

République Algérienne Démocratique Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université d'Ibn Khaldoun – Tiaret

Faculté des Mathématiques et de l'Informatique

Département Informatique

Thème

Clustering des tweets pour la recommandation

Pour l'obtention du diplôme de Master II

Spécialité : Réseaux et télécom

Réalisé par : TADJ Imane

BELEHZIEL Nouria

Dirigé par : MOKHTARI Ahmed

Année universitaire 2014-2015

Dédicaces

L'eau coule grâce à sa source

L'arbre pousse grâce à ses racines

A

Nos parents

*Pour les sacrifices déployés à nos égards ; pour leur patience
Leur amour et leur confiance en nous*

*Ils ont tout fait pour notre bonheur et notre réussite.
Qu'ils trouvent dans ce modeste travail, le témoignage de notre
Profonde affection et de notre attachement indéfectible.
Nulle dédicace ne puisse exprimer ce que nous leur devons
Que dieu leur réserve la bonne santé et une longue vie.*

A

Nos amis

*En témoignage de nos sincères reconnaissances pour les efforts
Qu'ils ont consentis pour nous soutenir au cours de nos études.
Que dieu nous garde toujours unis*

A

Toute personne qui nous 'a aidé à faire notre projet.

Remerciements

*Au terme de la rédaction de ce mémoire, nous tenons à remercier
notre
encadreur Mr : MOKHTARI Ahmed pour ses
précieus conseils et son aide durant toute la période du travail.*

*Nos vifs remerciements vont également aux membres du jury pour
l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner
notre
travail Et de l'enrichir par leurs propositions.*

Merci

Résumé

Depuis quelques années, les personnes n'hésitent plus à prendre contact sur les réseaux sociaux de type Facebook, Viadeo, LinkedIn ou même Twitter.

Les réseaux sociaux sont des plateformes de service web qui permettent à tout un chacun de créer son profil en présentant les informations liées à son activité personnelles ou professionnelle, son parcours et sa situation (dans notre cas Twitter).

Le principe de ces réseaux est de mettre en relation les personnes qui ont un intérêt commun.

Ainsi, la mise en relation s'effectue soit par une action volontaire de la personne présente sur le réseau social (c'est-à-dire que la personne demande à faire partie des contacts d'une autre personne) soit par suggestion du réseau lui-même par recommandation des tweets (le cas de Twitter) ou des amis (par exemple Facebook).

Dans ce mémoire nous avons étudié cette recommandation des tweets et/ou des amis à suivre sur Twitter en utilisant une classification des contenus des tweets.

Mots-clés : Le web social, Les réseaux sociaux, Twitter, tweets, classification, recommandation ...

Sommaire

Introduction Générale..... 1

Chapitre1 : Le web social et les réseaux sociaux

Introduction 2

I. Le web social 2

1. L'évolution du web social 3

2. Des projets de logiciels communautaire..... 4

3. Du web social à la vie réelle..... 4

II. Les réseaux sociaux 5

1. L'historique 5

2. L'objectifs des réseaux sociaux 6

3. Les principaux réseaux sociaux : description et fonction..... 6

3.1. Facebook : 6

III. Twitter 7

1. Fonctionnalités..... 8

1.1. Le tweet et le retweet 8

1.2. Abonnements 8

1.3. Fil d'actualité..... 8

1.4. Mentions 9

1.5. Hashtags – mot clic – mot dièse 9

1.6. Trending Topic..... 9

2. Twitter comme outil de connaissances	9
3. Twitter comme objet d'étude ou support d'études	10
4. Limitation de l'accès aux anciennes publications	10
Conclusion.....	10

Chapitre2 : la classification (Clustering) et la recommandation

Introduction	11
I. La classification	11
1. Applications de la classification.....	11
2. Etapes de la classification.....	11
2.1. Préprocessing	12
2.1.1. Extraction de données	12
2.2. Calcul de la distance de similarité sémantique	13
2.3. Classification.....	13
3. Les Techniques de classification automatique	14
3.1. Apprentissage non supervisé.....	14
3.2. Apprentissage supervisé.....	15
II. La recommandation	16
1. Les systèmes traditionnels de recommandation	16
2. Le filtrage collaboratif et la recommandation basée sur le contenu	17
3. Autres approches dans les médias sociaux	18
3.1. Recommandation des (Follower).....	18
3.2. Recommandation des hashtags	18

