

Proposition d'encadrement de projet

A préciser: *Projet de Découverte de la Recherche*

Laboratoire : CRAN

Équipe : Département ISET

Encadrant : Mario Lezoche (MdC HDR) et Mariano Ferrarione (PhD)

Bureau : 2.20 Téléphone : 06 24 58 97 37 E-mail : mario.lezoche@univ-lorraine.fr

Présentation du sujet (1 page max)

Titre : Création d'un logiciel pour transformer un Graphe de connaissance en une Ontologie et vice versa.

Description :

Un graphe est un ensemble d'entités reliées entre elles, composé de nœuds qui représentent les entités et d'arcs ou arêtes qui représentent les relations. Les relations ou arcs peuvent être enrichis par des attributs ou encore une valeur quantitative représentant le poids de la relation.

Le Knowledge Graph, terme introduit par Google, est une représentation de la connaissance relative à un domaine ou une entreprise sous une forme qui est facilement exploitable par la machine. Il est constitué d'entités et de relations, les entités étant des noms de personnes, des concepts, des objets. Bien que cette représentation graphique de la connaissance ne soit pas récente, elle a gagné en popularité et est un élément clé pour **des applications d'intelligence artificielle liés à la recherche rapide et contextuelle d'information ainsi qu'à la prise de décision**. Les données sont stockées dans des bases de données de type graphe et peuvent être structurées ou non-structurées.

Le graphe de connaissance est une base de connaissance sémantique qui permet de décrire la sémantique des sources d'information et rendre ainsi le contenu explicite. En effet, une entité ou un mot seul ne porte pas beaucoup de sens en soi, celui-ci se révèle quand l'entité est prise dans son contexte qui est défini par les propriétés de l'entité et les relations avec des entités auxiliaires. Chaque entité participe donc à la compréhension des entités auxquelles elle est liée. On dit alors que la base de connaissance est sémantique car elle encode le sens des données afin de mieux permettre aux machines de comprendre une requête introduite par l'humain. On peut donc utiliser le langage naturel pour interroger une base de connaissance et générer une réponse en langage naturel car les mots sont reliés aux concepts. C'est ce qui permet aux assistants vocaux tels que Google Assistant, Siri ou Alexa de comprendre la requête de l'utilisateur et de répondre avec précision.

Un graphe de connaissance ou knowledge graph se définit par trois éléments ("Towards a definition of knowledge graph", Lisa Ehrlinger and Wolfram Wöß) :

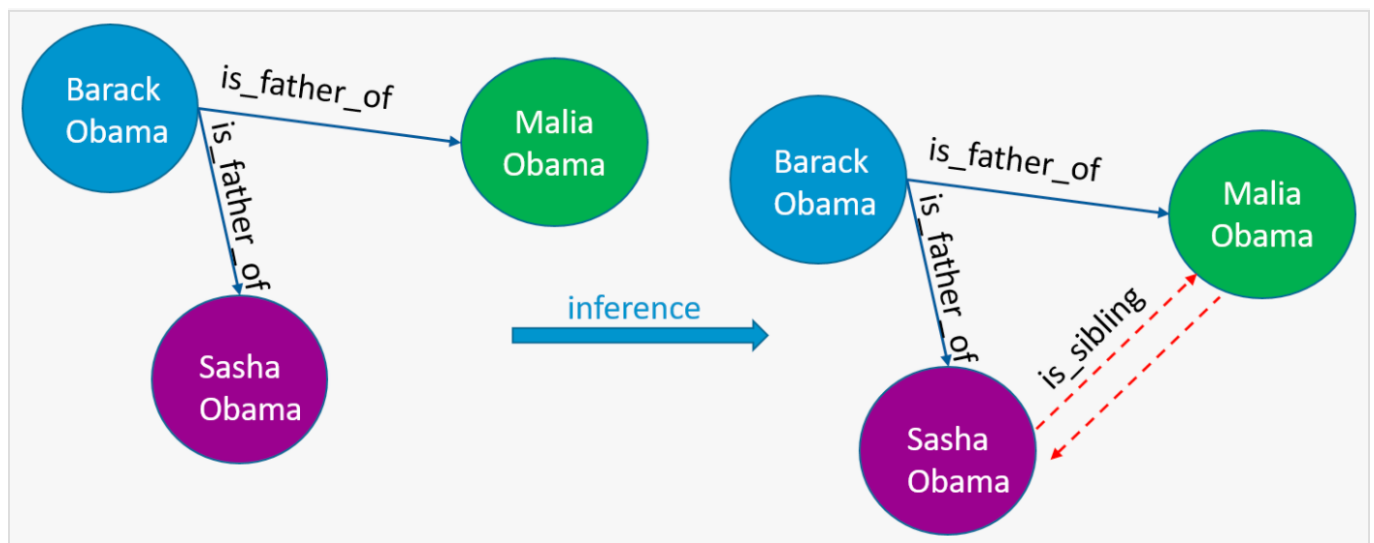
1. L'ingestion d'information provenant de multiples sources contenant des données structurées et non structurées
Dans une organisation où la connaissance (et les données y afférentes) est organisée en silos, les graphes de connaissances permettent d'avoir une vue globale de l'information. En particulier, ils

peuvent mieux que les bases de données relationnelles, capturer des données issues de multiples sources, hétérogènes et même incomplètes. Le principal avantage des bases de données graphes par rapport aux bases de données relationnelles est qu'elles permettent de représenter l'information de façon sémantique et permettent d'absorber une grande quantité de données

2. L'intégration de cette information à une ontologie.

L'ontologie permet, dans un domaine donné, de structurer les données extraites dans un graphe en décrivant les concepts (personne, lieu, ...), les types de relations entre ces concepts (enfant de), les propriétés, les règles business et les contraintes.

3. L'application d'un système d'inférence pour dériver de nouvelles connaissances à partir du graphe. Il est possible avec les graphes d'inférer de nouvelles relations entre entités, générant ainsi une nouvelle information.



Exemples de graphes de connaissance publiques :

- DBpedia
- Google Knowledge Graph
- Microsoft Concept Graph
- GeoNames
- Wikidata

L'objectif de ce projet est de créer un logiciel qui implémente un algorithme de transformation sans perte d'un graphe de connaissances vers une base de connaissances exprimée par une ontologie.

Environnement de travail (matériel, logiciel) :

On aura à disposition un serveur du CRAN, un bureau, des ordinateurs et l'accès à tous les documents scientifiques nécessaire pour la compréhension du problème et des outils à développer.

Le langage de programmation sera Python (environnement à choix).