Zhengzhou Commodity Exchange Information Technology Department I Internship Report

Intern: Jingfang Gu

Internship Project: Intelligent Data QA

Department: Information Technology Department I

Supervisors: Dr. Zhen Xu, Ms. Na Song

September 10, 2025

Abstract

This report provides a comprehensive summary of the work undertaken and achievements made during my internship at the Information Technology Department I of Zhengzhou Commodity Exchange (CZCE), specifically focusing on the "Intelligent Data QA" project. The internship concentrated on data layer optimisation to enhance the application capabilities of Large Language Models in the futures market.

The report outlines four core contributions: **(1) Named Entity Recognition (NER) Model Performance Benchmarking**: Through comprehensive evaluation of spaCy Chinese models, we selected an efficient small language model that balances performance and efficiency for the project; **(2) Futures Contract Code Recognition Enhancement**: Designed and implemented a customised solution combining regular expressions with spaCy NER models, significantly improving the system's recognition accuracy for diverse contract code formats (from 0% to 100% accuracy); **(3) FAQ Knowledge Base Semantic Retrieval and Intelligent Matching**: Introduced a semantic matching mechanism based on cosine similarity, creating a "rapid response channel" for frequently asked queries, optimising user experience; **(4) RAG Workflow Logic Optimisation**: Designed and implemented a parallel computing strategy based on FAQ and API matching similarity scores, effectively improving system response speed.

The report not only demonstrates specific technical implementation details and achievements but also shares the challenges encountered during this process, solutions developed, and personal growth in technical and problem-solving capabilities. This internship combined cutting-edge AI technology with real futures business scenarios, accumulating valuable practical experience for exploring the application of Large Language Models in the financial futures domain.

Acknowledgements

I would like to extend my heartfelt gratitude to Dr. Zhen Xu, the leader of Information Technology Department I at Zhengzhou Commodity Exchange, for selecting me from numerous candidates and providing this invaluable internship opportunity. Special thanks go to my supervisor Ms. Na Song for her dedicated guidance and patient assistance throughout the internship period. From initial project requirement briefings to technical solution discussions, and detailed code reviews, her profound professional knowledge, rigorous work attitude, and open-minded mindset have been tremendously beneficial to my development.

I am grateful to all the senior colleagues in the department who provided support and encouragement when I encountered challenges. The harmonious and collaborative team atmosphere enabled me to integrate quickly and dedicate myself fully to the work. This experience has not only enhanced my professional skills but also given me a fresh perspective on the financial technology industry. Once again, thanks to all the leaders and colleagues who helped me!

Though this journey ends, the bonds formed shall endure.

The breeze and bright moon invite me to drink, treading through all the dust in the world without forgetting the original aspiration!

Contents

1 Project Background and Personal Responsibilities

1.1 Project Background

With the rapid development of Large Language Model (LLM) technology, intelligent transformation in the financial industry has become an inevitable trend. Zhengzhou Commodity Exchange (CZCE), as an important national financial infrastructure, actively pursues cutting-edge artificial intelligence fields. The "Intelligent Data QA" project was initiated as one of the core research and development projects under this context. The project aims to construct an intelligent question-answering system capable of deeply understanding professional knowledge in the futures domain and converting users' natural language questions into precise SQL queries (i.e., Text2SQL), thereby empowering business personnel, researchers, and market participants to improve the efficiency of data acquisition and analysis.

The project's overall architecture is based on advanced Retrieval-Augmented Generation (RAG) technology, with workflows encompassing multiple key stages including question understanding/rewriting, entity recognition, knowledge base retrieval, SQL generation, and result presentation.

1.2 Personal Responsibilities and Work Focus

During this internship, I primarily focused on data layer optimisation work within the project. My core responsibility was to improve the accuracy of the system's understanding of user query intentions and optimise the efficiency of information retrieval and processing. Specific work focus areas included the following four aspects:

1. **NER Model Selection**: Evaluating the performance of different scales of NER models in futures scenarios to select the optimal technical solution for the project.

2. **Entity Recognition Enhancement**: Addressing the insufficient recognition capability of general models for domain-specific entities such as futures contract codes.

3. **FAQ Knowledge Base Semantic Retrieval**: Developing intelligent matching functionality to provide rapid and precise SQL answers for common questions.