3.3. Recommandation des tweets	18
Conclusion.....	19

Chapitre3 : la modélisation

Introduction	20
I. La description de notre système :.....	20
1. Téléchargement :	21
1. 2 Elimination de stop words :	21
2. Le Clustering :	21
2. 1 L'algorithme de Clustering :.....	21
2. 2 La procédure de distance de similarité sémantique	23
2. 3 L'explication de l'algorithme de Clustering	23
2. 4 L'explication de la procédure de similarité sémantique	24
3. La recommandation :.....	24
3. 1 L'algorithme de recommandation :.....	25
3. 2 L'explication de l'algorithme de recommandation :.....	25
Conclusion :.....	28

Chapitre4 :L'implémentation

Introduction	29
I. les outils de programmation.....	29
1. Qu'est-ce que Java ?	29

➤ Java et l'api.....	29
2. Netbeans 8	30
3. Api Twitter for java	30
II. Description de l'application	32
1. Page d'accueil.....	32
2. La barre de menu	32
2.1. Le bouton fichier.....	32
2.2. Le bouton connexion.....	33
2.3. Le bouton téléchargement.....	33
2.4. Le bouton traitement	33
3. La fenêtre de connexion	34
4. La fenêtre de chargement	34
3.1. Les résultats de chargement	35
5. La fenêtre de Clustering.....	37
5.1. Les résultats de Clustering	37
6. La fenêtre de recommandation	40
Conclusion.....	42
Conclusion générale	43
Bibliographie.....	44
Webographie	45
Liste du figures.....	46

Liste des tableaux 47

Introduction Générale

Avec de l'évolution l'informatique dans le monde, ces dernières années, est apparue la nécessité d'informatisation et de la communication, afin d'assurer une bonne qualité de service dans la société.

Il est clair que la communication joue un rôle prépondérant au sein des sociétés, leurs objets doivent être de plus en plus performants, car elles sont soumises virulente concurrence.

Ce mémoire est constitué de quatre(4) chapitres (partie).

Le 1^{er} chapitre :

Dans le premier chapitre intitulé «Le web social et les réseaux sociaux» nous définissons quelques concepts jugés nécessaire sur le web social, on a déterminé les différentes réseaux sociaux et ses fonctionnalités principale on a ensuite donnée une vue spécial à notre thème de notre étude (le Twitter)

Le 2^{ème} chapitre :

Au niveau de deuxième chapitre intitulé «*la classification (Clustering) et la recommandation* » nous définissons les notions de bases concernant la classification et la recommandation dans le Twitter.

Le 3^{ème} chapitre :

Au niveau de troisième chapitre intitulé «*La modélisation* » nous avant modéliser notre travaille les phases de notre système et quelques algorithmes et procédures.

Le 4^{ème} chapitre :

Finalement dans le dernier chapitre qu'on a nommé « Implémentation » nous présentons les outils de développement qui nous ont servi pour la réalisation de notre projet et quelque capture d'image concernant notre application.

Chapitre1 : Le web social et les réseaux sociaux

Introduction

Le sujet que nous avons décidé d'étudier porte sur le web social et en particulier les réseaux sociaux.

Nous avons choisi d'analyser ce sujet assez vaste car c'est une sorte de monde virtuel extrêmement présent en ce moment, principalement pour notre génération, mais aussi chez les personnes plus âgées.

Ce sujet nous intéresse beaucoup car nous aimerions en savoir plus concernant ces sites internet.

I. Le web social

Le web social fait référence à une vision d'Internet considéré comme un espace de socialisation, un lieu dont l'une de ses fonctions principales est de faire interagir les utilisateurs entre eux afin d'assurer une production continue de contenu, et non plus uniquement la distribution de documents.

