Table 1: The prompt of Selector agent.

system: You are a helpful, respectful and honest assistant. Your task is to output the IDs of the candidate Documents (0,1,2,...,K-1) which are helpful in answering the Question.

assistant: Okay, I will provide the ID of candidate Documents which are helpful in answering the Question.

user: Question: {content of Question}
Document0: {content of Document0}
⋮
Document(K-1): {content of Document(K-1)}

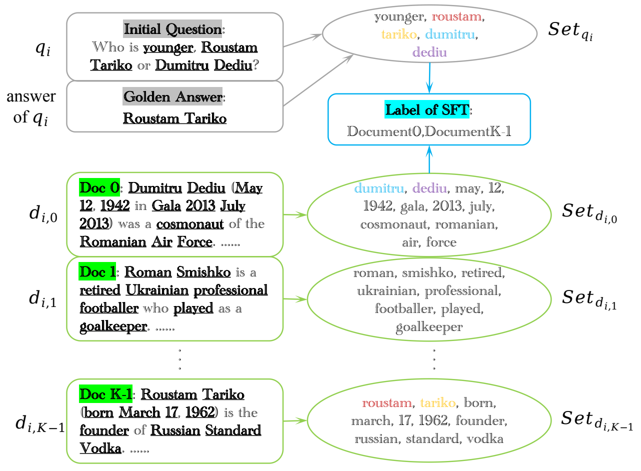
assistant: OK, I received the Question and the candidate Documents.

user: Now, output the IDs of the candidate Documents (0,1,2,...,K-1) which are helpful in answering the Question: {content of Question}, for example, in the following format: Document0,Document4,Document6,Document7.

4.3.1 Query Rewriter. In Rewrite-Retrieve-Read [25], a small language model was trained using PPO to effectively rewrite queries for RAG. Building on this approach, we utilize the publicly available query rewriting data from Rewrite-Retrieve-Read as the SFT dataset to warm start the Query Rewriter in MMOA-RAG.

4.3.2 Selector. The task of the Selector is to choose a subset D_{selected} that are helpful for answering a question from a given set D with K candidate documents. The output format of the Selector is the IDs of the documents in D_{selected} (e.g., Document0, Document4, Document6, Document7), as shown in the prompt of Selector in Table 1. Therefore, to construct SFT data for the Selector, the ground truth should be the IDs of documents that are truly useful for answering the question. One method to obtain the ground truth is to employ advanced LLMs, such as GPT-4o, to provide the ground truth. However, we have found that this approach does not yield results as good as expected. Additionally, BGM [19] introduced and optimized a bridge module which is similar to the Selector module. They proposed a method called synthesis silver passage sequence (Synthesis SPS) to construct the ground truth for SFT data. However, the Synthesis SPS method requires examining each candidate document $d_{i,j} \in D_i$ (candidate documents of question i), invoking the LLM for each check, and comparing the utility values before and after the check, making it a complex and costly method.

We propose a convenient heuristic approach for constructing SFT data, aimed at LLMs to effectively follow instructions and output in the desired format. As illustrated in Figure 2, for a given question q_i and its golden answer, there are K candidate documents denoted as $d_{i,j}$, where $j \in \{0, 1, \cdots, K-1\}$. First, by removing certain insignificant stop words and punctuation marks from q_i and its golden answer, and converting the words to lowercase, we obtain the set Set_{q_i}. Similarly, we perform the same operation on the K candidate documents $d_{i,j}$ to obtain $Set_{d_{i,j}}$. Finally, if any word from Set_{q_i} appears in $Set_{d_{i,j}}$, the ID of corresponding



q_i
Initial Question: Who is younger, Roustam Tariko or Dumitru Dediu?
answer of q_i
Golden Answer: Roustam Tariko
younger, roustam, tariko, dumitru, dediu
Set_{q_i}
Label of SFT: Document0,DocumentK-1
$d_{i,0}$
Doc 0: Dumitru Dediu (May 12, 1942 in Gala 2013 July 2013) was a cosmonaut of the Romanian Air Force. ......
dumitru, dediu, may, 12, 1942, gala, 2013, july, cosmonaut, romanian, air, force
$Set_{d_{i,0}}$
$d_{i,1}$
Doc 1: Roman Smishko is a retired Ukrainian professional footballer who played as a goalkeeper. ......
roman, smishko, retired, ukrainian, professional, footballer, played, goalkeeper
$Set_{d_{i,1}}$
$d_{i,K-1}$
Doc K-1: Roustam Tariko (born March 17, 1962) is the founder of Russian Standard Vodka. ......
roustam, tariko, born, march, 17, 1962, founder, russian, standard, vodka
$Set_{d_{i,K-1}}$

Figure 2: The convenient approach to construct the SFT data for Selector.

document j is included in the final output as the Label of SFT. With this approach, we can rapidly and cost-effectively construct the Selector's ground truth labels during the SFT stage. Given our focus on the subsequent joint optimization of multiple modules, this straightforward data construction method can adequately meet our requirements.

4.3.3 Generator. The Generator is responsible for producing the final answer, Ans_{predict}, based on the D_{selected} provided by the Selector. Therefore, the ground truth for the SFT data of Generator is the golden answer Ans_{golden}.

With these approaches, the SFT data used for the warm start training of these three modules—Query Rewriter, Selector, and Generator—can be obtained. All modules can be fine-tuned using the typical loss function of SFT presented in Equation (10).

$$\mathcal{L}_{\text{SFT}}(\theta) = -\sum_{n=1}^{N} \log P(Y_i \mid X_i; \theta) \tag{10}$$

In Equation (10), N represents the number of samples in the SFT dataset, while θ denotes the parameters of the LLM. The variable X_i corresponds to the input content of each module. Meanwhile, Y_i signifies the output content of each module.

4.4 Multi-Agent Optimization

After undergoing SFT, the LLM demonstrates an improved ability to follow instructions while executing the functions of Query Rewriter, Selector, and Generator. The RAG system also achieves relatively satisfactory warm-start performance. To further enhance the performance of the RAG system, which is modeled as a fully cooperative multi-agent system, it is crucial to conduct joint training of multiple agents to strengthen collaboration among them.

We adopt a setup similar to Multi-Agent PPO [45] in Starcraft II, where multiple agents share a global reward, that is to optimize $\mathcal{G}$ = {Query Rewriter (QR), Selector (S), Generator (G)} with R_{shared}. To reduce computational overhead, we apply the parameter-sharing mechanism among agents, allowing QR, S, and G to utilize the same LLM.