
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 2 est l'amélioration et l'ajout des fonctionnalités de l'application CATS. Dans cette version, nous allons ajouter de nouvelles fonctionnalités pour l'analyse de corpus ainsi que pour les méthodes de filtrages. De plus, nous allons améliorer l'interface web afin d'obtenir de meilleurs aperçus des résultats (par exemple lemmatisation des tweets). Enfin, durant la réunion pré-sprint avec notre encadrant, de nouvelles fonctionnalités nous ont été demandées. Nous devons dans ce sprint créer des modules qui pourront être intégrés à l'application. Un module peut par exemple être un affichage d'un nuage de mots selon un corpus donné.

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
Repérer les entités nommées et leur assigner un type	1/2 journée	Anthony,Abdelkader
Lemmatisation des tweets	1 journée	Anthony,Abdelkader
Entités nommées avec les fréquences	1/2 journée	Anthony,Abdelkader
Expressions régulières pour filtres	1/2 journée	Nathanaël ,Hicham
Création de modules externe au projet	1,5 journée	Selim,Cihan
Marquage de tweets détecté comme robot	1/2 journée	Selim,Cihan
Téléchargement d'un sous-corpus	1 journées	Nathanaël ,Hicham
Amélioration du design	1 journée	Selim
Correction bug de collecte	1/2 journée	Anthony
Visualisation d'un sous-corpus	1/2 journée	Nathanaël
Pagination dans un corpus	1/2 journée	Cihan
Administration de la VM	1 journée	Anthony
Mise en place création modules	1,5 journée	Nathanaël ,Hicham

3. LIVRABLES TECHNIQUES

A la fin du sprint 2, les nouvelles fonctionnalités doivent être ajoutées à l'application et les bugs présent doivent être corrigés. Nous allons ensuite présenté le travail réalisé à notre encadrant Monsieur Rico et aux doctorants de l'Université Lumières Lyon 2.

4. ETAT D'AVANCEMENT DES TACHES

TACHES	PREVISIONS	AVANCEMENTS
Repérer les entités nommées et leur assigner un type	100 %	90 %
Lemmatisation des tweets	100 %	80 %
Entités nommées avec les fréquences	100 %	90 %
Expressions régulières pour filtres	100 %	90 %
Création de modules externe au projet	100 %	90 %
Marquage de tweets détecté comme robot	100 %	20 %
Téléchargement d'un sous-corpus	100 %	90 %
Amélioration du design	100 %	20 %
Correction bug de collecte	100 %	100 %
Visualisation d'un sous-corpus	100 %	100 %
Pagination dans un corpus	100 %	100 %
Administration de la VM	100 %	100 %
Mise en place création modules	100 %	100 %

A la fin du sprint 2, les fonctions avancées n'ont pas été toutes terminées. La lemmatisation des tweets n'a pas pu être complètement implémentée. En effet, il n'est toujours pas fonctionnel pour le français. De plus, toutes ces fonctions doivent être mis en place dans un module. Ainsi, même si les fonctions fonctionnent, il faudra les implémenter dans un seul module unique. Nous n'avons pas eu le temps de le faire.

La création de modules a été mis en place. Il est maintenant possible d'ajouter des modules dans l'application. Ainsi, nous avons créé un module qui affiche un nuage de mots selon un corpus de tweets donné. Ce module fonctionne parfaitement avec un jeu de données de test. Nous n'avons pas eu le temps de le tester avec un vrai corpus.

De plus, le téléchargement d'un sous-corpus et sa visualisation sont fonctionnels.

Enfin, tout les bugs de l'ancienne version ont été corrigés. La VM a été configuré correctement. Les tâches non réalisées durant le sprint 1 telle que la pagination ont été faite. Cependant, le design n'a pas pu être amélioré, et le marquage de tweets détectés comme robot n'a pas pu être implémenter par manque de temps.

5. MODIFICATIONS DU PROJET ET ACTIONS CORRECTIVES

Nous avons décidé de créer des modules en REST pour l'application.

De plus, nous avons corrigés des problèmes de redondances dans la base de données.

Enfin, nous avons aussi décidé d'utiliser Stanford NLP pour la lemmatisation et les entités nommés.