



Rapport de projet innovant

Recherche et extraction d'information sur le réseaux social Twitter



Iliasse Ben ammar

Nalan Ramakichenin

Année universitaire : 2021-2022

Tuteur pédagogique : Nasreddine.bouhai

Table des matières

I.	Introduction	3
II.	Problématique	4
III.	État de l'art.....	5
A.	Article 1 : Extraction d'information du réseau social pour la commercialisation de produits agricole.....	6
B.	Article 2 : CredFinder Un système d'évaluation de la crédibilité des tweets en temps réel 7	
C.	Article 3 : Extraction de données commerciales à partir des réseaux sociaux.....	9
IV.	Présentation de Twitter.....	11
V.	La structure de twitter.....	12
VI.	Les différentes méthodes d'extraction de données.....	13
VII.	Définition d'une API.....	14
VIII.	La stratégie et la méthode adopté.....	15
IX.	Les outils utilisés.....	16
X.	Architecture du projet.....	19
XI.	Présentation de l'application	20
A.	Démonstration du Projet.....	22
XII.	Conclusion	23
XIII.	Références.....	24

I. Introduction

De nos jours, des informations utiles sont intégrées un peu partout sur internet comme dans une page html ou les réseaux sociaux par exemple. Et si l'on pouvait récupérer et utiliser ces informations à diverses fins ? La croissance d'Internet nous a permis de constater que l'obtention ainsi que la diffusion d'informations deviennent de plus en plus accessibles, ce qui signifie que les utilisateurs ont accès aux données depuis n'importe où et n'importe quand. Les réseaux sociaux sont de nos jours présents et utilisés très fréquemment. Pour de la prise d'information ou pour relayer de l'actualité provenant des médias, certains réseaux sociaux sont même plus rapides que les médias, comme Twitter par exemple.

Cependant, ces informations peuvent tout aussi bien être des fakes news, visant à diffamer des entreprises, des marques, des personnalités connues, des communautés.

C'est pourquoi les réseaux sociaux ont un certain rôle à jouer, d'une part, ils se doivent de laisser la liberté d'expression, d'autre part, ils se doivent de supprimer les cas trop graves de fake news. Aujourd'hui, près de 80% des marques sont présentes sur les réseaux sociaux. Elles utilisent ces moyens de communication pour gérer la relation et la communication avec leurs cibles. Leur contenu peut donc être analysé pour être ensuite utilisé dans un domaine d'activité comme la commercialisation.



II. Problématique

L'extraction de données du Web est un problème important qui a été largement étudié à l'aide de différents outils et applications scientifiques. De nombreuses approches ont été proposées et conçues pour résoudre des problèmes spécifiques et fonctionner dans divers domaines. Aujourd'hui, les entreprises, les marques, les influenceurs, les médias fournissent des informations utiles sur les réseaux sociaux comme Facebook, Instagram, Twitter ou YouTube. Les petites organisations qui n'ont pas de site web personnel s'appuient sur les réseaux sociaux pour offrir leurs services.

Il y a également certains médias qui ont connu le jour grâce aux réseaux sociaux. C'est pourquoi nous visons à extraire des données provenant des petits media comme "mediavenir" et "cerfia" qui sont des médias provenant de twitter suivis par des millions d'utilisateurs. Ces médias sont toujours à la pointe de l'actualité et sont généralement plus rapides et pertinents que certains gros médias de la télévision ou autres.

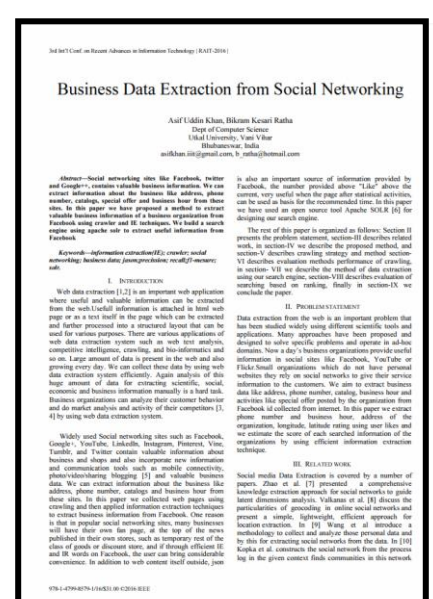
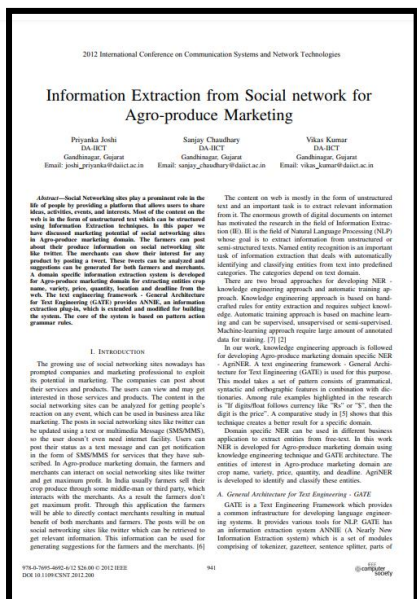
Une fois les données extraites nous voulons les incorporer de manière à les présenter sur un site intuitif et facile à interpréter et ainsi stocker toutes les informations dans la base de données. Il se peut que les informations mises à disposition ne proviennent pas forcément de sources fiables. La véracité de celles-ci peut varier car il n'y pas toujours de vérification effectuée derrière.