4. **Workflow Logic Optimisation**: Designing and implementing dynamic response logic based on FAQ and API matching scores, and optimising the retrieval process to improve user experience.

2 Core Work and Technical Implementation

2.1 Named Entity Recognition (NER) Model Performance Benchmarking

Entity recognition is a crucial step in understanding user query intentions. To select the most suitable NER engine for this project, I conducted a comprehensive performance benchmark test of spaCy's Chinese small (sm), medium (md), and Transformer large model (trf).

Testing Method: I constructed a corpus containing 68 typical sentences from the futures domain, covering various entity types including exchanges, futures companies, commodities, and contract codes. Testing dimensions included average latency, processing speed, total entity count discovered, and model size.

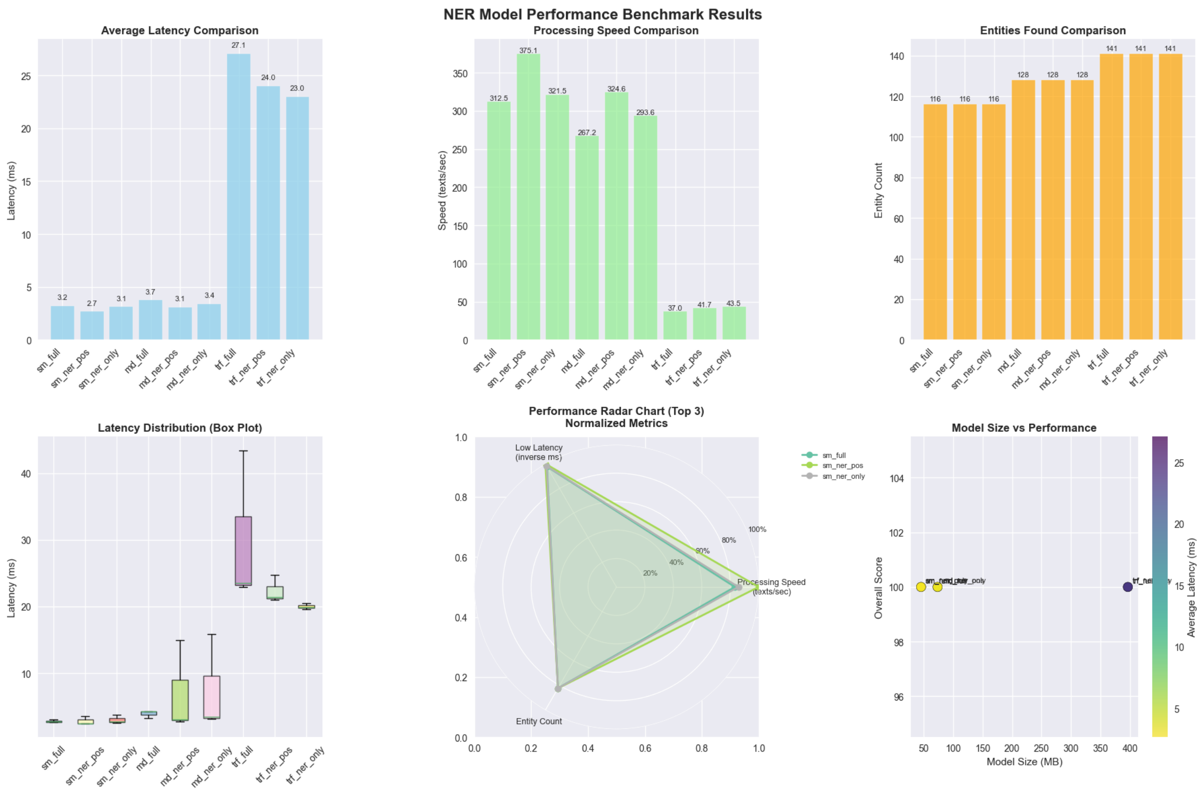

Figure 1: spaCy Chinese NER Model Performance Benchmark Comparison Results

Testing Results: The test results (as shown in Figure 1 above) clearly demonstrate:

- **Large Model (Transformer)** performed best in entity discovery quantity but had the highest latency, slowest speed, and largest model size.
- **Small Model (sm)** excelled in latency and speed performance. Although it discovered slightly fewer entities than the medium and large models, its comprehensive performance advantages were extremely significant.

