Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

RAC : Clarification Augmentée par la Recherche pour une Recherche Conversationnelle Fidèle

Ahmed Rayane Kebir[1, 2] Vincent Guigue[3]
Lynda Said Lhadj [2] Laure Soulier[1]
(1) Sorbonne Université, CNRS, ISIR, F-75005 Paris, France
(2) Ecole nationale Supérieure d'Informatique (ESI), Alger, Algérie
(3) AgroParisTech, UMR MIA-PS, Palaiseau, France
kebir@isir.upmc.fr

RÉSUMÉ

Les questions de clarification permettent aux systèmes de recherche conversationnelle de lever les ambiguïtés des requêtes utilisateurs. Les travaux existants se concentrent principalement sur la fluidité et l'alignement avec l'intention, tandis que l'ancrage des clarifications dans le corpus reste peu étudié. Sans cet ancrage, les systèmes risquent de poser des questions auxquelles les documents disponibles ne permettent pas de répondre. Nous proposons RAC (Retrieval-Augmented Clarification), un cadre visant à générer des clarifications fidèles au corpus. Après avoir comparé plusieurs stratégies d'indexation, nous ajustons finement un modèle de langue afin d'exploiter le contexte de recherche et de favoriser des questions fondées sur des preuves. Une optimisation contrastive par préférences privilégie les questions appuyées par des passages récupérés. Les évaluations sur quatre jeux de données de référence montrent des améliorations significatives. Nous introduisons de nouvelles métriques basées sur la NLI et le data-to-text pour mesurer l'ancrage contextuel et la fidélité.

ABSTRACT

RAC : Retrieval-Augmented Clarification for Faithful Conversational Search

Clarification questions help conversational search systems resolve ambiguous or underspecified user queries. While prior work has focused on fluency and alignment with user intent, especially through facet extraction, less attention has been paid to grounding clarifications in the underlying corpus. Without such grounding, systems risk asking questions that cannot be answered from available documents. We introduce RAC (Retrieval-Augmented-Clarification), a framework for generating corpus-faithful clarification questions. After comparing several indexing strategies for retrieval, we fine-tune a large language model to make optimal use of research context and to encourage the generation of evidence-based question. We then apply contrastive preference optimization to favor questions supported by retrieved passages over ungrounded alternatives. Evaluated on four benchmarks, RAC demonstrate significant improvements over baselines. In addition to LLM-as-Judge assessments, we introduce novel metrics derived from NLI and data-to-text to assess how well questions are anchored in the context, showing that our approach consistently enhances faithfulness.

MOTS-CLÉS : Recherche conversationnelle, Clarifications, RAG.

KEYWORDS: Conversational Search, Clarifying Questions, RAG.

ARTICLE ACCEPTÉ À : ECIR 26. (https://arxiv.org/abs/2601.11722)

Actes de la 21e Conférence en Recherche d'Information et Applications.
Volume 2 : articles déjà publiés, pages 14–14.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.