

***Réalisé par :***

Medhi Beauvillain

Yanis Cherifi

Samy Hamiche

Damien Jacquet

**Cahier**

**Des**

**Charges**

**Fonctionnel**

***Encadré par :***

Daniel Alban

Lazhar Labiod

Nicoleta Rogovschi

Interface web interactive pour le Text Mining :

*Analyse exploratoire des Tweets*

**Projet**

**Pluridisciplinaire**

Table des matières

[I. Présentation générale du problème 2](#_Toc447389465)

[A. Projet 2](#_Toc447389466)

[1. Le demandeur 2](#_Toc447389467)

[2. Finalité 2](#_Toc447389468)

[3. Espérance de retour sur investissement 2](#_Toc447389469)

[B. Contexte 3](#_Toc447389470)

[1. Suites prévues 3](#_Toc447389471)

[2. Nature des prestations demandées 3](#_Toc447389472)

[3. Parties concernées par le déroulement du projet et ses résultats 3](#_Toc447389473)

[4. Caractère confidentiel 3](#_Toc447389474)

[C. Énoncé du besoin 4](#_Toc447389475)

[II. Expression fonctionnelle du besoin 5](#_Toc447389476)

[A. Fonctions de service et de contrainte 5](#_Toc447389477)

[1. Principales fonctionnalités attendues 5](#_Toc447389478)

[2. Fonctionnalités annexes 5](#_Toc447389479)

[3. Diagramme de Cas d’Utilisation 5](#_Toc447389480)

[4. Contraintes 6](#_Toc447389481)

[B. Critères d’appréciation 7](#_Toc447389482)

[III. Cadre de réponse 8](#_Toc447389483)

[A. Pour chaque fonction 8](#_Toc447389484)

[1. Solution proposée 8](#_Toc447389485)

[B. Pour l’ensemble du produit 11](#_Toc447389486)

[1. Prévisions de fiabilité 11](#_Toc447389487)

[2. Perspective d’évolution technologique 11](#_Toc447389488)

# I. Présentation générale du problème

## A. Projet

### 1. Le demandeur

Dans le cadre de notre Projet Pluridisciplinaire (PPD), Nicoleta Rogovschi et Lazhar Labiod, tous deux professeurs à l’Université Paris Descartes, ont proposé le sujet suivant : la réalisation d’une interface web interactive pour l’analyse exploratoire de tweets.

### 2. Finalité

Notre objectif est de réaliser une plateforme qui permet d’automatiser la structuration de documents peu organisés. Cette nouvelle interface offrira à ses utilisateurs la possibilité de générer différentes informations et statistiques à partir d’un texte afin d’enrichir celui-ci.

Cette plateforme comporte plusieurs aspects principaux : extraction des données à partir du réseau sociale Twitter, prétraitement et transformation du texte, analyse exploratoire et affichage interactif des résultats.

### 3. Espérance de retour sur investissement

L’objectif principal de ce projet est de mettre en place une plateforme qui rend accessible l’analyse exploratoire à tous les utilisateurs, qu’ils soient novices ou confirmés dans la matière.

De plus, Twitter étant un flux de données très dense et fréquemment renouvelé, l’interface mise en place permettra d’analyser un grand volume de données afin de déterminer des tendances sur Twitter.

## B. Contexte

### 1. Suites prévues

La plateforme développée devra être aisée à maintenir afin de pouvoir proposer à ses utilisateurs des perspectives d’améliorations comme l’ajout de nouvelles statistiques et représentations graphiques ou de nouveaux filtres de recherches.

### 2. Nature des prestations demandées

Le prestataire assure la rédaction du présent cahier des charges à partir de documents de présentation du sujet et des spécifications du besoin.

L’équipe réalisant la prestation s’engage à fournir lors de la livraison les livrables suivants :

* le livrable contenant le code de la plateforme
* la présentation fonctionnelle de la plateforme

### 3. Parties concernées par le déroulement du projet et ses résultats

Les parties concernées par le projet sont les suivantes :

* Clients :
  + Nicoleta ROGOVSCHI
  + Lazhar LABIOD
* Maîtrise d’Ouvrage :
  + Daniel ALBAN
* Maîtrise d’Œuvre :
  + Medhi BEAUVILLAIN
  + Yanis CHERIFI
  + Samy HAMICHE
  + Damien JACQUET

### 4. Caractère confidentiel

La plateforme à développer ne comporte aucune section confidentielle.

