

Initiation au NoSQL

Exercices 04 – Le graphe mis en pratique

Twitter en graphe

Description

Dans le cadre d'une étude de positionnement de la communication de Dataswati sur Tweeter, je souhaite étudier le réseau d'influence et contrôler que je suis placé sur les bons mots-clés. Pour rappel: Nous faisons de l'IA pour l'Industrie.

Votre travail consistera à analyser les mots-clés qui sont rattachés directement et indirectement à nos tweets au travers des likes et des retweets à 2 niveaux de profondeur.

Installations préalables

Synchro git

- Synchroniser votre git dans le dossier source
- Un dossier **talend_components** doit apparaître

Installation Talend

1. Depuis Talend, cliquez sur Fenêtre -> Préférences
2. Dans l'arborescence choisissez Talends > Composants
3. Dans le champ Dossier des composants utilisateurs, placez le chemin du dossier **talend_components**
4. Relancer Talend
5. La bibliothèque Twitter est disponible.

Création d'un compte Twitter

- Créer / Connectez-vous à un compte Twitter

Création d'une apps Twitter

- Ouvrir l'URL: <https://apps.twitter.com/>
- Cliquer sur le bouton Create New App
- Remplir les éléments obligatoires
- Valider

Création d'un token

- Depuis l'interface cliquer sur l'onglet Keys and Access Token

Récupération des informations

Les informations à collecter pour se connecter depuis Talend sont :

- API Key
- API Secret
- Access Token
- Access Token Secret

Informations pratiques

Collecte et structuration des données

Au travers de Talend : Collectez les Tweets liés à Dataswati et construisez le graphe correspondant.

Recherche dans le graphe

Présenter les tweets qui ont été Tweeté et Retweeté

Identifier le hashtag le plus utilisés

Identifier le plus court chemin entre le User: Dataswati et les autres utilisateurs

Trouver la requête permettant d'identifier les mots-clés liés à Dataswati à 2 niveaux de profondeur

Création de connaissances supplémentaires

Faire un graphe secondaire à partir des informations à disposition dans le graphe.

Faire une série de requêtes permettant de lier HASHTAG et USER via la notion d'INFLUENCE au travers de leur tweet ou retweet. 1 tweet vaut 3 points et 1 retweet 1.

Présenter le hashtag qui a le plus d'influence

Présenter le hashtag le plus utilisé/populaire

Agrandissement du graphe

Ouvrir le graphe de Twitter à l'expression #IA

Identifier le hashtag le plus utilisé

Identifier le plus court chemin entre les éléments User:Dataswati et le hashtag IA

Recréer les influences sur les mots-clés (pensez à supprimer les anciennes valeurs)

Identifier les utilisateurs les plus proches du hashtag #IA

Identifier le chemin pour y accéder avec Dataswati (pensez à rechercher le parcours le plus court avant)