

Text mining et RI

M2 BI&A
2025-2026

Préambule

► Objectif pédagogique

► Contexte

- Temporalité de l'enseignement
- Enjeu d'évaluation

► Challenge

- sur données réelles
- avec mise en situation
- dans un temps très limité
- avec une organisation en groupe de 4-5 personnes (avec répartition des personnes ayant suivi le cours d'ingénierie des données en M1 de préférence)
- en mobilisant travail sur données textuelles
- restitution orale vendredi 16 janvier après-midi : 20 minutes par groupe

Mise en situation

- ▶ Dans le contexte de la recherche scientifique
 - ▶ Internationalisation de la recherche
 - ▶ Enjeu des communautés nationales à exister
 - ▶ Communauté Extraction et Gestion des Connaissances (EGC)
 - ▶ Défi EGC 2020 : « 20 ans d'histoire pour quel avenir ? »
 - ▶ <https://www.egc.asso.fr/manifestations/defi-egc/defi-egc-2020-20-ans-dhistoire-pour-quel-avenir.html>
 - ▶ Objectif : donner des éléments à l'Association par rapport à la production scientifique en fournissant des tableaux de bord

Enjeux du travail

- ▶ Mise en œuvre de vos compétences en BI en articulant avec un travail sur le texte
- ▶ Organisation d'un travail en groupes / répartition des tâches
- ▶ Apprentissage
- ▶ Travailler sur l'adaptation à de nouvelles méthodes
- ▶ Mobilisation de ressources possibles (articles sur le défi)
- ▶ Choix d'outils libres
- ▶ Communication orale sur les résultats en fin de séance du vendredi 16 janvier !

Présentation des données

- ▶ 1 fichier CSV des éditions de la conférence : `edition_EGC.csv`
 - ▶ Year
 - ▶ City
 - ▶ Country
- ▶ 1 fichier CSV des publications : `export_articles_EGC_2004_2018.csv`
 - ▶ Series : nom revue
 - ▶ Booktitle : EGC
 - ▶ Year
 - ▶ Title
 - ▶ Abstract
 - ▶ Authors (séparé par des virgules)
 - ▶ Link pdf1page
 - ▶ Lien pdfarticle

Objectifs

- Explorez, investiguez, surprenez-moi !

Bon travail !