La source des données étant une source de donnée publique (le réseau social Twitter), l’ensemble des résultats de la plateforme est public et accessible à tous les utilisateurs.

## C. Énoncé du besoin

Le projet une fois réalisé devra offrir une interface de recherche de tweets avec différents filtres. L’application proposera donc une recherche de tweets avec les possibilités suivantes :

* Choisir la quantité de tweets extraite
* Définir la fréquence minimale d’un mot pour qu’il soit traité
* Rechercher à partir de mots-clés
* Rechercher un utilisateur
* Définir le nombre de mots analysés
* Sélectionner la langue des tweets
* Télécharger les données
* Options de recherche avancée (exclusion de mots, préciser une fréquence minimale)

# II. Expression fonctionnelle du besoin

## A. Fonctions de service et de contrainte

L’application est disponible en libre accès sur le web. Il n’y a pas de restrictions d’accès pour les utilisateurs, pas d’authentification demandée pour accéder aux différents services de l’application. On peut citer différents exemples d’usage en fonction du profil des différents utilisateurs :

* **Visiteur** : Visite par curiosité avec analyse des tweets lié à un thème ou analyse de ses propres tweets ou des tweets de d’autres utilisateurs.
* **Académie** : Exemple d’analyse possible avec R et fouille de données.
* **Journalistes** : Analyse des tweets liés à phénomène ou d’une tendance.

### 1. Principales fonctionnalités attendues

L’ensemble des utilisateurs de l’application ont accès aux fonctionnalités suivantes :

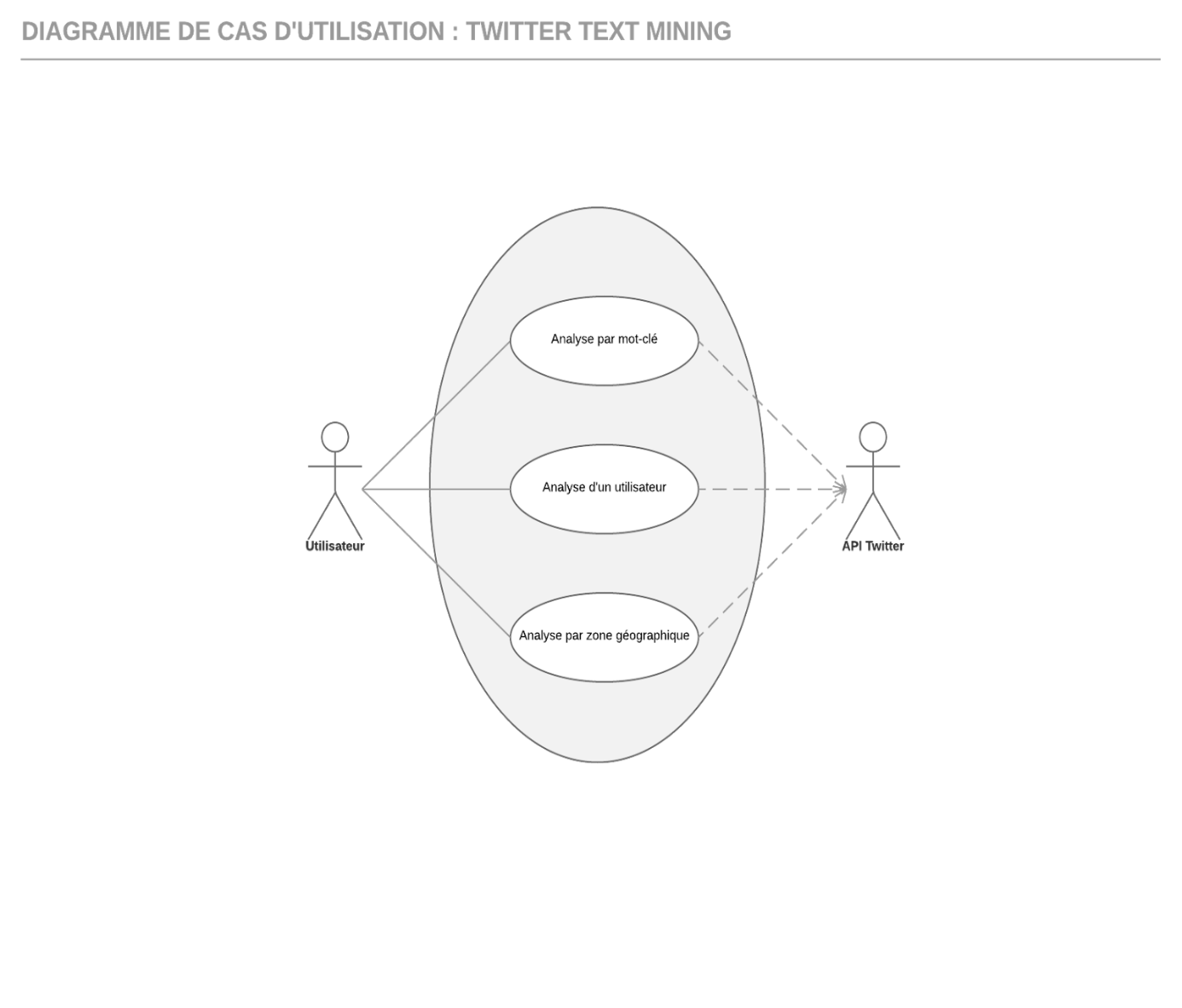
* *F1 : Analyse des tweets liés à un thème*
* *F2 : Analyse des tweets liés à un utilisateur*
* *F3 : Analyse des tweets liés à un lieu*

