A New Query Expansion Approach for Enhancing Information Retrieval via Agent-Mediated Dialogic Inquiry

Wonduk Seo, Hyunjin An, Seunghyun Lee
AI Research, Enhans

1. MOTIVATION

LLM-based reformulations often collapse into homogeneity, missing diverse context. Lacking dynamic feed-back keeps noise. Term document-only tweaks skip intent structure. We reveal a dialogic process that widens perspective while preserving relevance.

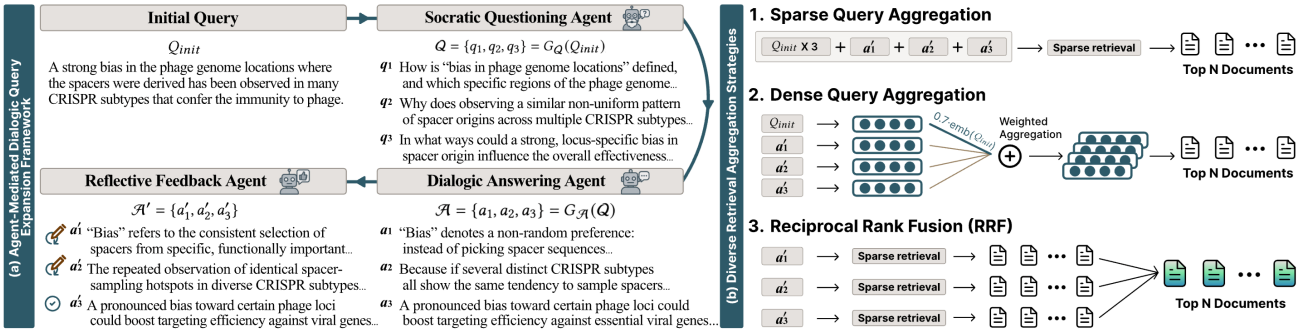
2. SOCRATIC QUESTIONING

We reformulate Q_{init} into 3 targeted sub-questions.

Query Type	Role in Sub-question
Clarification	Crafts a sub-question to refine intent and ensure accurate interpretation.
Assumption Probing	Decomposes the query by surfacing implicit assumptions, adding diversity.
Implication Probing	Explores downstream effects to expand with relevant and diverse information.

3. AMD FRAMEWORK

AMD orchestrates three agents to interrogate, answer, and refine queries. This Socratic-to-feedback loop crafts richer representations and plugs into sparse, dense, and RRF retrieval, boosting effectiveness without task-specific finetuning.



4. EXPERIMENTAL SETUP

1. **Benchmarks**: BEIR (6 sets) and TREC DL'19/'20.
2. **LLM**: Qwen2.5-7B-Instruct (temp 0.5, max context length: 512).
3. **Retrievers**: BM25 (bm25s) and E5-base.
4. **Baselines**: Q2D, Q2C, GenQREnsemble, GenQRFusion.

Sparse Retrieval:

$$Q^*_{\text{sparse}} = \sum_{i=1}^{3} Q_i \oplus \sum_{j=1}^{|\mathcal{S}|} a'_j,$$

Dense Retrieval:

$$q^* = 0.7 \cdot \text{emb}(Q_{init}) + 0.3 \cdot \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \text{emb}(a'_i).$$

Reciprocal Rank Fusion:

$$\text{score}(d) = \sum_{i=1}^{|\mathcal{S}|} \frac{1}{k + r_{i,d}},$$

5. RESULTS

AMD lifts averages:
(1) BEIR nDCG@10 0.4352 (sparse), **0.4707** (dense), 0.4113 (RRF).
(2) **TREC DL'19** avg 0.5422/**0.6605**/0.4883;
(3) TREC DL'20 avg 0.5433/**0.6488**/0.5077.

Methods	BEIR Benchmark (nDCG@10)							TREC DL'19			TREC DL'20		
	Webis	SciFact	TREC-COVID	DBPedia	SciDocs	Fiqa	Avg. Score	nDCG@10	R@1000	Avg. Score	nDCG@10	R@1000	Avg. Score
Sparse Retrieval Results													
BM25	0.2719	0.6694	0.5868	0.2831	0.1592	0.2326	0.3672	0.4239	0.3993	0.4116	0.4758	0.4240	0.4500
Q2C (2023)	0.3546	0.6876	0.6954	0.3252	0.1661	0.2595	0.4147	0.5439	0.4814	0.5127	0.5357	0.4941	0.5149
Q2D (2023)	0.3679	0.6794	0.6957	**0.3378**	0.1637	0.2712	0.4193	0.5732	0.4890	0.5311	0.5486	0.4958	0.5222
GenQREnsemble (2024)	0.2887	0.5560	0.5104	0.2302	0.1058	0.2017	0.3155	0.4109	0.4110	0.4110	0.4261	0.4163	0.4207
AMD* (Sparse, Ours)	**0.3896***	**0.7021***	**0.7115***	0.3352	**0.1834***	**0.2896***	**0.4352***	**0.5870***	**0.4974***	**0.5422***	**0.5818***	**0.5047***	**0.5433***
Dense Retrieval Results													
E5-Base	0.1786	0.6924	0.7098	0.4002	0.2326	0.3808	0.4324	0.7020	0.5185	0.6103	0.7029	0.5648	0.6339
Q2C (2023)	0.1841	0.7028	0.7238	0.4250	0.2595	0.4057	0.4502	0.5517	0.4891	0.5204	0.7084	0.5715	0.6400
Q2D (2023)	0.1931	0.7108	0.7284	0.4229	0.2712	0.4094	0.4560	**0.7472**	0.5565	0.6519	0.6971	0.5799	0.6385
AMD* (Dense, Ours)	**0.1985**	**0.7324***	**0.7493***	**0.4435***	**0.2871***	**0.4135***	**0.4707***	0.7458	**0.5752***	**0.6605***	**0.7128***	**0.5847***	**0.6488***
RRF Fusion (BM25-based) Results													
GenQRFusion (2024)	**0.3815**	0.6518	0.6594	0.2726	0.1436	0.2293	0.3897	0.4418	0.4205	0.4312	0.4375	0.4654	0.4515
AMD* (RRF, Ours)	0.3749	**0.6764***	**0.6703***	**0.3078***	**0.1749***	**0.2632***	**0.4113***	**0.5041***	**0.4724***	**0.4883***	**0.5325***	**0.4819***	**0.5077***

6. CONCLUSION

In this paper, we presented AMD, a novel agent-mediated framework for query expansion that combines Socratic Questioning, Dia logic Answering, and Reflective Feedback Agents. By reformulating queries into diverse sub-questions and iteratively refining pseudo- answers, AMD effectively produces robust, intent-aligned query representations. Extensive experiments on benchmark datasets confirm that our specially designed agent collaboration leads to consistent and significant improvements in retrieval performance.