Il est considéré comme un aspect très important du Web 2.0. En particulier, il est associé à différents systèmes sociaux tels que le réseautage social, les blogs ou les wikis. [1]



Figure 1 : Le web social de 2000 à 2010

1. L'évolution du web social

Le web social va rapidement devenir une manière de penser : beaucoup de personnes vont sur les réseaux sociaux au moins une fois par jour, et le temps de visite moyen de Myspace en 2008 était autour de 36 minutes (la longueur d'une série télévisée).

La frontière entre les réseaux sociaux et les médias sociaux est de plus en plus floue avec des sites comme Facebook et Twitter incorporent davantage de photos, vidéos, et d'autres fonctionnalités typique des médias sociaux sur les profils publiques des utilisateurs, comme les médias sociaux ont intégré plusieurs caractéristiques des réseaux sociaux dans leur propre interface. Un changement notable est apparu avec la fusion des réseaux et médias sociaux, qui transforme les applications du web social en un écosystème centré sur l'utilisateur.

Bien qu'il y ait eu de nombreuses discussions autour du sens des communautés sur le web avant ces innovations, les logiciels du web social moderne offrent un plus large choix d'interactions sociales, comme « devenir ami » ou « suivre » une personne, ou même envoyer des bisous et des cadeaux virtuels. Les applications du web social sont généralement construites avec de la programmation orientée objet, utilisant des combinaisons de plusieurs

langages de programmation, comme Ruby, PHP, Python, et/ou Java. Des API sont souvent utilisées pour attacher des sites Internet non-sociaux à des sites sociaux. [1]

2. Des projets de logiciels communautaire

Grâce à l'utilisation du Web social, nombre de développeurs de logiciels ont choisi de participer à des projets communautaires des logiciels libres. Les évolutions du système d'exploitation GNU/Linux sont des parfaits exemples de l'efficacité et de l'efficience de ce genre de collaboration. [2]

3. Du web social à la vie réelle

Le web social a changé la manière dont les individus peuvent partager des informations personnelles, se rencontrer, faire leurs courses ou rechercher du travail. On partage désormais ses préférences, ses opinions et ses activités avec des groupes d'amis que l'on n'aurait parfois jamais rencontrés sans le web social.

Un bon nombre de sites utilisent l'interaction sociale en ligne en vue d'une interaction dans la vie réelle. Des relations interindividuelles sont formées via Internet et peuvent devenir plus personnelles qu'à travers les autres formes de communication.

Le web social facilite les mises en relation dans la vie réelle : les rencontres sur Internet peuvent être effectuées par le biais des sites de rencontres ou ailleurs, de manière moins formelle : certains sites permettent de trouver des partenaires pour faire des sorties (*OVS*), faire du sport (*RencontreSportive.com*, *WidiWici*) ou pour organiser des tournois entre gamers (*GamersBand*).

D'autres sites offrent la possibilité d'organiser des rencontres dans un but d'apprentissage linguistique (en) (*LanguageForExchange*,...). Quelques sites sont également spécialisés dans les échanges de services avec son voisinage.

D'autres permettent d'organiser des rencontres thématiques où les inscrits échangent des réflexions ou des expériences sur un sujet donné (*Bar camp*, *First Tuesday* (en), *Meetup.com*). Le site *DreamShake* se propose quant à lui de réunir des personnes dans l'aide à la réalisation de projets personnels. [2]

II. Les réseaux sociaux

Le terme désigne un site internet permettant à l'internaute de s'inscrire et d'y créer une carte d'identité virtuelle appelée le plus souvent « profil ». Le réseau est dit social en ce qu'il permet d'échanger avec les autres membres inscrits sur le même réseau : des messages publics ou privés, des liens hypertexte, des vidéos, des photos, des jeux... L'ingrédient fondamental du réseau social reste cependant la possibilité d'ajouter des « amis », et de gérer ainsi une liste de contacts. [3]

L'émulation des réseaux sociaux fonctionne ensuite sur deux principes que l'on peut résumer ainsi :

1. « Les amis de mes amis sont mes amis ».
2. « Les personnes qui partagent les mêmes centres d'intérêts que moi sont mes amis ».