L’ensemble de ces fonctionnalités seront détaillées dans la partie Cadre de réponse.

### 2. Fonctionnalités annexes

Aucune fonctionnalité annexe n’est prévue dans le cadre de ce projet.

### 3. Diagramme de Cas d’Utilisation



1. DCU de l'application

### 4. Contraintes

La réalisation de ce projet est soumise aux contraintes suivantes :

* **C1 :** La plateforme devra être développée avec le langage R étudié dans le cadre du module de Fouille de Données et utiliser les libraires *shiny*, *textmining*, *wordcloud*, et *twitteR*.
* **C2 :** L’interface devra être une application Web client-serveur développée avec *Rstudio*.
* **C3 :** La source des données utilisée devra être Twitter.

## B. Critères d’appréciation

Afin de pouvoir juger de la qualité de la réponse proposée, certains critères ont été fixés à partir de l’analyse des besoins exprimés par le client.

La facilité de la prise en main de la plateforme est un élément important à prendre en compte. La pertinence des résultats et la profondeur d’analyse des sources de données est un critère essentiel étant donné que la plateforme a pour objectif principal de collecter, et traiter un grand volume de données.

# III. Cadre de réponse

## A. Pour chaque fonction

### 1. Fonction F1

**FONCTION F1 :** Analyse des tweets liés à un thème

**OBJET :** Création d’un wordcloud, d’un tableau des tweets, d’un tableau de fréquences d’apparition des mots et d’une analyse sur le caractère des tweets (positif, neutre ou negatif).

**UTILISATEUR :** Tout le monde

**ENTREES :** Formulaire avec les critères de recherche

**SORTIES :** Wordcloud, Tableau des tweets, Tableau de fréquences d’apparition des mots et graphique sur le caractère des tweets (positif, neutre ou negatif).

**MODE DE FONCTIONNEMENT :**

L’utilisateur rempli un formulaire avec les différents critères de recherche suivant :

* Le mot clé à rechercher : CHAINE DE CARACTERES.
* Nombre de tweets voulus : NOMBRE ENTIER.
* Langue des tweets : CHAINE DE CARACTERES.
* Fréquence minimale d’un mot : NOMBRE ENTIER.
* Période du tweet, date inférieure : DATE.
* Période du tweet, date supérieure: DATE.

### 2. Fonction F2

**FONCTION F2** : Analyse des tweets liés à un utilisateur

**OBJET :** Création d’un wordcloud, d’un tableau des tweets, d’un tableau de fréquences d’apparition des mots et d’une analyse sur les tweets d’un utilisateur (positif, neutre ou negatif).

**UTILISATEUR :** Tout le monde

**ENTREES :** Formulaire avec les critères de recherche

**SORTIES :** Wordcloud, Tableau des tweets, Tableau de fréquences d’apparition des mots et graphique sur le caractère des tweets (positif, neutre ou negatif).

**MODE DE FONCTIONNEMENT :**

L’utilisateur rempli un formulaire avec les différents critères de recherche suivant :

* Le nom de l’utilisateur : CHAINE DE CARACTERES.
* Nombre de tweets voulus : NOMBRE ENTIER.
* Langue des tweets : CHAINE DE CARACTERES.
* Fréquence minimale d’un mot : NOMBRE ENTIER.
* Période du tweet, date inférieure : DATE.
* Période du tweet, date supérieure: DATE.

