الجزائرية الديمقراطية الشعبية الجمهورية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 – بسيدي بلعباس
Ecole Supérieure en Informatique
Mai 1945- Sidi Bel Abbes 08-

THESIS

To obtain the diploma of **Engineer**

Field: **Computer Science**

Specialty: **Computer Systems Engineering (ISI)**

Theme

Developing a Retrieval-Augmented Generation (RAG) Chatbot Enhanced by Knowledge Graphs

Presented by:

- Ms BAIAI Abir
- Ms BENNACER Nor Farah

Submission Date: **02/07/2024**

In front of the jury composed of:

- Ms. TAOUILI Amina	President
- Ms. BOUSMAHA Rabab	Supervisor
- Ms. DIF Nassima	Co-Supervisor
- Ms. Neggaz Imene	Examiner
- Mr. AIT-HAMLAT Adel	Invited

Academic Year: 2023 / 2024

Abstract

In todays rapidly evolving digital landscape, businesses are increasingly integrating large language models (LLMs) with chatbots to enhance customer interactions. While LLMs excel in understanding human-like language, they often struggle with accessing and providing real-time information accurately.

To overcome these challenges, our project focuses on developing a cutting-edge chatbot utilizing Retrieval-Augmented Generation (RAG) and knowledge graphs. By leveraging RAG to dynamically retrieve information and employing knowledge graphs as structured data repositories.

Our approach incorporates sophisticated techniques such as Semantic Chunking for efficient information retrieval, Query Reformulation for accurate capture of user intent, Complex Query Handling for generating optimal relevant responses, and Semantic Routing for precise answers.

Evaluation results demonstrate significant improvements in response quality and user satisfaction, positioning our RAG-based chatbot as an intelligent and user-friendly virtual assistant. This report provides detailed insights into the design, implementation, and evaluation of our chatbot, showcasing its potential across various domains.

Keywords: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Chatbots, Knowledge Graphs, Text Embedding, Semantic Chunking ,Query Handling, Semantic Router.

Résumé

Dans un monde numérique en constante évolution, les entreprises intègrent de plus en plus des modèles de langage avancés (LLMs) avec des chatbots pour améliorer les interactions avec les clients. Bien que les LLMs excellent dans la compréhension du langage humain, ils rencontrent souvent des difficultés pour accéder et fournir des informations en temps réel de manière précise.

Pour relever ces défis, notre projet se concentre sur le développement d'un chatbot innovant utilisant la Génération Augmentée par Récupération (RAG) et les graphes de connaissances. En exploitant la RAG pour récupérer dynamiquement des informations et en utilisant les graphes de connaissances comme des bases de données structurées.

Notre approche intègre des techniques avancées telles que le Fragmentation Sémantique pour une récupération efficace d'informations, la Reformulation de Requête pour une capture précise de l'intention de l'utilisateur, la Gestion de Requêtes Complexes pour générer des réponses pertinentes et optimales, et le Routage Sémantique pour des réponses précises.

Les résultats de l'évaluation démontrent des améliorations significatives de la qualité des réponses et de la satisfaction des utilisateurs, positionnant notre chatbot basé sur la RAG comme un assistant virtuel intelligent et convivial. Ce rapport offre des insights détaillés sur la conception, la mise en uvre et l'évaluation de notre chatbot, mettant en lumière son potentiel dans divers domaines.

Mots-clés : modèles de langage avancés, Génération Augmentée par Récupération, chatbot, les graphes de connaissances, Fragmentation Sémantique, Reformulation de Requête, Requêtes Complexes, Routage Sémantique.

ملخص

في ساحة الرقمية المتطورة بسرعة في الوقت الحالي، تقوم الشركات بدمج النماذج اللغوية الكبيرة (LLMs) مع الروبوتات الدردشة بشكل متزايد لتعزيز التفاعل مع العملاء. بالرغم من تفوق النماذج اللغوية الكبيرة في فهم اللغة البشرية، إلا أنها غالبًا ما تواجه صعوبات في الوصول إلى المعلومات وتقديمها بدقة في الوقت الحقيقي.

للتغلب على هذه التحديات، يركز مشروعنا على تطوير روبوت دردشة متطور باستخدام التوليف المعزز بالاسترجاع (RAG) والرسوم البيانية للمعرفة. من خلال استغلال RAG لاسترجاع المعلومات ديناميكيًا واستخدام الرسوم البيانية للمعرفة كمستودعات بيانات منظمة.

تضمن نهجنا تقنيات متطورة مثل الفصل الدلالي لاسترجاع المعلومات بكفاءة، وإعادة صياغة الاستعلام لالتقاط نية المستخدم بدقة، ومعالجة الاستعلامات المعقدة لتوليد ردود مناسبة ومثالية، والتوجيه الدلالي للإجابات الدقيقة.

تظهر نتائج التقييم تحسنًا كبيرًا في جودة الردود ورضا المستخدمين، مما يجعل روبوتنا القائم على RAG مساعدًا افتراضيًا ذكيًا وسهل الاستخدام. يقدم هذا التقرير رؤى مفصلة حول التصميم والتنفيذ والتقييم لروبوتنا، مما يبرز إمكانياته في مجالات مختلفة.

كلمات مفتاحية: النماذج اللغوية الكبيرة, التوليف المعزز بالاسترجاع, روبوتات الدردشة, الرسوم البيانية للمعرفة, الفصل الدلالي, .إعادة صياغة الاستعلام, معالجة الاستعلامات المعقدة, التوجيه الدلالي للإجابات