On en revient donc à se dire : **Est-il possible de réaliser une application web monopage en utilisant des technologies récentes (API avec des technologies d'extraction de données, etc.), pour stocker/lire ces données via une base de données ?**

III. État de l'art

Notre recherche d'articles s'est faite majoritairement sur le site **IEE explorer**. On a sélectionné 3 articles qui nous semblaient pertinents.

- Extraction d'informations du réseau social pour commercialisation des produits agricoles
- Un système d'évaluation de la crédibilité des Tweets en temps réel
- Extraction de données commerciales à partir des réseaux sociaux



A. Article 1 : Extraction d'information du réseau social pour la commercialisation de produits agricole.

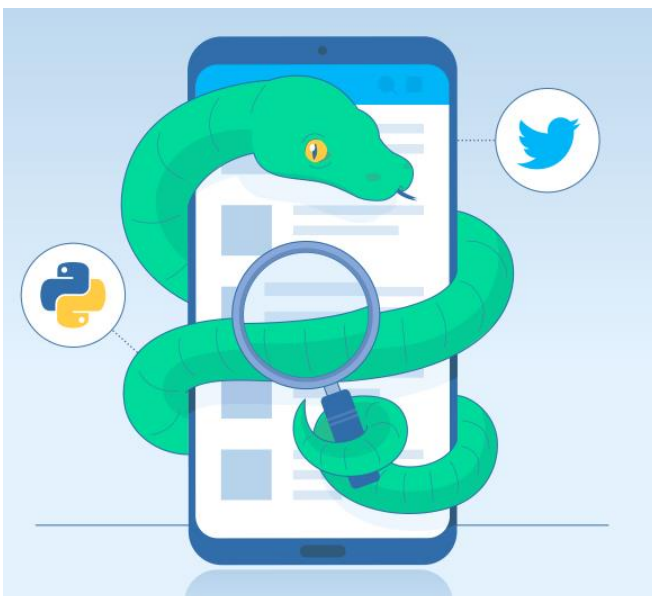
Les réseaux sociaux jouent un rôle important dans la vie des gens en fournissant une plate-forme qui permet aux utilisateurs de partager des idées, activités, des événements et des intérêts. La plupart des contenus sur le web se présentent sous la forme de texte non structuré, qui peuvent être structurés à l'aide de techniques d'extraction de données.

Dans cet article, il sera question du potentiel marketing des plateformes de réseaux sociaux dans le domaine de la commercialisation des produits agroalimentaires. Les agriculteurs peuvent poster des informations à propos de leurs produits sur une plateforme de réseaux social comme twitter. Les commerçants peuvent montrer leur intérêt pour un produit en particulier en publiant un tweet. Ces tweets peuvent donc être analysés et des suggestions peuvent être générées pour les agriculteurs et les commerçants.

L'objectif de cet article est d'exploiter le potentiel marketing des sites des réseaux sociaux au profit de la société. Il y a un utilisateur central et les autres utilisateurs qui sont abonnés à cet utilisateur central.

Les Tweets ont une structure prédéfinie : "@CentralUser AgroSell", Type de produit, sous-type de produit, Prix unitaire minimum, Quantité, Date etc.

Les tweets sont récupérés sur le réseaux social Twitter. Le processus de récupération des données utilise les API Twitter avec une architecture REST (Representational State Transfer). Les API renvoient les tweets en format JSON, XML, RSS et ATOM. Chaque tweet est associé à un utilisateur.



B. Article 2 : CredFinder Un système d'évaluation de la crédibilité des tweets en temps réel

Twitter est l'un des réseaux sociaux les plus privilégiés pour diffuser de l'information aux personnes autour du globe. Cependant, le principal défi auquel sont confrontés les utilisateurs est de savoir comment vérifier la véracité des informations publiées sur ce réseau. Dans cet article, un système d'évaluation de la crédibilité du contenu nommé "CredFinder" sera présenté. Ce système est capable de mesurer la fiabilité des informations par l'analyse des utilisateurs et l'analyse du contenu. Il est également capable de fournir un score de crédibilité pour chaque tweet.

Par conséquent, il offre aux utilisateurs la possibilité de juger la crédibilité de l'information plus rapidement. Souvent, les personnes malveillantes utilisent Twitter comme moyen de répandre des rumeurs visant à diffamer des marques ou même en cas de politique. Ces informations peuvent être présentées sous la forme d'images qui ont été modifiées pour s'adapter à la stratégie d'attaque ciblée. Les informations de cette nature sont généralement difficiles à vérifier, ce qui peut conduire des utilisateurs naïfs à propager des informations qui n'ont aucune crédibilité. Dans certains cas, la presse écrite pourrait finir par être impliquée dans un tel scénario, à cause d'un manque de temps pour vérifier correctement les sources dont ils tirent leurs informations. Les chercheurs ont déjà établi que le réseau social Twitter pouvait être utilisé en cas d'urgence compte tenu de sa capacité extrêmement rapide à réduire le temps nécessaire à la communication des informations. La difficulté est donc d'évaluer la crédibilité des informations publiées, ce qui est l'objectif du travail présenté dans cet article.

1) Qu'est-ce que le système CredFinder ?

Le système CredFinder est un outil d'extraction de données qui évalue la crédibilité des tweets en temps réel. Le système CredFinder peut classer les tweets en fonction de leur fiabilité. Ainsi, il calcule un score de crédibilité pour les tweets. Le processus de calcul du score de crédibilité implique une communication entre trois parties : (I) une extension de navigateur Web, (II) le serveur système et (III) et Twitter.

Le système fonctionne en 2 parties, le côté client et le côté serveur.

- Côté client :

L'utilisateur doit télécharger une extension de navigateur web nommée "CredFinder".