### 3. Fonction F3

**FONCTION F3 :** Analyse des tweets liés à un lieu

**OBJET :** Création d’un wordcloud, d’un tableau des tweets, d’un tableau de fréquences d’apparition des mots et d’une analyse sur le caractère des tweets (positif, neutre ou negatif).

**UTILISATEUR :** Tout le monde

**ENTREES :** Formulaire avec les critères de recherche

**SORTIES :** Wordcloud, Tableau des tweets, Tableau de fréquences d’apparition des mots et graphique sur le caractère des tweets (positif, neutre ou negatif).

**MODE DE FONCTIONNEMENT :**

L’utilisateur remplit un formulaire avec les différents critères de recherche suivant :

* Le lieu où les tweets sont envoyés : CHAINE DE CARACTERES.
* Nombre de tweets voulus : NOMBRE ENTIER.
* Langue des tweets : CHAINE DE CARACTERES.
* Fréquence minimale d’un mot : NOMBRE ENTIER.
* Période du tweet, date inférieure : DATE.
* Période du tweet, date supérieure: DATE.

## B. Pour l’ensemble du produit

### 1. Prévisions de fiabilité

L’application doit être conçue pour être accessible à tout moment depuis n’importe quel poste à n’importe quel endroit. Cela dit il n’existe pas de plage de service garantie.

### 2. Perspective d’évolution technologique

L’interface doit pouvoir être évolutive afin de pouvoir mettre à jour assez simplement les formulaires de recherche et ajouter facilement de nouveaux outils d’analyse statistique avec différents graphiques, ou de nouvelles fonctions.

### 3. Planning

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Lundi** | **Mardi** | **Mercredi** | **Jeudi** | **Vendredi** | **Commentaires** |
| **11/01/2016** |  |  |  |  |  | Documentation sur les logiciels et les libraires à utiliser. |
|  |  |  |  |  | PPD |
| **18/01/2016** |  |  |  | PPD |  |
|  |  |  |  |  | PPD |
| **25/01/2016** |  |  |  |  |  |
|  |  |  |  |  | PPD |
| **01/02/2016** |  |  |  | PPD |  | Répartition des rôles et initialisation du projet |
|  |  |  |  |  | PPD |
| **08/02/2016** |  |  |  |  |  | **Equipe Fonctionnelle :** Création du dépôts pour le projet Rédaction des comptes rendus Initialisation du Cahier des charges  **Equipe Technique :** Initialisation du projet |
|  |  |  |  | PPD |  |
| **15/02/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **22/02/2016** |  |  | PPD |  |  |
|  |  |  |  |  | PPD |
| **29/02/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **07/03/2016** |  |  |  |  |  | **Equipe Fonctionnelle :** Rédaction des comptes rendus Rédaction du Cahier des charges  **Equipe Technique :** Développement F1 Phase de tests |
|  |  |  |  |  |  |
| **14/03/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **21/03/2016** |  |  |  | PPD |  |
|  |  |  |  |  |  |
| **28/03/2016** | Férié |  |  |  |  |
|  | Férié |  |  |  | PPD |
| **04/04/2016** |  |  |  |  |  |
|  |  |  |  |  | PPD |
| **11/04/2016** |  |  |  |  |  |
|  |  |  |  |  | PPD |
| **18/04/2016** |  |  |  |  |  |
|  |  |  |  |  | PPD |
| **25/04/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **02/05/2016** |  |  |  | Férié |  | **Equipe Fonctionnelle :** Rédaction des comptes rendus Rédaction du Cahier des charges Finalisation du Cahier des charges Tests fonctionnels  **Equipe Technique :** Développement F2 Développement F3 Phase de tests |
|  |  |  |  | Férié |  |
| **09/05/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **16/05/2016** | Férié |  |  |  |  |
|  | Férié |  |  |  |  |
| **23/05/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **30/05/2016** |  |  |  |  |  |
|  |  |  |  |  | PPD |
| **06/06/2016** |  |  |  |  |  |
|  |  |  |  |  |  |
| **13/06/2016** |  |  |  |  |  |
|  |  |  |  | PPD | PPD |
| **20/06/2016** |  |  |  |  |  | **Présentation du Projet** |
|  |  |  |  | **PPD-Sout** |  |

2. Planning

**Equipe Fonctionnelle :** BEAUVILLAIN Medhi, HAMICHE Samy

**Equipe Technique :** CHERIFI Yanis, JACQUET Damien

# Annexes :

## Table des illustrations

[1. DCU de l'application 8](#_Toc450819340)

[2. Planning 15](#_Toc450819341)