

# TP3 – La recherche

M2 informatique – Université Paris Diderot

Année 2018-2019

## **Exercice 1** *Requête*

La requête de l'utilisateur consiste en un ensemble de mots.

1. Enlever de cet ensemble les mots redondants et les mots « vides » (voir TP 1).  
(*Optionnel*) Pour chaque mot, si nécessaire en corriger l'orthographe si vous avez programmé ces fonctions au TP 1. Sinon, on considérera seulement les requêtes à base de mots corrects.
2. Donner un algorithme *efficace* qui, à partir de la relation mots-pages triée, énumère, dans l'ordre de pagerank décroissant, toutes les pages contenant *tous* les mots de la requête.
3. Programmer cet algorithme et le tester.

## **Exercice 2** *Site*

Programmer un serveur et réaliser un site web fonctionnel où l'utilisateur entre sa requête et où s'affiche la liste des résultats de la recherche.

Penser à écrire une fonction de conversion des liens Wikipédia donnés dans le fichier `frwiki.xml` en liens `html` pour qu'on puisse suivre les liens donnés par votre moteur de recherche.

## **Exercice 3** *Améliorations possibles*

Quels sont les défauts de votre moteur de recherche ? Comment les améliorer ? (Inutile de programmer ces améliorations.)

Lors de la phase d'exploration du graphe, vous avez stocké la fréquence d'apparition des mots dans les pages. Vous en êtes-vous servi ? Comment la prendre en compte ?