Cette extension collecte donc les ID des tweets que les utilisateurs publient par rapport à un sujet ou un événement spécifique. CredFinder génère alors une liste avec tous les tweets qui sont en correspondance avec le sujet souhaité avec une note en crédibilité, ici des étoiles. Une étoile sur cinq signifie que le tweet est d'une fiabilité médiocre.

- Côté serveur :

La première étape côté serveur consiste à générer des “clés uniques” (token key). Ensuite, 2 paramètres (TID, TokenKey) sont envoyés à une fonction de validation, qui vérifie la clé. Si la clé est invalide alors un message est envoyé à l'utilisateur pour lui demander de générer une autre clé, en d'autres termes l'ID du tweet est utilisé pour aller chercher le tweet original.

Résultat :

Le système CredFinder a montré au cours des tests qu'il a réalisé d'excellentes performances en termes de temps de réponse. Cependant, comme le système n'en est qu'à ses débuts, l'analyse est limitée à seulement 70 utilisateurs à ce jour. Les notes sont généralement bonnes. Mais le système doit tout de même être testé par plusieurs utilisateurs.



C. Article 3 : Extraction de données commerciales à partir des réseaux sociaux

L'extraction de données peut s'avérer très utile pour l'extraction d'informations pertinentes et concrètes. Ces informations seront alors transformées et pourront être utilisées à des fins diverses. Les informations peuvent être extraites de manière différente, chaque méthode sont des méthodes d'extraction unique. Par exemple, un système d'extraction d'analyse de texte. Cette méthode consiste à extraire des données par rapport à un texte donné.

Les données sont présentes partout sur internet et dans notre quotidien également. On peut les récupérer en utilisant ces systèmes d'extraction de données de manière efficace. Par exemple, extraire des données scientifiques, sociales ou encore des informations économiques et commerciales manuellement est une tâche qui n'est pas évidente. Grâce à cela les entreprises peuvent utiliser ces données à des fins commerciales comme analyser le comportement de leur client, etc.

Les réseaux sociaux comme Twitter, Facebook, YouTube, Instagram, regorgent de données plus ou moins exploitables comme par exemple des données sur le business ou sur le commerce. Les données extraites peuvent être l'adresse, le numéro de téléphone, etc.

Sur cet article, les données ont été extraites via la technique du "data crawling".

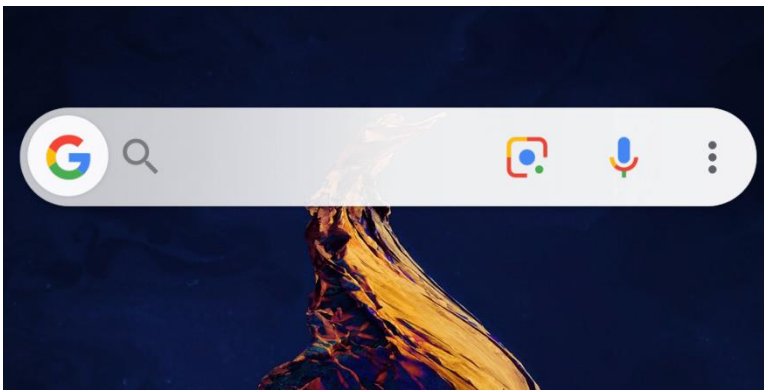
Le data crawling consiste à creuser profondément dans les recoins du web pour récupérer des éléments sur lesquels on aurait pu passer à côté. Ce sont des bots (des robots) qui fouillent le web pour trouver tout ce qui est pertinent par rapport à ce que l'on cherche.

Les bots agissent sur un algorithme pour suivre des instructions. Le data crawling fonctionne un peu comme le moteur de recherche Google. Le processus suit des liens vers de nombreuses pages différentes. Ils ne se contentent pas de parcourir les pages, ils collectent toutes les informations pertinentes en les indexant au cours du processus, ils recherchent également tous les liens vers les pages pertinentes du processus. Ils peuvent extraire des informations en double d'un article de blog qui ont peut-être été copié-collé car ils ne connaissent pas la différence.

Résultat :

Le data crawling a été utilisé pour collecter les ID Facebook.

La stratégie “crawler” est divisé en deux blocs, un pour obtenir un les ID Facebook professionnel et l’autre pour le web content et le json.



IV. Présentation de Twitter

Twitter est un réseau social se focalisant sur des messages appelés "tweets" privés et publics.