Considering the real-time question-answering system's requirements for low latency and high performance, as well as deployment costs, we ultimately selected the **small model (zh_core_web_sm)** as the project's foundational NER engine, achieving the optimal balance between performance and resource consumption.

2.2 Futures Contract Code Recognition Enhancement

During testing, we discovered that general NER models had difficulty accurately recognising futures contract codes with diverse formats in the futures market. To address this issue, I designed and implemented a hybrid enhancement solution combining regular expressions with NER models. The core logic is shown in Algorithm 1.

Algorithm 1 Contract Code Recognition Enhancement Algorithm (‘entity_recognition‘ & ‘_expand_short_contracts‘)

Input: Original query string $query$
Output: Enhanced query string $enhanced\_query$

$enhanced\_query \leftarrow$ RegexExpandShortContracts($query$) ▷ Step 1: Expand short codes, e.g., AP509 → AP2509
$entities \leftarrow$ SpaCyNERModel($enhanced\_query$) ▷ Step 2: General entity recognition
for all $entity$ in $entities$ **do**
  if $entity.label =$ ’Contract’ **then**
    $normalized\_code \leftarrow$ ToUpper(RemoveSeparators($entity.text$)) ▷ Step 3: Format normalisation
    $enhanced\_query \leftarrow$ Replace($enhanced\_query, entity.text,$ "(normalized_code)(Contract)")
  end if
end for
return $enhanced\_query$

The implementation of this solution enabled the system to accurately recognise most contract code formats (as shown in Table 1), providing accurate entity input for subsequent table positioning and SQL generation, effectively improving the robustness of the entire data question-answering process.

Table 1: Contract Code Recognition Enhancement Effect Examples

User Original Input	Enhanced Recognition Result
Apple futures AP501 position volume	Apple(Variety) futures AP2501(Contract) position volume
Cotton cf_2409 contract price	Cotton(Variety) CF2409(Contract) price
Rebar rb-409 opening price	Rebar(Variety) RB2409(Contract) opening price
Iron ore i/2501 highest price	Iron ore(Variety) I2501(Contract) highest price
Shanghai copper cu.2409 lowest price	Shanghai copper(Variety) CU2409(Contract) lowest price

2.3 FAQ Knowledge Base Semantic Retrieval and Intelligent Matching

To improve user experience for frequently asked queries and reduce model calling costs, I was responsible for designing and implementing the FAQ (Frequently Asked Questions) semantic retrieval module. The core objective was: when user queries are highly similar to existing questions in the knowledge base (similarity $>=90\%$), directly return pre-configured, validated SQL statements. The core logic of this functionality is shown in Algorithm 2.

Algorithm 2 FAQ Semantic Retrieval Algorithm (‘semantic_search_faq‘)

Input: User query string $query$, return count k
Output: Top-K similar FAQ result list $top\_k\_results$

$query\_embedding \leftarrow$ Embedder.GetEmbedding$(query)$
$similarities \leftarrow []$
for all faq in $FAQ\_KnowledgeBase$ **do**
    $score \leftarrow$ CosineSimilarity$(query_embedding, faq.embedding)$
    Add $(faq, score)$ to $similarities$
end for
$sorted\_faqs \leftarrow$ SortByScore$(similarities,$ descending$)$
$top\_k\_results \leftarrow sorted\_faqs[0 \ldots k-1]$
return $top\_k\_results$

The introduction of this module established an efficient "rapid response channel" for the system.

2.4 RAG Workflow Logic Optimisation and Parallelisation

Building upon the FAQ semantic retrieval implementation, I optimised the core ‘do_generate‘ workflow logic and introduced parallel processing mechanisms to reduce latency.

2.4.1 Dynamic Routing Based on FAQ Similarity Scores

I designed a dynamic response strategy based on FAQ matching similarity scores (as shown in the lower right of Figure 2), dividing the question-answering process into three pathways: high-score direct access, medium-score reference, and low-score standard pathway. This mechanism achieved a balance between system intelligence and resource efficiency.