1. L'historique

L'étymologie du mot "réseau" provient de "rete" en latin qui signifie "filet". Un réseau est composé par définition de liens interconnectés qui s'entrecroisent pour former un filet, un amas de nœuds reliés entre eux. Un réseau "social" est constitué de regroupements d'individus, de clubs ou d'organisations liés par des interactions sociales. Le terme désigne aujourd'hui les sites Internet communautaires qui regroupent des profils d'individus interconnectés.

Les différents réseaux sociaux sont spécialisés dans des domaines spécifiques. Facebook, le plus populaire actuellement, se base principalement sur les relations amicales, les amis perdus de vue ou encore les relations intimes. Myspace est un outil plus particulièrement utilisé par les musiciens et les artistes. LinkedIn est un réseau essentiellement professionnel, alors que Flickr est un outil prévu pour publier des photographies.

Il n'est pas rare de retrouver une même personne simultanément sur chacun d'eux, voire d'autres encore. [3]



Figure 2 : *Les réseaux sociaux*

2. L'objectifs des réseaux sociaux

Les réseaux sociaux ont pour but de créer des interactions sociales entre les utilisateurs : c'est ce qui explique leur succès. Une majorité d'utilisateurs s'inscrivent d'abord pour voir les photos d'amis perdus de vue ou pour visionner les photos dans lesquelles ils ont été "tagués" ou marqués, c'est-à dire révélés (leur nom est inscrit en dessous de la photo).

La curiosité étant un vilain défaut propre à la nature humaine, le nombre de personnes connectées sur ces réseaux augmente de façon exponentielle. [3]

3. Les principaux réseaux sociaux : description et fonction

3.1. Facebook :

C'est le réseau social le plus connu : plus de 500 000 000 inscrits. Un film a même été réalisé sur son créateur : Mark Zuckerberg.

Le principe est d'échanger avec sa communauté d'amis sur tout et n'importe quoi. L'inscription est obligatoire pour l'utiliser. Une fois cette démarche faite vous pouvez commencer à écrire, mettre photo et vidéo sur votre « mur », « wall » pour les anglophones.

Pour être amis sur Facebook avec une personne, il faut lui envoyer une demande et que cette dernière l'accepte. Facebook permet également de réagir sur les commentaires et news postés par ses amis via le « Like » ou J'aime.

C'est un moyen pour dire que l'on a trouvé un commentaire ou un post à son goût. Il est devenu fréquent d'entendre le verbe « liker » dans une conversation.

Vous pouvez d'ailleurs aimer cet article grâce au bouton situé en haut de l'article. Facebook permet beaucoup d'autre chose : discussion instantanée, envoi de message direct, identifier des amis sur une photo. [4]

III. Twitter

C'est une plateforme de microblogging. Comme Facebook, Twitter permet de partager avec d'autres.

Le fonctionnement est toutefois différent de Facebook : une limitation à 140 caractères par message, la possibilité de suivre d'autres comptes, pas de demande d'invitation, vous suivez qui vous voulez et vous suit qui veut, le partage de photo, de vidéo ou d'article se fait par l'utilisation de lien. En gros, vous publiez des SMS avec des liens.

Sur Twitter, contrairement à Facebook, vous n'avez pas un mur mais une timeline ou fil. Alors que Facebook est destiné à échanger avec ses amis, Twitter, lui, possède un côté plus pro et relationnel. La notion d'influence y est très présente de par sa logique de suivi d'informations données par d'autres.

Avec Twitter, vous lisez les tweets de personnes que vous suivez et pouvez les Retweeter pour les partager avec vos followers. Le temps réel + le retweet en fait un outil de propagation d'information mais aussi de rumeur très puissant. Twitter est en grande partie la raison du développement des URLS raccourcie, ex : <http://bit.ly/h6aB1a> est l'url raccourcie pour ce billet. Vous pouvez me retrouver sur Twitter. [5]

1. Fonctionnalités

1.1. Le tweet et le retweet

Chaque court message, le tweet, publié est limité à 140 caractères il contraint les utilisateurs à être concis dans leur rédaction. Initialement, Twitter pouvait être utilisé par l'intermédiaire des SMS. Ceux-ci étant limités à 160 caractères, Twitter prend cette limite et conserve 20 caractères pour ajouter son nom d'utilisateur. La capacité limitée à 140 caractères par message sur le service a favorisé l'émergence de plateformes de contenu, telle TwitPic, qui permet de poster des images et photos ; bit.ly pour raccourcir les liens.