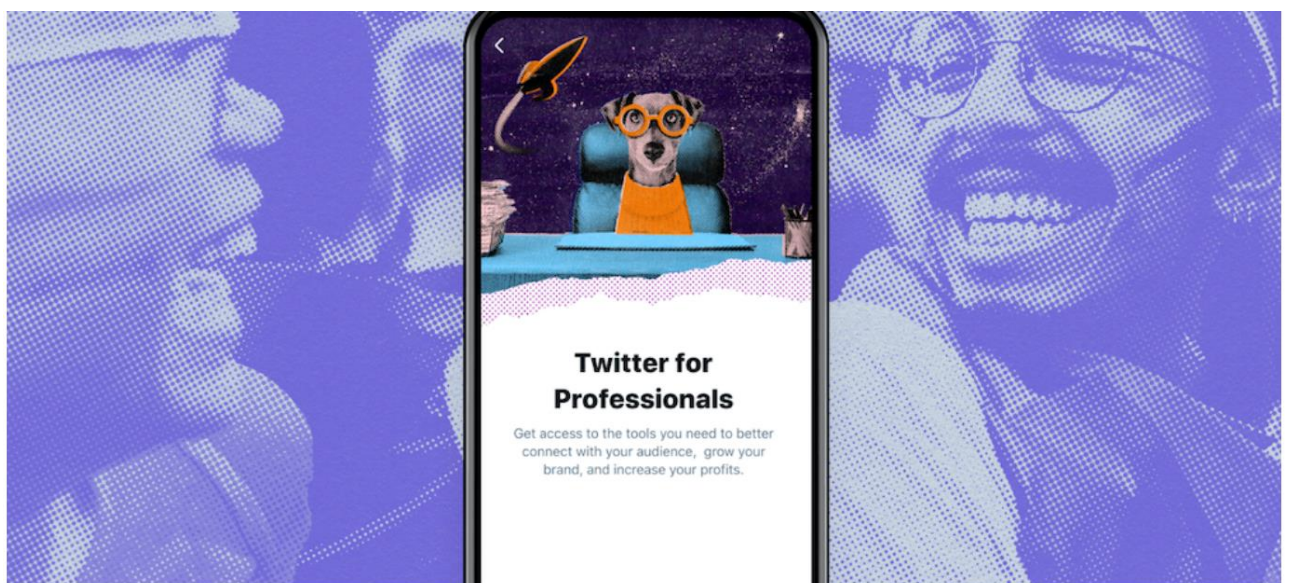
Ces messages peuvent notamment contenir du texte, des liens, des photos, vidéos.

Fonctionnalités principales :

- "Tweeter" = publier un message public
- "Retweeter" = partager le message public de quelqu'un d'autre
- Répondre à un tweet public
- Écrire un message privé à une personne ou une société
- S'abonner à d'autres personnes ou sociétés pour suivre les messages qu'ils publient.

Contrairement à d'autres réseaux sociaux comme Facebook, Snapchat ou autres, Twitter ne possède qu'une seule entité : le compte Twitter. Que vous soyez une personne, une association, une entreprise, le compte est le même. Une certification est présente pour le compte de personnalité connue mais le contenu du compte reste le même.

Pour la création ainsi que la connexion à un compte Twitter, il suffit d'une adresse e-mail et d'un mot de passe propre à chaque compte.

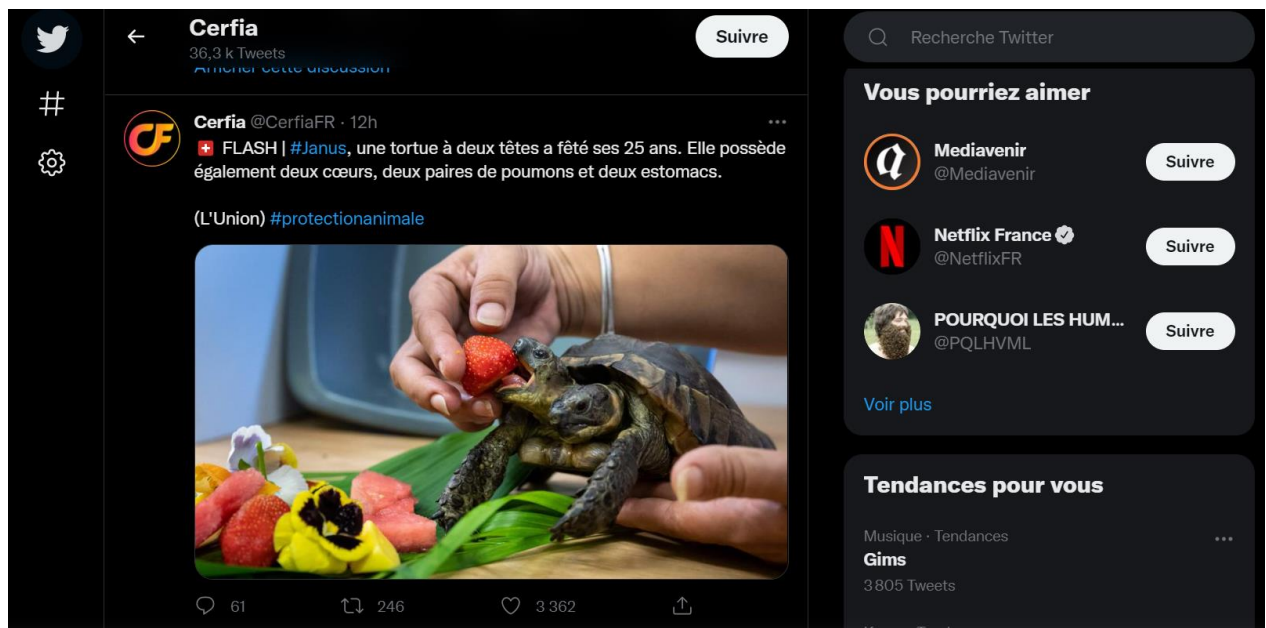


V. La structure de twitter

Le réseau social twitter est composé de plusieurs tweets venant du monde entier affiché sur une tweet List ou time line. Un tweet est un message de 140 caractères au maximum, émis sur une plateforme de micro-blogging ouverte : Twitter.

Un tweet ne peut être émis que par un compte Twitter et n'a pas de durée de péremption et seul l'auteur d'un tweet peut décider de le supprimer.

Exemple de Time Line :



La timeline est la fonctionnalité principale de twitter, celle-ci nous montre tous les tweets des comptes que l'on suit dans un ordre chronologique.

Exemple de Tweet :



VI. Les différentes méthodes d'extraction de données

Il existe plusieurs méthodes d'extraction de données nous allons en voir quelques-unes :

Qu'est-ce que le Data Scrapping ?

