

Web sémantique

Mini-projets

Master 2 Datascale – 2023/2024

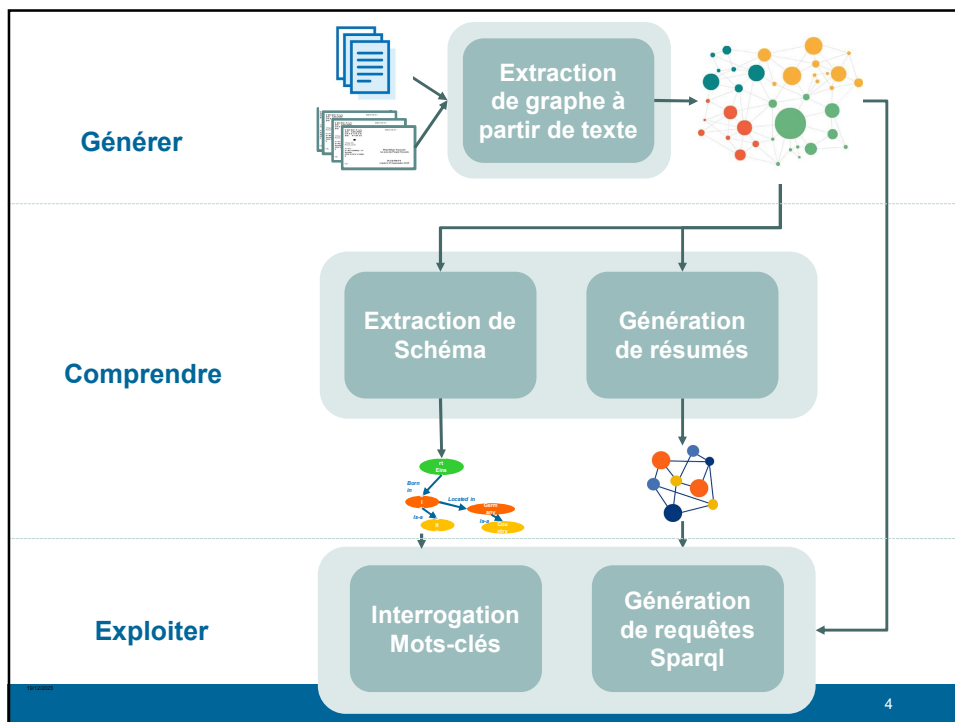


Présentation

- Objectif des mini-projets
 - Mise en pratique des concepts vus en cours
 - Réaliser une étude bibliographique sur un sujet, avantages et inconvénients des approches étudiées
 - Proposer une solution au problème posé
 - Implementation et test de la solution
- A faire par groupes (entre 2 et 4 étudiants)
- Résultats attendus
 - Rapport, présentation et démo
- Dates clés
 - Présentations avec démo et rapport : Semaine du 5/2/2024

Contexte

- Un environnement pour faciliter l'utilisation de sources RDF



Sujet 1 – Extraction de graphes de connaissances à partir de texte

- Génération automatique d'un graphe RDF à partir d'un texte en langage naturel
 - Extraction d'entités nommées (les sommets / ressources)
 - Librairie NLP SpaCy
 - Extraction de liens entre ces entités (les arcs / propriétés)
 - Ontologies, LLM

Zoubida Kedad

5

Sujet 2 – Génération du schéma d'un graphe de connaissances

- Extraction automatique des classes / types et propriétés d'un graphe RDF
 - Approches de clustering
 - Approches statistiques
- Annotation des classes / types
 - Bases de connaissances ou ontologies de références

Zoubida Kedad

6

Sujet 3 – Génération d'un résumé d'un graphe de connaissances

- Produire automatiquement une représentation concise d'un grand graphe RDF mettant en évidence ses concepts les plus « importants »
 - Comment identifier les thèmes saillants ?
 - Comment identifier les ressources importantes (e.g. centralité) et les classer ?
 - Comment construire une synthèse (sous-graphe) contenant toutes les ressources importantes ?
 - Comment classer les sous-graphes obtenus s'il y en a plusieurs (e.g. PageRank) ?

Zoubida Kedad

7

Sujet 4 - Recherche par mots clés dans un graphe RDF

- $Q = \{kw_1, kw_2, \dots, kw_n\}$
- Pour chaque mot clé, identification de fragments pertinents dans le graphe RDF
 - Définir un processus de matching permettant d'identifier les ressources / propriétés / sous-graphes
- Construction d'un résultat à partir de fragments pertinents
 - Comment "relier" les fragments pertinents : recherche de chemins, utilisation d'ontologies

Zoubida Kedad

8

Sujet 5 – Génération de requêtes Sparql

- Etant donné un besoin exprimé comme une ontologie, générer automatiquement une requête Sparql permettant de retrouver les triplets instances de cette ontologie
 - Identifier les triplets candidats
 - Générer les BGP représentatifs de ces triplets candidats
 - Générer la requête Sparql correspondante