Il existe deux manières de « retweeter » : soit comme initialement en copiant/collant intégralement le tweet lu en le précédant de la mention « RT @Bob », soit comme depuis fin 2009 en le « retweetant » automatiquement pour l'afficher à ses abonnés tel qu'on l'a vu soi-même, avec l'avatar de l'auteur d'origine, sauf si l'auteur d'origine utilise un compte protégé. [6]

1.2. Abonnements

Après s'être connecté à Twitter en tant que membre inscrit, on accède aux tweets (mini-messages) en fil postés par ses propres abonnements, c'est-à-dire par les comptes d'utilisateurs que l'on a choisi de « suivre ». Si l'utilisateur Alice « suit » l'utilisateur Bob, on dit qu'Alice est une abonnée de Bob et que Bob est un abonnement d'Alice.

Dans la version française de l'interface, un follower est appelé initialement « suiveur » remplacé par « abonné » et un following, par « suivis » puis « abonnement ». Twitter est un réseau social asymétrique, c'est-à-dire n'engageant pas de réciprocité.

Il est possible pour un utilisateur de restreindre la lecture de ses mini-messages en gardant privé l'accès à son compte, en évitant donc de le rendre public. Les messages sont alors visibles par l'abonné uniquement après validation d'une requête d'ajout à sa liste d'abonnement par l'utilisateur qui a appliqué un accès privé. [7]

1.3. Fil d'actualité

Le fil d'actualité, ou « timeline » en anglais (abrégé « TL »), est la page principale sur laquelle apparaissent les tweets des comptes auxquels l'utilisateur s'est abonné.

1.4. Mentions

Un nom précédé d'arobase « @ » est un lien vers le compte Twitter de l'utilisateur de ce nom (qui permet de voir tous ses tweets, sauf s'ils sont protégés). Chaque utilisateur peut consulter les mentions qu'il a reçues dans l'onglet « @ Connect ». Si un tweet débute par une mention, seuls les followers suivant le compte mentionné verront le tweet dans leur fil d'actualité (par exemple @Eve rédige un tweet en commençant par @Bob, donc parmi les followers de @Eve, seuls ceux qui suivent également @Bob liront le tweet depuis leur fil d'actualité). [7]

1.5. Hashtags – mot clic – mot dièse

Un mot précédé du signe « # » (croisillon) est un hashtag. Au Canada en janvier 2011 l'Office québécois de la langue française a créé et proposé le terme « mot-clic ». En France, la commission générale de terminologie et de néologie a proposé en janvier 2013 le terme « mot-dièse ». Il s'agit d'un sujet attribué au message, Twitter peut afficher tous les tweets comportant un hashtag précis, et établit un classement des mots ou bien des hashtags du moment les plus utilisés (les trending topics, désormais disponibles pour les tweets rédigés en français.) — d'où parfois des détournements du système, sur le même principe que le bombardement Google.

En pratique Twitter a un vocabulaire propre, les utilisateurs du service de microblogage créent de nombreux hashtags sous forme d'abréviations. [7]

1.6. Trending Topic

Les trending topics, abrégés « TT » sur Twitter, sont les sujets tendances. Ce sont des mots, des hashtags ou des phrases qui ont été tweetés de multiples fois durant une période. Il est possible d'afficher les tendances par pays, par ville ou encore dans le monde entier. [8]

2. Twitter comme outil de connaissances

Les fonctionnalités de Twitter permettent à ses utilisateurs d'obtenir, filtrer, trier, échanger et classer le flot d'un grand nombre d'informations et de données. En ce sens que Twitter peut être assimilé à un outil d'acquisition et gestion des connaissances. [8]

3. Twitter comme objet d'étude ou support d'études

La base de données de Twitter est un substrat statistique et sémantique dont les scientifiques ou des entreprises commerciales ou de communication peuvent tirer des informations.