Le data scrapping (ou capture, collecte ou extraction de données en français) est une technique par laquelle un programme informatique extrait des données depuis une source lisible par un être humain et produite par un autre programme informatique.

Le data scraping permet d'extraire des données et de les structurer depuis une source prévue à l'origine pour être lue par un humain et où l'information n'est donc pas structurée, pas documentée, et pas optimisée pour être extraite facilement. En général, le data scraping est utilisé pour interfacer un programme avec un programme plus ancien qui ne propose pas d'API ou pour extraire des données depuis une source tierce, parfois sans son accord.

Qu'est-ce que le Data Extractor ?

Les données nécessaires pour une analyse sont présentes sur Internet, ces données contiennent une grande quantité d'informations et il n'est pas possible pour un analyste par exemple de prendre manuellement ces données pour effectuer une extraction de données. Les données dans les sites web ne sont pas structurées, ce qui signifie qu'elles peuvent être des données indésirables.

Pour extraire des données volumineuses et non structurées depuis internet, il existe des techniques et des outils d'extraction de données qui peuvent Extraire facilement de données en grande quantité et non structurées et les convertir en un format significatif et structuré.

Le data extractor est un outil pour extraire les données. Le processus commence par l'indexation. Il explore les données des sites Web en faisant du data crawling. Après avoir creusé en profondeur les sites web, il nous renvoie les liens des sites disponibles pour nous. Dans ces pages Web, beaucoup de données indésirables et utiles ont été mixées, nous devons donc extraire les données utiles et les convertir dans un format souhaité. L'extracteur de données extrait les données nécessaires dans les sites Web qu'il a explorés afin de les séparer et utiliser les informations qu'on désire.

VII. Définition d'une API

Une API, ou interface de programmation d'application, est un ensemble de définitions et de protocoles qui facilitent la création et l'intégration de logiciels d'applications. C'est comme un regroupement de fonctionnalités que l'on peut facilement intégrer à notre application.

À quoi sert une API ?

Une API permet d'échanger des données ou des fonctionnalités à l'intérieur ou à l'extérieur des entreprises. On peut voir une API comme une porte d'entrée de données.

Par exemple, pour notre cas dans une application web. Grâce à l'API Twitter on a pu intégrer toutes les fonctionnalités de celle-ci (les tweets, la timeline, les retweets etc.). L'API va se connecter à la base de données où se trouvent ces informations pour les afficher dans notre application.

Les API procurent ainsi de nombreux avantages, que ce soit côté utilisateurs ou côté fournisseurs.

On peut donc intégrer les fonctionnalités de l'API très aisément dans son application ou son site web.



VIII. La stratégie et la méthode adopté

Pour commencer, nous avons choisi d'extraire les données à l'aide de l'API Twitter, comme dans l'article. Car l'API est bien fournie et documentée, elle nous donne accès à la plupart des informations disponibles sur le site de Twitter. Ensuite, il nous a fallu créer un côté Backend car nous avons remarqué qu'une communication directe entre le côté Frontend et l'api n'était pas possible du fait des erreurs de CORS.

Qu'est-ce que le CORS ?

Le Cross-Origin Response Sharing (CORS) est un mécanisme de navigateur qui permet un accès contrôlé aux ressources situées en dehors d'un domaine donné. Cependant, il offre également un potentiel pour les attaques inter-domaines.

Pourquoi bloquer le CORS ?

Le blocage du CORS permet d'empêcher la falsification des requêtes intersites. Imaginons que vous vous connectiez à twitter.com et que votre navigateur stocke le jeton d'authentification afin que vous soyez connecté automatiquement à l'avenir. Et pour chaque requête vers l'origine twitter.com, ces en-têtes auth-token seront présents. Imaginez maintenant un scénario dans lequel vous cliquez sur un pop-up ouvrant un site malveillant. Ce site aura également la capacité d'effectuer des requêtes vers twitter.com. Dans ce cas, le site malveillant pourrait envoyer des requêtes en votre nom et vous pirater. Pour éviter cela, l'erreur CORS a été introduite.

Le Backend

La partie Backend est l'endroit où va se jouer le plus gros du projet, c'est la partie invisible d'un site Web, toute l'algorithmie, la technique va se jouer dans cette partie. Dans ce Backend nous allons y mettre toutes les requêtes vers l'api Twitter. Nous allons aussi y mettre les requêtes de stockage vers la base de données et y définir les différentes routes que notre côté Frontend va récupérer.

L'Api Twitter

L'api Twitter va donc nous donner une clef privée qui va nous permettre à nous développeurs d'envoyer des requêtes en notre nom et récupérer les informations disponibles sur l'api.

Le Frontend

Le Frontend quant à lui va s'occuper d'afficher correctement les différents tweets, étant donné que l'on veut faire un site disponible au grand public, il se doit d'être esthétique, ergonomique et facile à prendre en main.

IX. Les outils utilisés

NodeJS :



NodeJs est un environnement d'exécution JavaScript à code source ouvert et multiplateforme. C'est un outil populaire pour presque tous les types de projets.

NodeJs fait tourner le moteur JavaScript V8, le noyau de Google Chrome, en dehors du navigateur. Cela permet à NodeJs d'être très performant.

Une application NodeJs fonctionne dans un seul processus, sans créer un nouveau thread pour chaque requête. NodeJs fournit un ensemble de primitives asynchrones dans sa bibliothèque standard qui empêche le code JavaScript de se bloquer et, en général, les bibliothèques de NodeJs sont écrites en utilisant des paradigmes non-bloquants, ce qui fait du comportement bloquant l'exception plutôt que la norme.

