

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
المدرسة العليا للإعلام الآلي 80 ماي 5491 - سيدي بلعباس

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Higher School of Computer Science – May 8, 1945 – Sidi Bel Abbès



Master's Thesis

In partial fulfillment of the requirements for the Master degree in Computer Science

Specialization: Information Systems and Web

Topic

AI-Assisted Semantic Data Integration

A Study of Artificial Intelligence Approaches for Enhancing Semantic Web Data Interoperability

Presented by

Hakim ALLEM

Defended in: 10 March 2025

In front of the jury composed of:

Mr. Mohamed Yacine Kazi Tani

President of the Jury

Mr. Mohamed Bedjaoui (ESI-SBA)

Academic Supervisor

Mrs. Christelle Pierkot (Data Terra)

Professional Supervisor

Mrs. Khayra Benchrif

Examiner

Abstract

The rapid growth of heterogeneous and distributed data has created a critical need for advanced integration approaches that go beyond traditional syntactic and structural matching. **AI-assisted Semantic Data Integration** (AI-SDI) combines Semantic Web standards with artificial intelligence techniques to automate and enhance the alignment, merging, and enrichment of data from diverse sources.

Problematic: Despite significant progress in Semantic Web technologies, existing integration processes remain largely manual and time-consuming. The challenge lies in automating schema and ontology matching, entity resolution, and semantic enrichment while maintaining high accuracy, interoperability, and scalability.

Methodology: To address this, we conducted a comprehensive study combining theoretical research, technology review, and experimental analysis. We examined Semantic Web standards (RDF, OWL, SKOS, SPARQL) and explored AI methods such as machine learning, natural language processing, and knowledge graph embeddings. The methodology involved identifying integration challenges, reviewing state-of-the-art approaches, and analyzing the role of human-in-the-loop strategies for improving precision and explainability.

Results: The study highlights how AI can improve ontology alignment, schema matching, and knowledge enrichment, while also identifying key limitations related to scalability and ethical considerations. By synthesizing the literature and outlining promising research directions, this work provides a reference framework for future developments in AI-SDI.

Keywords— Semantic Data Integration, Artificial Intelligence, Ontology Alignment, Knowledge Graphs, Semantic Web, Machine Learning

Résumé

La croissance rapide des données hétérogènes et distribuées crée un besoin crucial de méthodes d'intégration avancées dépassant les approches purement syntaxiques ou structurelles. **L'intégration sémantique des données assistée par l'IA** (AI-SDI) combine les standards du Web sémantique avec des techniques d'intelligence artificielle pour automatiser et améliorer l'alignement, la fusion et l'enrichissement de données issues de sources diverses.

Problématique : Malgré les avancées des technologies du Web sémantique, les processus d'intégration restent majoritairement manuels et chronophages. Le défi consiste à automatiser la correspondance de schémas et d'ontologies, la résolution d'entités et l'enrichissement sémantique tout en garantissant précision, interopérabilité et montée en charge.

Méthodologie : Pour répondre à ce défi, nous avons mené une étude combinant recherche théorique, analyse technologique et exploration expérimentale. L'étude s'appuie sur les standards du Web sémantique (RDF, OWL, SKOS, SPARQL) et explore l'apport de l'IA à travers l'apprentissage automatique, le traitement du langage naturel et les plongements de graphes de connaissances. La méthodologie inclut l'identification des défis, la revue des approches de pointe et l'analyse du rôle des stratégies *human-in-the-loop* pour améliorer la précision et l'explicabilité.

Résultats : Les résultats montrent comment l'IA peut améliorer l'alignement d'ontologies, la correspondance de schémas et l'enrichissement des connaissances, tout en identifiant des limites liées à la montée en charge et aux considérations éthiques. Ce travail fournit ainsi un cadre de référence pour les développements futurs en AI-SDI.

Mots-clés— Intégration sémantique des données, Intelligence artificielle, Alignement d'ontologies, Graphes de connaissances, Web sémantique, Apprentissage automatique

Arabic abstract

إن النمو السريع للبيانات الموزعة والمتنوعة المصدر يفرض حاجة ملحة إلى أساليب متقدمة لدمج البيانات تتجاوز المطابقة التركيبية أو النحوية التقليدية. تكامل البيانات الدلالية بمساعدة الذكاء الاصطناعي يجمع بين معايير الويب الدلالي وتقنيات الذكاء الاصطناعي لأتمتة وتحسين عمليات مواءمة البيانات ودمجها وإثرائها من مصادر مختلفة.

الإشكالية: على الرغم من التقدم الكبير في تقنيات الويب الدلالي، إلا أن عمليات الدمج الحالية لا تزال في الغالب يدوية وتستغرق وقتاً طويلاً. التحدي هو أتمتة مطابقة المخططات والأنطولوجيات، وحل الكيانات، والإثراء الدلالي مع الحفاظ على الدقة، وقابلية التشغيل البيئي، وقابلية التوسع.

المنهجية: لمواجهة هذا التحدي، أجرينا دراسة شاملة تجمع بين البحث النظري، وتحليل التقنيات، والتجارب العملية. اعتمدت الدراسة على معايير الويب الدلالي (SPARQL، SKOS، OWL، RDF) واستكشفت دور الذكاء الاصطناعي من خلال التعلم الآلي، ومعالجة اللغة الطبيعية، وتضمين الرسوم البيانية المعرفية. شملت المنهجية تحديد التحديات، ومراجعة أحدث الأساليب، وتحليل دور استراتيجيات human-in-the-loop لتحسين الدقة وقابلية التفسير.

النتائج: أظهرت النتائج كيف يمكن للذكاء الاصطناعي تحسين مواءمة الأنطولوجيات، ومطابقة المخططات، وإثراء المعرفة، مع تحديد القيود المرتبطة بقابلية التوسع والاعتبارات الأخلاقية. يوفر هذا العمل إطاراً مرجعياً لتطويرات مستقبلية في مجال تكامل البيانات الدلالية بمساعدة الذكاء الاصطناعي.

الكلمات المفتاحية— تكامل البيانات الدلالية، الذكاء الاصطناعي، مواءمة الأنطولوجيات، الرسوم البيانية المعرفية، الويب الدلالي، التعلم الآلي