

**Sujet conception d'un entrepôt de données de type GED,
Analyse via Text Mining.**

Soutenance du projet le 1 juin 2021

Module : Entrepôts de Données - Master2 MIAGE



Objectifs

Construire un entrepôt de données non structurées de type Gestion Electronique de Documents et y réaliser un certains nombres d'études via les principe du Text Mining.

Données sources

L'ensemble des documents de cours que vous avez cumulés tout au long de vos années de formations. Vous pouvez les ramener tous au format PDF si cela simplifie votre gestion.

Modèle de l'entrepôt

- Modèle conceptuel avec indexation des documents et métadonnées d'enrichissement.
- Modèle logique relationnel normalisé ou dénormalisé.
- Modèle physique MariaDB ou PostgreSQL (sous Docker)
- Repository documents FileSystem.

Alimentation de l'entrepôt (Ingestion)

Moyens possibles :

- Chargement dans l'entrepôt à l'aide d'un outil de type EAI (en utilisation des fichiers de métadonnées accompagnant le document)
- Avec un script qui accède aux métadonnées du document et les charge dans l'entrepôt (Ingestion Custom)
- Vous prendrez en compte la gestion de l'absence des métadonnées dans le documents et la notion de mise à jour du document dans les deux cas.

Analytics au choix

- Vous mettrez en place une solution basé sur du Text Mining qui permettra d'exploiter le contenu de votre entrepôt, conception d'un nuage de mots ou classification automatique des documents (vous présenterez une solution open source pour réaliser cette analyse)
- Ou Mise en place avec un moteur Elasticsearch ou Solr d'un moteur de recherche simple sur cet entrepôt.

GIT :

Vous mettrez à disposition et en partage avec le compte tondeur.herve@yahoo.fr un GIT soit sur GitLab ou GitHub où vous déposerez vos sources et documents de travaux préparatoires, cet espace collaboratif vous permettra de créer votre équipe projet et d'améliorer vos échanges et partages.

Soutenance

Est attendu une présentation organisationnelle de votre équipe autour du projet, technique du projet démontrant les principes techniques et technologiques mis en jeu pour réaliser ce projet
Cette présentation se réalisera selon un temps limité de 30 minutes (5 minutes de mise en place, 20 minutes de présentation du projet avec démo et support de présentation à l'appui), 5 minutes pour les questions.