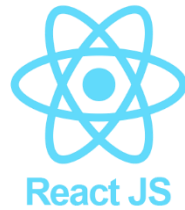
Lorsque NodeJs effectue une opération, comme la lecture du réseau, l'accès à une base de données ou au système de fichiers, au lieu de bloquer le thread et de gaspiller des cycles de CPU en attendant, NodeJs reprendra les opérations lorsque la réponse reviendra.

Cela permet à NodeJs de gérer des milliers de connexions simultanées avec un seul serveur sans introduire le fardeau de la gestion de la simultanéité des threads, qui pourrait être une source importante de bogues.

NodeJs a un avantage unique car des millions de développeurs frontaux qui écrivent du JavaScript pour le navigateur sont maintenant capables d'écrire le code côté serveur en plus du code côté client sans avoir besoin d'apprendre un langage complètement différent.

Dans NodeJs, les nouvelles normes ECMAScript peuvent être utilisées sans problème, car on n'a pas besoin d'attendre que tous vos utilisateurs mettent à jour leurs navigateurs - on est en charge de décider quelle version ECMAScript utiliser en changeant la version de NodeJs, et vous pouvez également activer des caractéristiques expérimentales spécifiques en exécutant NodeJs avec des drapeaux.

React :



Pour notre projet, nous avons donc choisi d'utiliser React pour le côté Front. React est une bibliothèque de développement d'interface utilisateur basée sur JavaScript. Facebook et une communauté de développeurs open-source l'exploitent. Bien que React soit une bibliothèque plutôt qu'un langage, elle est largement utilisée dans le développement web. La bibliothèque est apparue pour la première fois en mai 2013 et est maintenant l'une des bibliothèques Front les plus utilisées pour le développement web.

React propose diverses extensions pour une prise en charge architecturale complète des applications, comme Flux et React Native, au-delà de la simple interface utilisateur. La popularité de React a aujourd'hui éclipsé celle de tous les autres frameworks de développement Front. Voici pourquoi :

- Création facile d'applications dynamiques : React facilite la création d'applications web dynamiques car il nécessite moins de codage et offre plus de fonctionnalités, contrairement à JavaScript, où le codage devient souvent très vite complexe.
- Amélioration des performances : React utilise le DOM virtuel, créant ainsi des applications web plus rapidement. Virtual DOM compare les états précédents des composants et ne met à jour que les éléments du Real DOM qui ont été modifiés, au lieu de remettre à jour tous les composants, comme le font les applications web classiques.
- Des outils dédiés pour un débogage facile : Facebook a publié une extension Chrome qui peut être utilisée pour déboguer les applications React. Cela rend le processus de débogage des applications web React plus rapide et plus facile.
- Composants réutilisables : Les composants sont les éléments constitutifs de toute application React, et une seule application se compose généralement de plusieurs composants. Ces composants ont leur logique et leurs contrôles, et ils peuvent être réutilisés dans toute l'application, ce qui réduit considérablement le temps de développement de l'application.
- Flux de données unidirectionnel : React suit un flux de données unidirectionnel. Cela signifie que lors de la conception d'une application React, les

développeurs imbriquent souvent les composants enfants dans les composants parents. Comme les données circulent dans une seule direction, il devient plus facile de déboguer les erreurs et de savoir où se situe un problème dans une application au moment considéré.

- Courbe d'apprentissage réduite : React est facile à apprendre, car il combine principalement des concepts HTML et JavaScript de base avec quelques ajouts bénéfiques. Néanmoins, comme c'est le cas avec d'autres outils et frameworks, vous devez passer un peu de temps pour bien comprendre la bibliothèque de React.
- Il peut être utilisé pour le développement d'applications web et mobiles : Nous savons déjà que React est utilisé pour le développement d'applications web, mais ce n'est pas tout ce qu'il peut faire. Il existe un framework appelé React Native, dérivé de React lui-même, qui est extrêmement populaire et qui est utilisé pour créer de magnifiques applications mobiles. Donc, en réalité, React peut être utilisé pour faire à la fois des applications web et mobiles.

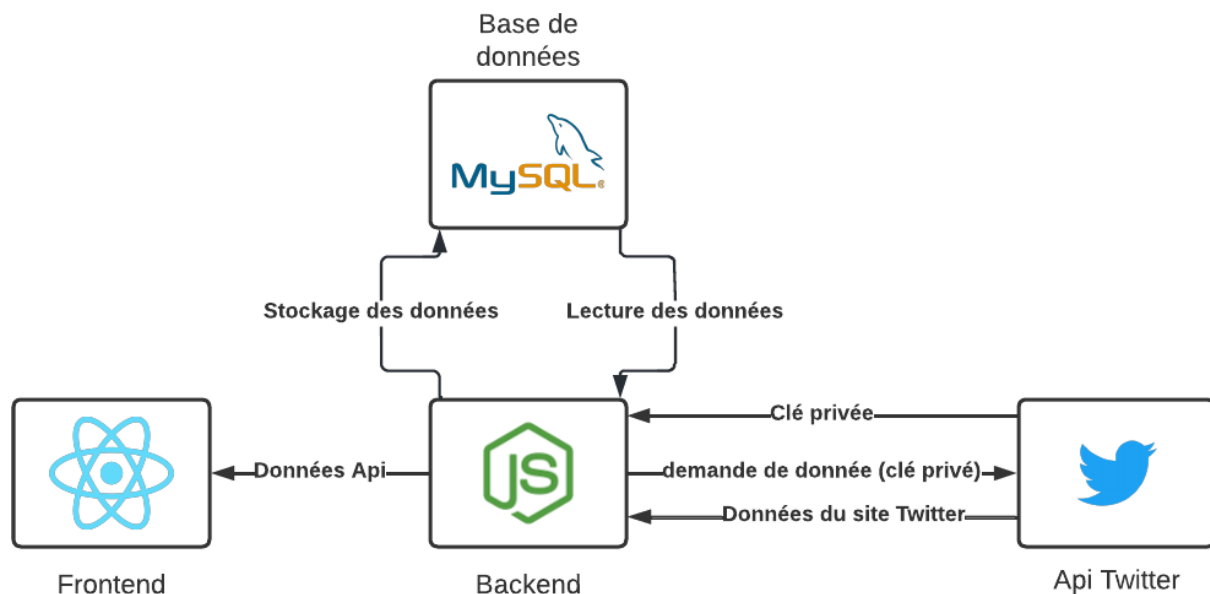
MySQL :



