

Programme du cours :

L'objectif de ce cours est d'approfondir les acquis du cours du semestre précédent et qui portait sur les différents formats de représentation des connaissances et notamment les ontologies, pour surtout étudier leurs différentes applications, en NLP notamment. On couvrira d'abord les méthodes permettant d'automatiser la création d'ontologies et de bases de connaissances avant de présenter leurs applications principales en NLP, mais pas que, avec notamment un focus sur le monde du renseignement.

Objectifs du cours :

- * comprendre les méthodes proposées pour automatiser la construction d'ontologies
- * étudier en détail diverses tâches NLP associées (Keyword Extraction, Open Information Extraction, Topic Modeling)
- * étudier les graphes (de connaissance notamment) et leurs propriétés
- * élargir ses connaissances en data science en étudiant des algorithmes connus, utiles pour les entretiens.
- * étudier le chemin de vie de la donnée, de sa collecte jusqu'à son exploitation
- * voir comment les graphes de connaissance et les ontologies sont utilisés au sein d'application.

Programme indicatif :

Date	Programme de la séance
28/01	1) Introduction : aux origines des ontologies
04/02	2) Automatiser la création d'ontologies
11/02	3) Qu'est ce qu'une relation saine ?
18/02	4) Ingestion, indexation, requête
04/03	5) Les graphes et leurs applications
11/03	6) Le clustering : algorithmes, et applications
18/03	7) ?
25/03	8) ?
02/04	9) ? Contrôle ?
10/04	10) Big Data is watching us

Encore quelques séances au contenu non déterminé. À vos votes !

Plusieurs sujets possibles parmi :

- * Les tourments de la déduplication ou comment séparer Dupont et Dupond (entity resolution)
- * Les applications des graphes de connaissance en NLP : analyse de corpus et RAG
- * Les graphes au service de la lutte contre la fraude
- * Les Agents IA et leurs liens possible avec les ontologies
- * Ontology as a code : les liens avec le paradigme orienté objet
- * Ontologies et raisonnements automatiques

Modalités d'évaluation :

=> Un contrôle de connaissance en fin de semestre (40%)

Porte sur le cours ET les supports fournis. Questions de cours, application de certains algorithmes, réflexion sur des cas pratiques.

=> Présentation en groupe de 3 d'un article scientifique autour d'un thème proche de ceux abordé. Nécessite de présenter le contexte de l'article, l'expérience menée et les résultats de façon critique.

Présentation de 15/20 mn + 10 mn de questions / discussion / reprise

Parmi les thèmes proposés (à finaliser) :

- * L'embedding d'ontologies

<https://arxiv.org/pdf/2406.10964>

- * Une nouvelle approche d'extraction de relations

<https://github.com/SapienzaNLP/relik>

<https://arxiv.org/abs/2408.00103>

- * LLM on graph

<https://arxiv.org/pdf/2312.02783> ou <https://openreview.net/pdf?id=IuXR1CCrSi>

- * Le Graph RAG par clustering

<https://github.com/microsoft/graphrag>

<https://arxiv.org/html/2404.16130v1>

- * La fin des LLM ? Large Concepts Models

<https://arxiv.org/pdf/2412.08821>

- * Entity linking as Question Answering ?

<https://arxiv.org/abs/2110.02369>

=> Travaux pratiques en python 40 %

- * TP 1 sur l'automatisation de la création d'ontologies

- * TP 2 sur les bases de données à base de graphes

- * TP 3 : selon les choix des séances 7 et 8 ou travail de groupe pour prolonger la présentation d'article