

# AlasQA : Système Neurosymbolique de Questions-Réponses sur Graphes de Connaissances

Baptiste Amice\*, Peggy Cellier\*, Sébastien Ferré\*

\*Univ Rennes, CNRS, Inria, Insa Rennes, IRISA  
prenom.nom@irisa.fr

**Résumé.** Nous nous intéressons à la tâche d’interrogation d’un graphe de connaissances en langue naturelle. Les graphes de connaissances peuvent être interrogés de façon fiable via des langages formels comme SPARQL. Toutefois cela exige une traduction complexe du langage naturel vers le langage formel. Les grands modèles de langue (LLM) sont capables de répondre directement aux questions en langue naturelle mais n’offrent aucune garantie concernant la validité des réponses qu’ils génèrent. Nous proposons un système neurosymbolique, appelé AlasQA, qui répond à des questions posées en langue naturelle en combinant la fiabilité d’un langage formel comme SPARQL et la puissance des LLMs. La proposition s’appuie sur un outil intermédiaire de construction interactive de requêtes SPARQL. Des expérimentations menées sur les jeux de données QALD et TEXT2SPARQL valident l’intérêt de cette approche hybride.

## 1 Introduction

Le développement de systèmes de questions-réponses pour l’interrogation de graphes de connaissances (KGQA) est une tâche importante à la croisée du traitement automatique du langage naturel (TALN) et du web sémantique. Contrairement aux moteurs de recherche traditionnels, qui renvoient une liste de documents pertinents, les systèmes de KGQA visent à fournir directement une réponse concise et exacte à une question formulée en langue naturelle en s’appuyant sur les connaissances contenues dans des graphes comme DBpedia ou Wikidata. La tâche de KGQA nécessite donc à la fois de comprendre la question formulée en langage naturel et d’extraire l’information pertinente d’un graphe de connaissances pour y répondre.

Des approches symboliques ont été proposées. Elles traduisent les questions en langage naturel en requêtes SPARQL via des patrons, des règles linguistiques ou des modèles de traduction syntaxique, e.g. (Unger et al., 2012). Si ces approches garantissent précision et traçabilité, elles sont sensibles à la variabilité du langage et difficilement adaptables à grande échelle. Des travaux plus récents s’intéressent au développement de systèmes hybrides croisant LLMs et web sémantique (Pan et al., 2024). Des systèmes tels que QAnswer (Ruseti et al., 2015) et SPARQL-QA (Borroto et al., 2022) figurent parmi les approches KGQA les plus performantes, exploitant respectivement les lexicalisations de Wikipédia et la traduction automatique neuronale pour générer des requêtes SPARQL. Toutefois, des défis restent ouverts. En effet, si les LLMs facilitent la compréhension du langage, ils manquent de robustesse et de fiabilité. De