MySQL est la base de données open source la plus populaire au monde. Selon DB-Engines, MySQL se classe au deuxième rang des bases de données les plus populaires, derrière Oracle Database. MySQL alimente de nombreuses applications parmi les plus utilisées, notamment Facebook, Twitter, Netflix, Uber, Airbnb, Shopify et Booking.com.

MySQL étant open source, il comprend de nombreuses fonctionnalités développées en étroite collaboration avec les utilisateurs depuis plus de 25 ans. Il est donc très implanté dans le milieu informatique, il a une grande maturité et de nombreux outils permettant son utilisation.

X. Architecture du projet

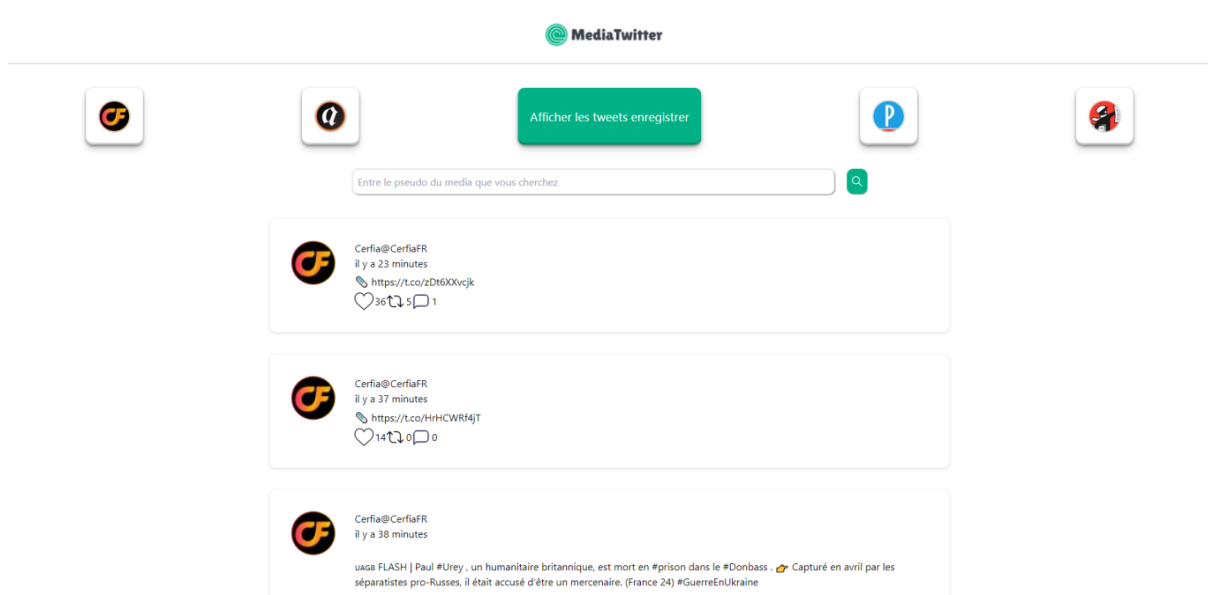


Pour l'architecture du projet, nous avons décidé d'utiliser React pour le côté Front grâce à sa facilité d'apprentissage et aussi sa capacité à pouvoir réutiliser des composants et pour la gestion de nos tweets. Ensuite pour le Backend, on a décidé d'utiliser NodeJS pour notre Api Back afin d'avoir un projet entièrement en JavaScript et aussi grâce à sa forte documentation sur internet. Pour finir, nous avons choisi d'utiliser MySQL car nous maîtrisons déjà cette technologie grâce à nos cours à l'Université.

Le projet va avoir un Front qui va discuter avec notre Api Backend qui elle va discuter avec l'Api Twitter et va traiter les données, les stocker dans la base de données puis les renvoyer au Frontend.

XI. Présentation de l'application

Premièrement sur l'application, nous allons avoir une image de présentation du site pour le header. Ensuite, nous allons avoir 4 boutons raccourcis de média disponibles. En effet, les liens qui sont présent dans les tweets sont automatiquement redus cliquable.



Les raccourcis vont permettre d'afficher les tweets du média, de sélectionner en cliquant dessus comme on peut le voir sur l'image ci-dessus.

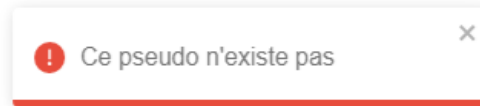
Ce système fonctionne de la façon suivante :

- On clique sur le bouton (Le bouton possède deux états Ouvert/Fermé comme un interrupteur)
- Le bouton va lancer une requête vers l'api en indiquant le nom du bouton
- Le serveur va lancer deux requêtes vers l'api Twitter, l'une pour récupérer la photo de profil ainsi que le pseudonyme, et l'autre va récupérer les derniers tweets en utilisant le pseudonyme récupérer dans la première requête.
- Toutes ces données vont être traité, assembler puis stocker en base de données sans insérer de doublons et envoyer au Front.
- Si l'on reclique cela ferme



Deuxièmement, on va avoir une barre de recherche disponible pour pouvoir rechercher les tweets d'un média qui n'est pas disponible dans nos raccourcis. Il faut savoir qu'un pseudo Twitter ne peut dépasser 15 caractères, alors un message apparaîtra dès lors que nous dépassons ce cota.

