

Vers des LLM moins gourmands pour interroger des graphes de connaissances en langue naturelle

Marion Schaeffer*^{**} Christophe Bouvard*
Jean-Philippe Kotowicz*^{**} Cecilia Zanni-Merk**

*Wikit

prenom@wikit.ai, <https://www.wikit.ai/>

**LITIS - INSA Rouen Normandie

prenom.nom@insa-rouen.fr, <https://www.litislab.fr/>

Résumé. Les graphes de connaissances sont largement utilisés pour définir et représenter un domaine. Ils permettent de réunir des données issues de diverses sources et de les stocker de façon structurée. L'accès à ces ressources peut cependant s'avérer complexe car il faut maîtriser le langage de requête compatible avec le formalisme utilisé pour représenter le graphe de connaissances. Dans cet article, nous proposons un système conversationnel de questions-réponses utilisant les grands modèles de langage pour interroger des graphes de connaissances en langue naturelle. Nous comparons différentes techniques afin de combiner au mieux frugalité et performance dans le but de maîtriser les ressources utilisées. Notre travail se base sur un graphe de connaissances décrivant les Objectifs du Développement Durables définis par les Nations Unies.

1 Introduction

Un graphe de connaissances est défini par Hogan et al. (2020) comme un graphe de données destiné à accumuler et à transmettre des connaissances sur le monde réel, dont les nœuds représentent des entités d'intérêt et dont les arêtes représentent des relations entre ces entités. Ils constituent une riche source d'informations et permettent de construire diverses applications, comme celles listées par Hogan et al. (2020). Le schéma définit la structure de haut niveau du graphe de connaissances et permet de lier des données issues de sources variées.

Malgré leur utilité, les graphes de connaissances sont difficilement accessibles pour des utilisateur·rice·s finaux·ales car ils nécessitent de maîtriser le langage de requête du formalisme dans lequel ils sont construits. C'est pour cela que des systèmes de questions-réponses (*Question Answering*, QA) ont été largement développés, comme expliqué par Liang et al. (2020). Ils permettent aux utilisateur·rice·s finaux·ales de poser leur question directement en langue naturelle. Nous nous intéressons aux systèmes de QA sur des domaines fermés utilisant des graphes de connaissances spécifiques à un domaine, contrairement aux systèmes de QA en domaine ouvert qui utilisent plutôt des graphes de connaissances génériques comme DBpedia, Lehmann et al. (2015), ou Wikidata, Vrandečić et Krötzsch (2014).

Un cas concret de cette difficulté d'accès à l'information est le SustainGraph, proposé par Fotopoulou et al. (2022), qui regroupe des connaissances autour des Objectifs du Développe-