

Représentations conditionnelles entité-centrées pour le raisonnement multi-saut dans les systèmes de question-réponse multi-document

Romain Bourgeois^{1,2} Adrian Chifu¹ Sébastien Fournier¹

(1) Aix-Marseille Université, CNRS, LIS, Marseille, France

(2) Nextra, Marseille, France

romain.bourgeois@lis-lab.fr | romain.bourgeois@nextra-partners.com

adrian.chifu@lis-lab.fr, sebastien.fournier@lis-lab.fr

RÉSUMÉ

Les systèmes de question-réponse multi-document (MD-QA) nécessitent un raisonnement multi-saut fondé sur des informations éparses à travers plusieurs documents. Pour structurer cette information, de nombreuses approches s'appuient sur des graphes de connaissances où les passages textuels sont représentés comme des nœuds reliés par des relations lexicales, sémantiques ou symboliques. Dans ce contexte, ce papier propose EntEmbed, un encodeur conçu pour représenter un passage de manière conditionnelle à une entité spécifique qu'il contient. Cette représentation entité-centrée vise à capter les dimensions sémantiques associées à l'entité, tout en maintenant une contextualisation fine du passage. L'objectif est d'explorer comment ces représentations peuvent être construites et de les utiliser pour améliorer le raisonnement multi-saut dans les systèmes MD-QA.

ABSTRACT

Entity-centric conditional representations for multi-hop reasoning in multi-document question answering systems

Multi-document question-answering (MD-QA) systems require multi-hop reasoning based on sparse information across multiple documents. To structure this information, many approaches rely on knowledge graphs where textual passages are represented as nodes connected by lexical, semantic, or symbolic relations. In this context, this paper proposes EntEmbed, an encoder designed to represent a passage conditionally on a specific entity it contains. This entity-centered representation aims to capture the semantic dimensions associated with the entity, while maintaining fine-grained contextualization of the passage. The objective is to explore how these representations can be constructed and to use them to improve multi-hop reasoning in MD-QA systems.

MOTS-CLÉS : Systèmes de question-réponse multi-document, Recherche d'informations, Génération augmentée de récupération, Raisonnement multi-saut, Représentation sémantique, Désambiguïsation d'entités, Systèmes hybrides texte-graph.

KEYWORDS: Multi-document question answering, Information retrieval, Retrieval-augmented generation, Multi-hop reasoning, Semantic representation, Entity disambiguation, Hybrid text-graph system.
