

Mini Projet

- Travaillez en binômes.
- Faire un rapport technique.
- Envoyez le rapport sous formats pdf et doc en utilisant la plateforme **piazza** avant le **29 Mai**.
- Il y aura une présentation des projets le **mardi 31 mai** à partir de 08h30.
- La note du projet compte 40% de la note finale.

Solr

1. Installez et configurez nutch (<http://nutch.apache.org/>) pour qu'il fasse l'indexation des universités marocaines.
2. La mise à jour de l'index doit être réalisé tous les jours à minuit.
3. Installez et configurez **solr** (<http://lucene.apache.org/solr/>)
4. Intégrez **nutch** avec **solr**.
5. Proposez une interface web personnalisé qui contient le nom et le logo de la faculté ainsi que le nom du Master M2I. Cette interface remplacera l'interface classique fournie par solr.
6. (question optionnelle) Proposez un logo pour le Master.
7. Pour les documents et la recherche en langue arabe :
 - intégrez Alkhalil (<http://oujda-nlp-team.net/?p=1299&lang=en>) pour pouvoir utiliser le lemme (partie requête et partie indexation).
 - utilisez la liste fournie des mots-vides (envoyer via piazza).
8. Décrivez les différentes composantes suivantes :
 - crawler ;
 - indexation ;
 - endroit de la base de données des indexes ;
 - méthode de classement utilisée ;
 - les outils utilisés (stemming, lemmatisation, transformation de documents, ...);