2.4.2 Data Question-Answering Retrieval Process Optimisation

To further optimise performance, I parallelised FAQ retrieval and API positioning, two independent time-consuming operations. The final workflow logic is shown in Figure 2 and Algorithm 3.

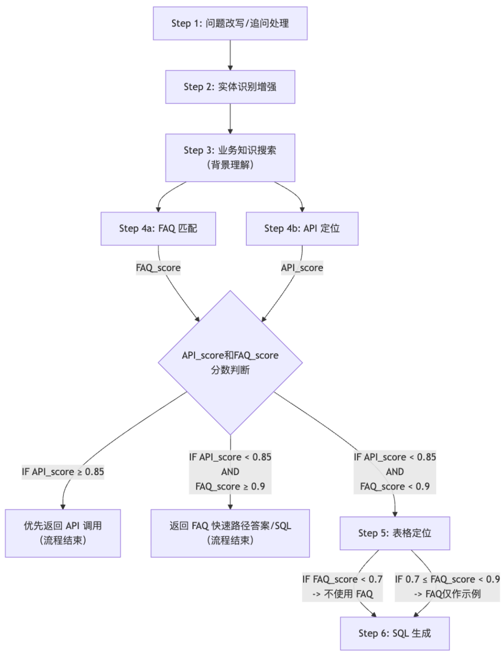

Figure 2: RAG Workflow Logic

Algorithm 3 Data Question-Answering Dynamic Routing and Parallelised RAG Workflow (‘do_generate‘)

Input: User query string $user\_query$
Output: Final answer $final\_answer$

$query \leftarrow$ QueryEnhancement$(user_query)$ ▷ Steps 1&2&3
in parallel do
    $(faq\_results, faq\_score) \leftarrow$ SemanticSearchFAQ$(query)$ ▷ Step 4a
    $(api\_results, api\_score) \leftarrow$ LocateAPI$(query)$ ▷ Step 4b
end parallel
if $api\_score \geq 0.85$ **then**
    return FormatApiResult$(api_results)$ ▷ Path 1: API priority
else if $faq\_score \geq 0.9$ **then**
    return $faq\_results[0].sql$ ▷ Path 2: High-score FAQ direct access
else
    $table\_schema \leftarrow$ LocateTable$(query)$ ▷ Step 5: Standard pathway
    if $faq\_score \geq 0.7$ **then**
        $prompt \leftarrow$ BuildPromptWithFAQ$(query, table_schema, faq_results)$
    else
        $prompt \leftarrow$ BuildPromptWithoutFAQ$(query, table_schema)$
    end if
    $final\_answer \leftarrow$ LLM.GenerateSQL$(prompt)$ ▷ Step 6
    return $final\_answer$
end if

3 Internship Summary and Personal Achievements

This internship at Zhengzhou Commodity Exchange has been a valuable experience that deeply combines theory with practice. Through participation in the "Intelligent Data QA" project, I not only applied knowledge from natural language processing and machine learning acquired at university to real financial business scenarios, but also developed engineering capabilities and innovative thinking through solving practical problems.

Main Achievements and Growth:

- **Technical Depth**: Gained deep understanding of RAG system core components and optimisation strategies, mastering the complete development process from model evaluation, data processing to algorithm optimisation. Particularly in handling domain-specific NLP tasks, I accumulated practical experience in combining general technology with business knowledge.
- **Engineering Capabilities**: Experienced the importance of code performance, robustness, and maintainability in industrial-grade projects. Through parallelisation transformation and other work, I gained more intuitive understanding of system performance optimisation.
- **Problem-Solving Abilities**: Faced with seemingly minor but crucial specific problems like "futures contract code recognition", I learned to analyse problems, consult literature, and creatively design hybrid solutions.
- **Industry Knowledge**: Through project requirements, I gained deeper understanding of futures market business logic, data characteristics, and information processing pain points, laying a solid foundation for my future learning and work.

Looking forward, this project still has broad optimisation potential, such as introducing more advanced re-ranking models to improve retrieval accuracy, exploring the possibility of fine-tuning foundational models for specific tasks, providing brief sentiment analysis reports for information and documents returned from user queries to assist decision-making processes, and more.

I believe that with the team's continued efforts in the future, the "Intelligent Data QA" project will become an important tool for empowering the digital transformation of the futures industry!