

UE INF2245

Algorithme du PageRank

Frédéric Raimbault

1 Implémentation de l'algorithme du pagerank

Complétez les programmes Map et Reduce fournis sur l'ENT pour implémenter l'algorithme du PageRank :

1. **ExtractLinksMR.java** est le driver du programme MR qui extrait les noms de page et les liens sortants des articles de wikipédia et formate les résultats pour servir d'entrée au programme **PageRankMR.java**.
2. **PageRankMR.java** est le driver du programme MR qui prend en entrée les données produites par **ExtractLinksMR.java** et calcule le PageRank de chaque page d'article de Wikipédia.

Testez votre programme sur des extraits puis sur une copie complète de Wikipédia disponible dans le répertoire `/data/wikipedia` du système de fichier HDFS du cluster.

Dans le rendu de votre TP, vous influerez un fichier `README.txt` contenant le nombre de liens trouvés par **ExtractLinks**, le nombre d'itérations et le PageRank moyen pour la totalité de Wikipedia.