

Université de Jendouba
Institut Supérieur de Gestion

Rapport de Recherche

**Intégration des données hétérogènes sur le Web sémantique
à l'aide des Knowledge Graphs et des modèles de langage**

Présenté par

Feriel Benhenda

Octobre 2025

Plan du Rapport

1. Hétérogénéité des données et types rencontrés
2. Importance de la gestion de l'hétérogénéité dans le Web sémantique
3. Technologies du Web sémantique et harmonisation des données
4. Rôle des Knowledge Graphs dans la représentation intégrée
5. Avantages du format graphe face aux bases relationnelles
6. Apport des Knowledge Graphs à l'apprentissage automatique
7. Interaction entre LLM et Knowledge Graphs
8. Défis liés à l'utilisation conjointe des KG et LLM
9. Contribution des LLM à la correction et au complétement des données
10. Collaboration entre modèles sémantiques et modèles statistiques
11. Scénarios métiers prometteurs

1. Qu'est-ce que l'hétérogénéité des données dans le contexte de l'intégration ?

L'hétérogénéité des données désigne la diversité des formats, structures et significations des informations issues de différentes sources. Elle se manifeste à trois niveaux :

- **Hétérogénéité structurelle** : différence dans l'organisation des données (ex. bases relationnelles, XML, JSON).
- **Hétérogénéité syntaxique** : variations dans les formats ou conventions.
- **Hétérogénéité sémantique** : divergence dans le sens attribué aux données.

2. Pourquoi est-il important de prendre en compte l'hétérogénéité pour réussir une intégration de données sur le Web sémantique ?

Ignorer l'hétérogénéité conduit à des incohérences et des doublons. La prise en compte des différences garantit l'interopérabilité et la compréhension uniforme entre systèmes, assurant une intégration réussie sur le Web sémantique.

3. Comment les technologies du Web sémantique (ex : RDF, ontologies) aident-elles à réduire les incohérences et à harmoniser des données hétérogènes ?

Les technologies du Web sémantique comme RDF et les ontologies structurent les données selon des modèles communs. RDF utilise des triplets (sujet, prédicat, objet) et les ontologies définissent les concepts et relations, assurant cohérence, harmonisation et réutilisation des informations.

4. Quel rôle joue un Knowledge Graph dans la gestion et la représentation des données intégrées provenant de sources variées ?

Un Knowledge Graph (KG) relie des entités issues de différentes sources. Il organise les connaissances sous forme de nœuds et relations, offrant une vue unifiée, interconnectée et contextuelle des données intégrées.

5. En quoi le format graphe facilite-t-il l'exploration et le raisonnement par rapport à des bases relationnelles classiques dans un contexte hétérogène ?

Le format graphe modélise des relations complexes sans schéma rigide. Il permet des inférences automatiques, des parcours dynamiques et des analyses sémantiques plus flexibles que les bases relationnelles classiques, surtout dans des contextes hétérogènes.

6. Comment les Knowledge Graphs peuvent-ils contribuer à améliorer les résultats et la contextualisation dans des applications d'apprentissage automatique ?

Les KG fournissent un contexte sémantique riche qui améliore la pertinence et la précision des modèles d'apprentissage automatique. Ils permettent d'expliquer les résultats, contextualiser les prédictions et enrichir les données d'entraînement.

7. De quelle manière les grands modèles de langage (LLM) peuvent-ils être utilisés pour enrichir ou interagir avec les connaissances structurées dans un Knowledge Graph ?

Les LLM peuvent générer, reformuler ou compléter des connaissances manquantes dans un KG. Ils facilitent la création de liens sémantiques, la génération de descriptions et l'interrogation en langage naturel, enrichissant ainsi les graphes de connaissances.

8. Quels sont les défis spécifiques liés à l'utilisation conjointe des Knowledge Graphs et des LLM pour la compréhension et le traitement des données hétérogènes ?

Les principaux défis sont :

- Assurer la cohérence entre connaissances formelles (KG) et probabilistes (LLM).
- Gérer la qualité et la complétude des données.
- Maintenir la scalabilité et la mise à jour dynamique des informations.
- Harmoniser le raisonnement symbolique et statistique.

9. Comment les LLM peuvent-ils aider à identifier, corriger ou compléter les données manquantes ou ambiguës dans un Knowledge Graph ?

Les LLM détectent les incohérences, proposent des liens manquants ou des définitions alternatives, et complètent les KG avec des informations plausibles, pertinentes et contextuelles, améliorant ainsi leur complétude et fiabilité.

10. En quoi une collaboration entre modèles sémantiques (comme les KG) et modèles statistiques de langage (LLM) peut ouvrir de nouvelles perspectives pour l'intelligence artificielle ?

La combinaison KG + LLM crée une intelligence hybride : rigueur du raisonnement symbolique et flexibilité linguistique. Elle permet une IA explicable, adaptable et capable d'une compréhension contextuelle avancée.

11. Quels sont les scénarios métiers où cette intégration entre Web sémantique, Knowledge Graphs et LLM est particulièrement prometteuse ?

Cette intégration est prometteuse dans :

- **Santé** : intégration de dossiers médicaux et aide au diagnostic.
- **Finance** : analyse de risques et détection de fraudes.
- **E-commerce** : recommandations personnalisées.
- **Recherche scientifique** : fusion et exploration de bases de connaissances hétérogènes.

Elle permet une exploitation plus efficace des données et une prise de décision plus éclairée.

Conclusion

L'hétérogénéité des données constitue un défi majeur pour l'intégration et l'exploitation intelligente des informations. Grâce au Web sémantique, aux ontologies, aux Knowledge Graphs et aux modèles de langage, il devient possible de construire des systèmes capables de raisonner, contextualiser et apprendre. La collaboration entre modèles symboliques et statistiques ouvre de nouvelles perspectives pour une intelligence artificielle explicable, flexible et performante.