An Evaluation Framework for Attributed Information Retrieval using Large Language Models - Abstract

Hanane Djeddal*†[1,2], Pierre Erbacher[3], Raouf Toukal[3], Laure Soulier[3], Karen Pinel-Sauvagnat[1], Sophia Katrenko[2], and Lynda Tamine[1]

[1]Institut de recherche en informatique de Toulouse (IRIT) – CNRS, University Paul Sabatier Toulouse III – 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

[2]Ecovadis – Ecovadis – France

[3]Institut des Systèmes Intelligents et de Robotique (ISIR) – Sorbonne Universités, UPMC, CNRS – Sorbonne-Université, Boite courrier 173 4 Place JUSSIEU 75252 Paris cedex 05, France

Résumé

With the growing success of Large Language models (LLMs) in information-seeking scenarios, search engines are now adopting generative approaches to provide answers along with in-line citations as attribution. While existing work focuses mainly on attributed question answering, in this work, we target information-seeking scenarios which are often more challenging due to the open-ended nature of the queries and the size of the label space in terms of the diversity of candidate-attributed answers per query. We propose a reproducible framework to evaluate and benchmark attributed information seeking, using any backbone LLM, and different architectural designs: (1) Generate (2) Retrieve then Generate, and (3) Generate then Retrieve. Experiments using HAGRID, an attributed information-seeking dataset, show the impact of different scenarios on both the correctness and attributability of answers.

Mots-Clés: Large Language Models, Retrieval Augmented Generation (RAG), Answer generation, Attributed information seeking

*Intervenant

†Auteur correspondant: fh.djeddal@gmail.com