
Collection and Analysis of Tweets made Simple



Membres du groupe

Abdelkader Messeguem
Selim Dogan
Cihan Cinar
Hicham Boubouh
Nathanaël Desjobert

Encadrant : Fabien Rico
Chef d'équipe: Anthony Deseille

SOMMAIRE

1. DECOMPOSITION EN TACHES	2
2. EVALUATION DE LA CHARGE DES TACHES ET REPARTITION	2
3. LIVRABLES TECHNIQUES	3
4. ETAT D'AVANCEMENT DES TACHES	3
5. MODIFICATIONS DU PROJET ET ACTIONS CORRECTIVES	4

1. DECOMPOSITION EN TACHES

L'objectif du sprint 1 est la réalisation de la version minimale de l'application CATS. Cette première version permet à un utilisateur de se connecter sur l'application avec un compte Twitter. Il faut donc gérer l'authentification via Twitter. Elle permet aussi de lancer une collecte avec un ou plusieurs filtres afin de récupérer des tweets. Enfin, elle permet de stocker les corpus de tweets dans une base de données avec la possibilité de les exporter au format CSV. Toutes ces fonctionnalités seront utilisables depuis des interfaces web que nous allons créer.

2. EVALUATION DE LA CHARGE DES TACHES ET REPARTITION

Dans le tableau suivant, nous avons détaillés l'évaluation de la charge des tâches avec leurs répartitions afin de visualiser quelle quantité de travail est prévue sur chaque tâche.

TACHES	CHARGE DE TRAVAIL	REPARTITION
Mise en place du projet	1/2 journée	Nathanaël
Formation Spring	1/2 journée	Selim, Cihan
Formation Lucene	1/2 journée	Abdelkader, Hicham
Création de la base de données	1 journée	Abdelkader, Anthony
Authentification Twitter	1,5 journée	Nathanaël
Collecte de Tweets	1 journée	Anthony
Manipulation de la base de données	2 journées	Anthony, Abdelkader, Hicham
Interface de connexion	1/2 journée	Nathanaël
Interface de création de collectes	1 journée	Nathanaël, Selim, Cihan
Interface de visualisation de corpus	1 journée	Selim, Cihan
Export en CSV	1/2 journée	Nathanaël
Documentations	1 journée	Selim, Cihan
Tests	1/2 journée	Selim, Cihan
Gestion du projet	2 journées	Anthony

3. LIVRABLES TECHNIQUES

A la fin du sprint 1, la première version de l'application doit être terminée afin d'être présentée à notre encadrant Monsieur Rico et aux doctorants de l'Université Lumières Lyon 2.

4. ETAT D'AVANCEMENT DES TACHES

TACHES	PREVISIONS	AVANCEMENTS
Mise en place du projet	100 %	100 %
Création de la base de données	100 %	100 %
Authentification Twitter	100 %	100 %
Collecte de Tweets	100 %	90 %
Manipulation de la base de données	100 %	100 %
Interface de connexion	100 %	90 %
Interface de création de collectes	100 %	90 %
Interface de visualisation de corpus	100 %	90 %
Import Export en CSV	100 %	100 %

A la fin du sprint 1, la collecte de tweets avec un ou plusieurs filtres dans une durée donnée a été implémentée. Cependant, un bug empêche son lancement. Le marquage automatique du corpus pour la détection automatique du sexe de l'auteur n'a pas été implémentée par manque de temps.

La base de données a été créée, et elle permet d'enregistrer des tweets et des utilisateurs. Comme prévu, l'authentification se fait via Twitter.

Les interfaces ont toutes été créées, mais doivent être améliorées au niveau du design. De plus, dans l'interface de visualisation de corpus, il manque la pagination pour plus de visibilité. Enfin, l'interrogation de corpus avec différents filtres n'a pas été implémentée par manque de temps. L'export et l'import en CSV est totalement fonctionnel.

5. MODIFICATIONS DU PROJET ET ACTIONS CORRECTIVES

Nous avons décidé d'utiliser la bibliothèque de recherche Lucene pour le traitement des données. Finalement, nous avons décidé d'utiliser SOLR, une plateforme logicielle de moteur de recherche qui s'appuie sur Lucene car cela nous semble plus facile et rapide.

De plus, nous avons revu nos objectifs au fil de la semaine. Ainsi, nous avons du créer la partie métier de façon plus modulaire ce qui nous a fait perdre du temps dans nos tâches initiales.

Enfin, nous avons aussi décidé de gérer les droits des utilisateurs avec Spring Security.