



Projet final

Indexation sémantique et recherche d'information

1 OBJECTIF

L'objectif de ce projet est de créer un moteur de recherche en utilisant Solr/Lucene.

2 DESCRIPTION

1. **Sélection du corpus** : Le corpus doit avoir les conditions suivantes : comporter au minimum 1.000 documents et au minimum 5 attributs par document correspondants aux prérequis de la question 5.
2. **Annotation du corpus** : Création des annotations en GATE afin de prétraiter le corpus (par exemple, en ajoutant un identifiant aux documents).
3. **Création du schéma** : Créez le schéma d'indexation de Solr.
4. **Indexation du corpus** : Création/modification des fichiers « xml » ou « csv » à partir du corpus afin d'être indexés par Solr.
5. **Création des facettes** : Proposez au moins cinq facettes (2 facettes de champ, 2 facettes d'intervalles, 1 facette pivot)
6. **Modification de Velocity** : Modifiez « Velocity » afin de visualiser le contenu des documents.
7. **Rédaction du rapport** : Ecrivez un rapport détaillant le travail effectué et en discutant chacune des étapes/décisions réalisées.

3 SOUTENANCE

8. **Fichier PowerPoint** : Le projet devra être soutenu oralement. Vous devrez créer un fichier PowerPoint ou équivalent contenant les diapositives de présentation et le sauvegarder en format PDF.
9. **Présentation** : La durée de la présentation est de 10 minutes + 5 minutes de questions & réponses. Durant la présentation, vous devrez reprendre succinctement chaque partie du rapport rendu et faire une démonstration en direct de votre moteur de recherche

4 SOUMISSION

L'ensemble des documents à fournir doit être contenu dans un même dossier portant votre nom, par exemple « Dupond ». Ce dossier principal doit comporter :

- Un sous-dossier contenant le corpus.
- Le fichier « schema.xml »
- Le fichier « solrconfig.xml »
- Les fichiers « *.vm » modifiés/créés.
- Le rapport écrit, au format PDF.
- La présentation, au format PDF.

Le dossier principal doit ensuite être compressé et envoyé via la plateforme pédagogique de notre faculté : <https://moodle-lettres.paris-sorbonne.fr/>

5 ÉVALUATION

- Groupes : Travail individuel.
- Date : **lundi 15 novembre 2021**