الجمهورية الشعبية الديمقراطية الجزائرية
وزارة التعليم العالي و البحث العلمي
المدرسة العليا للإعلام الآلي 08 ماي 1945 بسيدي بلعباس

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Speciality: **Artificial Intelligence and Data Science (IASD)**

Theme

Incorporation of ontologies into deep learning for classification based on texts and structured data.

Presented by:
HACHOUD Mohammed

Submission Date: **September, 2025**
In front of the jury composed of:

Mr. BENDAOUD Fayssal	President
Mr. KHALDI Belkacem	Supervisor
Ms. HARZALLAH Mounira	Supervisor
Ms. NAOUM Hanae	Examiner

Academic Year : 2024/2025

Abstract

This Master's thesis presents a novel approach to archaeological ceramic classification by integrating domain-specific ontologies with Graph Neural Networks (GNNs) for enhanced analysis of structured archaeological data. Working with the Spiridon database containing over 11,000 ceramic objects from the Laboratoire d'Archéologie Médiévale et Moderne en Méditerranée (LA3M), this research emerged from the observation of significant missing data problems and the hierarchical nature of archaeological ceramic classification during internship work with this database. These challenges in ceramology led to conducting research on how Graph Neural Networks combined with ontological knowledge can address the limitations of traditional tabular data approaches.

The proposed methodology transforms the structured archaeological database into a comprehensive Knowledge Graph, where ceramic objects are represented as nodes within networks of meaningful archaeological relationships. By incorporating formal ontologies encoding centuries of expertise, the system guides Graph Neural Networks to respect expert-defined relationships and cultural contexts that traditional machine learning approaches often overlook.

This research develops an ontology-enhanced GNN architecture for archaeological ceramic classification, using semi-supervised strategies to maximize limited labeled data while leveraging large unlabeled collections. The approach integrates structured archaeological information—including descriptive metadata, typological classifications, and hierarchical taxonomies—into a unified analytical framework that can handle missing data and preserve archaeological knowledge structures.

This Master's thesis provides a comprehensive theoretical foundation through background chapters covering ceramology, knowledge graphs, and graph neural networks, followed by a detailed state-of-the-art analysis of relational graph neural networks and ontology integration approaches. The research establishes the theoretical groundwork for understanding how archaeological domain knowledge can be effectively represented and integrated with modern graph-based learning architectures.

This research contributes to the broader field of neuro-symbolic AI by demonstrating how formal domain knowledge can be effectively integrated with modern deep learning architectures, offering a paradigm for AI systems that amplify human expertise while preserving the interpretive complexity that makes archaeological research both challenging and profoundly meaningful.

Keywords— Ontologies, Deep Learning, Graph Neural Networks, Knowledge Graphs, Ceramology, Cultural Heritage, Semi-supervised Learning, Multimodal Data

الملخص

تطرح هذه الرسالة منهجاً مبتكراً لتصنيف السيراميك الأثري عبر دمج الأنطولوجيات المتخصصة بالمجال مع شبكات الجراف العصبية ،(GNNs) بهدف تطوير تحليل أكثر عمقاً وفعالية للبيانات الأثرية المهيكلة. ينطلق البحث من قاعدة بيانات سبيريدون التي تضم أكثر من 11,000 قطعة سيراميك محفوظة في مختبر الآثار الوسطى والحديثة في البحر الأبيض المتوسط (LA3M). خلال فترة التدريب على هذه القاعدة، برزت تحديات أساسية تتعلق بوفرة البيانات المفقودة وبالطبيعة الهرمية لتصنيف القطع الأثرية، وهو ما دفع إلى استكشاف كيف يمكن لشبكات الجراف العصبية المدعمة بالمعرفة الأنطولوجية تجاوز قيود النهج التقليدي القائم على البيانات الجدولية.

تعيد المنهجية المقترحة صياغة القاعدة الأثرية المهيكلة على شكل جراف معرفة متكامل، حيث تُمثل القطع السيراميكية كعُقد مترابطة بعلاقات أثرية ذات معنى. ومن خلال إدماج أنطولوجيات رسمية تجسد خبرة تراكمت عبر قرون، يتم توجيه شبكات الجراف العصبية لاحترام العلاقات والسياقات الثقافية المحددة من قبل الخبراء، وهي أبعاد غالباً ما تغفلها طرق التعلم الآلي التقليدية.

