الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbes

Thesis

To obtain the diploma of **Master Degree**

Field of Study: **Computer Science**

Specialization: **Information Systems and Web (SIW)**

Theme

Knowledge Graph-Driven Question Answering for Intelligent Medical Diagnosis Systems : state of art

---

Presented by
Nassane Marwa

Defended on: **25 September 2025**
In front of the jury composed of

Dr.MAHAMMED Nadir	President of the Jury
Pr.MALKI Mimoun	Thesis Supervisor
Dr.MALKI Abdelhamid	Co-Supervisor
Dr.BELMEKKI Ghizelane Amira	Examiner
Dr.KECHAR Mohamed	Representative of the incubator,Examiner

Academic Year: 2024/2025

Abstract

The growing complexity and heterogeneity of medical data, combined with advances in Artificial Intelligence (AI), have driven the development of intelligent systems capable of providing personalized diagnoses and treatments.

Knowledge Graphs play a key role in structuring, organizing, and semantically representing medical knowledge by modeling relationships among symptoms, diseases, treatments, and patient data. This structured representation supports efficient reasoning, knowledge discovery, and integration of heterogeneous data sources., these graphs are often enriched with embedding techniques that capture semantic similarities and latent relationships in vector space.

Building on these strengths, KG-based Question Answering (QA) systems have emerged as powerful tools that allow users to interact with medical knowledge through natural language queries. These systems provide accurate, context-aware answers, improving the accessibility, usability, and interpretability of complex medical information for both healthcare professionals and patients.

In this research, we examine the current state of the art in the development of medical Question Answering (QA) systems that leverage Knowledge Graphs (KGs) in combination with various AI techniques, including Machine Learning (ML), Deep Learning (DL), embedding models, and Large Language Models (LLMs). We also explore hybrid approaches that integrate these techniques with KGs to enhance reasoning capabilities and knowledge representation.

Keywords:knowledge graph,Artificial Intelligence ,Machine learning,deep learning,question/answering system,Large Language Models ,Embedding Models,

Résumé

La complexité croissante et l'hétérogénéité des données médicales, combinées aux avancées de l'Intelligence Artificielle (IA), ont conduit au développement de systèmes intelligents capables de fournir des diagnostics et des traitements personnalisés.

Les graphes de connaissances jouent un rôle essentiel dans la structuration, l'organisation et la représentation sémantique du savoir médical, en modélisant les relations entre symptômes, maladies, traitements et données des patients. Cette représentation structurée favorise un raisonnement efficace, la découverte de nouvelles connaissances et l'intégration de sources de données hétérogènes. De plus, ces graphes sont souvent enrichis par des techniques d'embedding qui capturent les similarités sémantiques et les relations latentes dans un espace vectoriel.

En s'appuyant sur ces atouts, les systèmes de questions-réponses (QA) basés sur les graphes de connaissances (KG) ont émergé comme des outils puissants permettant aux utilisateurs d'interagir avec le savoir médical à travers des requêtes en langage naturel. Ces systèmes fournissent des réponses précises et contextualisées, améliorant ainsi l'accessibilité, l'utilisabilité et l'interprétabilité d'informations médicales complexes, tant pour les professionnels de santé que pour les patients.

Dans cette recherche, nous examinons l'état de l'art actuel dans le développement de systèmes de questions-réponses médicaux exploitant les graphes de connaissances (KG) en combinaison avec diverses techniques d'IA, notamment l'apprentissage automatique (ML), l'apprentissage profond (DL), les modèles d'embedding et les grands modèles de langage (LLM). Nous explorons également des approches hybrides qui intègrent ces techniques aux KG afin de renforcer les capacités de raisonnement et la représentation des connaissances.

Mots-clés : graphe de connaissances, intelligence artificielle, apprentissage automatique, apprentissage profond, système de questions-réponses, grands modèles de langage, modèles d'embedding

الملخص

إن التعقيد المتزايد وعدم التجانس في البيانات الطبية، إلى جانب التقدم في مجال الذكاء الاصطناعي (IA)، قد أدى إلى تطوير أنظمة ذكية قادرة على تقديم تشخيصات وعلاجات مخصصة.

تلعب الرسوم البيانية المعرفية دورًا رئيسيًا في هيكلة وتنظيم وتمثيل المعرفة الطبية دلاليًا، من خلال نمذجة العلاقات بين الأعراض والأمراض والعلاجات وبيانات المرضى. هذا التمثيل المنظم يدعم الاستدلال الفعّال، واكتشاف المعرفة، ودمج مصادر البيانات غير المتجانسة. وغالبًا ما تُثرَى هذه الرسوم بتقنيات الـ gniddebme التي تلتقط أوجه التشابه الدلالي والعلاقات الكامنة في الفضاء المتجهي.

استنادًا إلى هذه المزايا، برزت أنظمة الأسئلة والأجوبة (AQ) المعتمدة على الرسوم البيانية المعرفية كأدوات قوية تتيح للمستخدمين التفاعل مع المعرفة الطبية من خلال استعلامات بلغة طبيعية. وتوفر هذه الأنظمة إجابات دقيقة وواعية بالسياق، مما يحسن من سهولة الوصول إلى المعلومات الطبية المعقدة وقابليتها للاستخدام وتفسيرها، سواء من قبل المتخصصين في الرعاية الصحية أو المرضى.

في هذا البحث، نستعرض أحدث ما توصلت إليه الأبحاث في تطوير أنظمة الأسئلة والأجوبة الطبية التي تستفيد من الرسوم البيانية المعرفية بالتكامل مع تقنيات متعددة للذكاء الاصطناعي، بما في ذلك التعلم الآلي (LM)، التعلم العميق (LD)، نماذج الـ gniddebme، والنماذج اللغوية الكبيرة (MLL). كما نستكشف مناهج هجينة تدمج هذه التقنيات مع الرسوم البيانية المعرفية لتعزيز قدرات الاستدلال وتمثيل المعرفة.

الكلمات المفتاحية: الرسوم البيانية المعرفية، الذكاء الاصطناعي، التعلم الآلي، التعلم العميق، نظام الأسئلة والأجوبة، النماذج اللغوية الكبيرة، نماذج gniddebme