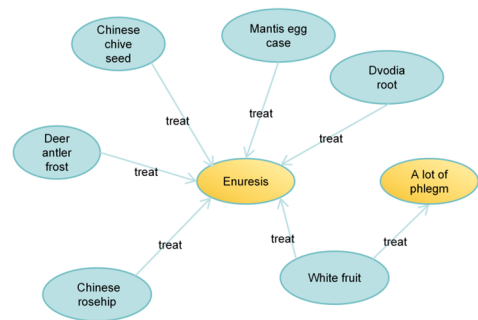


Fig 4. Visualization of Indirect Relationships

This knowledge graph-based visualization analysis not only helps researchers quickly understand the complex relationships among Chinese medicinal herbs but also provides robust data support for modern TCM research. The knowledge graph can also be integrated with clinical data,

pharmacology experimental data, and other sources to build a more comprehensive TCM knowledge service platform. For example, by combining patient symptom information, the system can recommend herbs with similar therapeutic effects to assist TCM clinical decision-making.

7. Conclusion

This paper proposes the KGMAL framework to address accuracy challenges in TCM knowledge graph construction through multi-agent collaboration. Different agents extract distinct entity types while assigning confidence scores to task completion, ensuring agent stability and entity extraction accuracy. Experimental validation demonstrates that multi-agent collaboration overcomes limitations of single-agent processing, particularly offering greater scalability for complex tasks. The resulting TCM knowledge graph is visualized using Neo4j, clearly illustrating relationships among medicinal herbs. This graph serves as a foundational database for TCM question-answering systems, integrating TCM knowledge to enhance learning and comprehension of TCM theories, thereby promoting the inheritance and development of TCM culture.

Although the KGMAL framework achieved an F1 score of 99.01% in experimental settings, practical deployment faces challenges such as high computational resource consumption and model hallucination. These issues primarily stem from the high computational demands of multi-agent coordination mechanisms and the inherent uncertainty of large language models. Training and inference costs escalate significantly when processing large-scale data. Model distillation techniques can be employed to reduce model size and computational overhead. The hallucination issue of large models cannot be overlooked during this process. When generating entities or relationships, false information may be produced, thereby affecting the quality of the knowledge graph. Beyond the confidence and clarity filtering methods used in this paper, knowledge augmentation techniques can be employed to validate extracted content by integrating authoritative knowledge bases.

The KGMAL framework offers substantial advantages for information extraction tasks. Its modular design enables strong transferability, allowing rapid adaptation to knowledge graph construction needs across diverse domains such as acupuncture points, meridian systems, and traditional

Chinese wellness practices. This scalability also facilitates future expansion to other agent types for constructing knowledge graphs in additional domains. For instance, integrating formula analysis agents and formula combination recognition agents could enable the construction of formula knowledge graphs. Looking ahead, with continuous technological advancements and accumulating data, the KGMAL framework is poised to play an increasingly vital role in the intelligent development of traditional Chinese medicine. It will support the inheritance, innovation, and international dissemination of TCM knowledge.

Acknowledgments

Research Project on the Reform of Graduate Education and Teaching in Sichuan University of Science and Engineering (No. JG202440).

References

- [1] Xindi Huang, Changsong Ding, Hao Liang, et al. Research on the construction of TCM pattern differentiation and treatment model on the theory of herbal properties[J]. Journal of Beijing University of Traditional Chinese Medicine, 2019, 42(6): 5. DOI:10.3969/j.issn.1006-2157.2019.06.003.
- [2] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. ArXiv, 2019, abs/1810.04805. DOI:10.18653/v1/N19-1423.
- [3] Carvalho A, de Paiva D S, Sichman J S, et al. Multi-agent systems for natural language processing[C]//Proceedings of the Second Iberoamerican Workshop on Distributed Artificial Intelligence and Multi-agent Systems. 1998.
- [4] Lu Y, Wang J. KARMA: Leveraging Multi-Agent LLMs for Automated Knowledge Graph Enrichment[J]. arxiv preprint arxiv:2502.06472, 2025.
- [5] GLM T, Zeng A, Xu B, et al. Chatglm: A family of large language models from glm-130b to glm-4 all tools[J]. arxiv preprint arxiv:2406.12793, 2024.
- [6] Yang M, Diao M, Luo J, et al. GLM-4 Based Method for Automatic Construction of Content Graph[J]. IEEE Access, 2025.
- [7] Liu A, Feng B, Xue B, et al. Deepseek-v3 technical report[J]. arxiv preprint arxiv:2412.19437, 2024.