يعتمد البحث على بناء بنية جراف عصبية معززة بالأنطولوجيا لتصنيف السيراميك الأثري، باستخدام استراتيجيات شبه مُراقبة لاستثمار الكم المحدود من البيانات المصنفة مع الاستفادة من الكميات الكبيرة غير المصنفة. كما يوحد هذا النهج البيانات الأثرية المهيكلة—من بيانات وصفية وتصنيفات نمطية وهرمية—ضمن إطار تحليلي قادر على معالجة النقص في البيانات والحفاظ على البنية المعرفية للمجال.

وتتضمن الرسالة أساساً نظرياً شاملاً من خلال فصول تمهيدية تغطي علم السيراميك، وجرافات المعرفة، وشبكات الجراف العصبية، يليها تحليل معمق لأحدث التطورات في الشبكات العلائقية ونهج دمج الأنطولوجيا. وبذلك يضع البحث أسساً لفهم كيفية تمثيل ودمج المعرفة الأثرية بشكل فعّال داخل بنى التعلم الجرافية الحديثة.

يساهم هذا العمل في الحقل الأوسع للذكاء الاصطناعي العصبي-الرمزي من خلال توضيح كيفية دمج المعرفة الرسمية للمجال مع تقنيات التعلم العميق، مما يفتح الطريق أمام أنظمة ذكاء اصطناعي قادرة على تعزيز الخبرة البشرية مع الحفاظ على البعد التفسيري الذي يمنح البحث الأثري قيمته العلمية والثقافية.

الكلمات المفتاحية: الأنطولوجيات، التعلم العميق، الشبكات العصبية للرسوم البيانية، التصنيف الأثري، رسوم المعرفة، علم الخزف، التراث الثقافي، التعلم شبه الخاضع للإشراف، البيانات متعددة الوسائط

Résumé

Ce mémoire de Master présente une approche novatrice pour la classification archéologique des céramiques en intégrant des ontologies spécifiques au domaine avec les réseaux de neurones graphiques (GNNs) pour une analyse améliorée des données archéologiques structurées. Travaillant avec la base de données Spiridon contenant plus de 11 000 objets céramiques du Laboratoire d'Archéologie Médiévale et Moderne en Méditerranée (LA3M), cette recherche a émergé de l'observation de problèmes significatifs de données manquantes et de la nature hiérarchique de la classification archéologique des céramiques lors du stage avec cette base de données. Ces défis en céramologie ont mené à conduire une recherche sur comment les réseaux de neurones graphiques combinés avec la connaissance ontologique peuvent adresser les limitations des approches traditionnelles de données tabulaires.

La méthodologie proposée transforme la base de données archéologique structurée en un graphe de connaissances compréhensif, où les objets céramiques sont représentés comme nœuds au sein de réseaux de relations archéologiques significatives. En incorporant des ontologies formelles encodant des siècles d'expertise, le système guide les réseaux de neurones graphiques à respecter les relations définies par les experts et les contextes culturels que les approches traditionnelles d'apprentissage automatique négligent souvent.

Cette recherche développe une architecture GNN améliorée par ontologie pour la classification archéologique des céramiques, utilisant des stratégies semi-supervisées pour maximiser les données étiquetées limitées tout en exploitant de grandes collections non étiquetées. L'approche intègre des informations archéologiques structurées—incluant les métadonnées descriptives, les classifications typologiques, et les taxonomies hiérarchiques—dans un cadre analytique unifié capable de gérer les données manquantes et de préserver les structures de connaissance archéologique.

Ce mémoire de Master fournit une base théorique complète à travers des chapitres de contexte couvrant la céramologie, les graphes de connaissances, et les réseaux de neurones graphiques, suivi d'une analyse détaillée de l'état de l'art des réseaux de neurones graphiques relationnels et des approches d'intégration d'ontologies. La recherche établit les fondements théoriques pour comprendre comment la connaissance du domaine archéologique peut être efficacement représentée et intégrée avec les architectures d'apprentissage graphique modernes.

Cette recherche contribue au domaine plus large de l'IA neuro-symbolique en démontrant comment la connaissance formelle du domaine peut être efficacement intégrée avec les architectures modernes d'apprentissage profond, offrant un paradigme pour les systèmes d'IA qui amplifient l'expertise humaine tout en préservant la complexité interprétative qui rend la recherche archéologique à la fois challengeante et profondément significative.

Mots clés: Ontologies, Apprentissage Profond, Réseaux de Neurones sur Graphes, Classification Archéologique, Graphes de Connaissances, Céramologie, Patrimoine Culturel, Apprentissage Semi-supervisé, Données Multimodales