Un pseudo twitter ne peut dépasser 15 caractères


De plus, si le pseudo n'existe pas, nous avons fait sorte que l'api renvoie une erreur 402 vers le Front et va donc afficher une bulle d'erreur.



Troisièmement, on a la possibilité de cliquer sur le bouton "Afficher les tweets enregistrer" qui va afficher la totalité des tweets enregistrer dans la base de données en les triant par date de créations.







Afficher les tweets enregistrer



le_Parisien@le_Parisien
il y a 4 minutes




Ségolène Royal reconnaît s'être «trompée» sur le bombardement de la maternité de Marioupol en #Ukraine ➡
<https://t.co/gYs4x0HQJA> <https://t.co/L2YaLCdTII>

 1  0  0



BFMTV@BFMTV
il y a 9 minutes

Emmanuel Macron: "Le choix le plus rationnel était de confirmer la fermeture" de la centrale de Fessenheim
<https://t.co/9qNyNu7V52>

 7  8  21

Rapport de projet innovant – Université Paris 8

21

Quatrièmement Afin de ne pas avoir des centaines de tweets sur une même page, nous avons mis en place un système de pagination.



Pour finir sur la présentation de l'application, nous avons mis en place un petit formulaire en bas de page permettant de nous faire des propositions de raccourcis.

Des Medias à nous proposer ?

Bonjour, Pouvez vous rajouter un raccourci pour BFMTV, merci !

Envoyer

Ces propositions vont nous être automatiquement envoyé par mail.



A. Démonstration du Projet

Lien de la vidéo : <https://youtu.be/WUqlmnZ78mk>

XII. Conclusion

Pour conclure, Tout au long de ce projet, on a développé des compétences tant techniques que transversales et pu découvrir de nouvelles notions comme le fonctionnement d'une API, les différentes méthodes d'extraction de données, le web scrapping, le web crawler etc...

Ce projet nous a permis d'apprendre énormément de compétences les plus importantes sur l'utilisation d'un API mais aussi de comprendre l'efficacité d'un Framework monopage.

Nous avons créé un programme permettant d'afficher correctement les derniers tweets de médias comme « CerfiaFR » et « MediaAvenir » en utilisant un Backend qui va récupérer les informations à partir de l'api Twitter. De plus, nous avons mis à disposition un formulaire de propositions de médias qui va nous envoyer un mail dès lors que l'on va recevoir une proposition. L'application web comporte également un système de pagination ainsi qu'un bouton qui permet de lire les tweets stocker localement dans la base de données.

Nous avons également mis à disposition des boutons correspondants à certains médias. Ainsi, il sera possible de cliquer sur un bouton correspondant au média de son choix afin que tous ses tweets apparaissent. Toutes ces informations seront ensuite stockées dans une base de données. Nous avons envisagé des possibilités futures comme un système de connexion via un compte pour s'abonner aux médias que l'on souhaite ainsi qu'une barre de favoris pour un accès plus rapide.

Pour finir, ce projet a été très enrichissant dans tous les domaines, on a pu découvrir et acquérir énormément de compétences qui nous seront très utiles dans le monde du travail et même sur le plan personnel. On voulait, en fin de ce rapport, remercier nos deux tuteurs Madame LYNDA SEDDIKI et M. NASREDDINE BOUHAÏ pour leur investissement et leur accompagnement très complet. Cela nous a permis d'atteindre nos espérances concernant l'aptitude à mener à bien le projet et à réaliser l'ensemble des fonctionnalités souhaitées.

XIII. Références

Article 1 Extraction de données commerciales à partir des réseaux sociaux :

<https://scihub.mkksa.top/https://ieeexplore.ieee.org/document/7507976>

- Article 2 Système d'évaluation de la crédibilité des Tweets en temps réel :

<https://sci-hub.mkksa.top/10.1109/asonam.2016.7752431>

- Article 3 Extraction d'information du réseau social pour la commercialisation de produits agricole :

<https://sci-hub.mkksa.top/10.1109/csnt.2012.200>

- React 18.0.0.0:

<https://reactjs.org/blog/2022/03/29/react-v18.html>

- React-moment: 1.1.2:

<https://www.npmjs.com/package/react-moment>

- React-toastify: 8.2.0:

<https://www.npmjs.com/package/react-toastify/v/8.2.0>

- React-linkify: 1.0.0:

<https://www.npmjs.com/package/react-linkify/v/1.0.0-alpha>

- React-paginate: 8.1.3:

<https://www.npmjs.com/package/react-paginate>

- React-hook-form: 7.30.0:

<https://www.npmjs.com/package/react-hook-form>

- Wikipedia Web scrapping:

https://fr.wikipedia.org/wiki/Web_scrapping

- Web crawling:

<https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203499-crawler-definition-traduction-et-acteurs/>

Doc CORS :

<https://developer.mozilla.org/fr/docs/Web/HTTP/CORS>

- ReactJS:

<https://fr.reactjs.org/>

- NodeJS:

<https://nodejs.org/en/docs/>

- Démonstration du projet :

<https://youtu.be/WUqlmnZ78mk>