Twitter, ses hashtags et smileys et ses utilisateurs deviennent donc des sujets d'étude et de recherche. [9]

4. Limitation de l'accès aux anciennes publications

Twitter limite la consultation aux seules publications récentes, il devient impossible d'accéder à certaines publications anciennes notamment par le moteur de recherche interne du service.

Cependant, le service a intégré depuis le mois de mars 2013, la possibilité de télécharger une archive contenant l'ensemble des publications d'un compte utilisateur. [10]

Conclusion

Grâce à nos recherches sur le web social et les réseaux sociaux nous avons beaucoup appris de ceux-ci, notamment sur leurs origines, leur importance et dans le chapitre suivant en va voire les méthodes qui nous idée et facilité l'utilisation du ces réseaux.

Et dans le chapitre suivant nous allons défini les notions de bases de la classification et la recommandation.

Chapitre2 : la classification (Clustering) et la recommandation

Introduction

Le but de notre projet est de réaliser une application qui fait la classification de tweets et après la recommandation .donc dans ce chapitre on va définir ces notions (classification et recommandation).

I. La classification

La classification automatique de documents est un problème connu en informatique, il s'agit d'assigner un document a une plusieurs catégories ou classes. Le problème est différent selon la nature des documents en question, en effet la classification de textes diffère de la classification de documents images, vidéo ou encore son.

On peut aussi imaginer des classifications selon des paramètres associés aux documents tels que par exemple l'auteur, la date de parution [11]

1. Applications de la classification

La classification automatique est une technique utilisée dans plusieurs domaines. Sa capacité prédictive la rend rapide et efficace. Parmi les applications où la classification est utilisée, nous trouvons le filtrage de spam, en effet il s'agit de traiter les messages électroniques textuels (dans notre cas les tweets), identifier leurs caractéristiques et les classer en groupes selon leurs contenus. [11]

Une autre application est la détermination automatique du sujet d'un texte pour le classer automatiquement afin de notifier des personnes intéressées par ce sujet de la présence d'un nouveau texte...

2. Etapes de la classification

La classification automatique de documents doit être obligatoirement précédée par une phase de préparation de données appelée « preprocessing ».

2.1. Préprocessing

2. 1.1. Extraction de données

L'extraction de données est une tâche qui s'est développée dans le domaine de Traitement Automatique des Langues (TAL).

Elle consiste à identifier et extraire d'un texte les éléments pertinents contenant des informations dont la nature est spécifiée à l'avance.

Elle vise donc à transformer un texte de son format initial (une suite de chaînes de caractères) à une représentation structurée et donc un format qui soit compréhensible par l'ordinateur. Elle se fait en reconnaissant dans le texte des unités lexicales particulières.

Il existe plusieurs techniques d'extraction de données telles que : • Les outils terminologiques : nous nous intéressons aux termes présents dans un texte, ces outils peuvent être linguistiques, statistiques ou mixtes, les méthodes linguistiques se basent sur des patrons lexicaux-syntaxiques et sur le découpage des textes en unités syntaxiques.

Les méthodes statistiques font des calculs sur les fréquences et les distributions des mots dans les textes.

Les méthodes d'extraction de relation terminologique : ici nous nous intéressons plutôt à la relation entre les différents termes et structures des textes.

Là aussi plusieurs approches peuvent être utilisées, on distingue les méthodes structurelles, contextuelles et distributionnelles Les méthodes structurelles se basent sur la structure syntaxique interne des termes. Dans le cas de méthodes contextuelles c'est plutôt le contexte dans lequel apparaît le terme qui est pris en compte.

Pour les méthodes distributionnelles, il s'agit de trouver des relations sémantiques entre les termes d'un texte à partir des statistiques de leur apparition et leur distribution.

La reconnaissance des entités nommées : le plus souvent il s'agit de noms de personnes, de lieux, de noms d'organisations, etc. Selon le domaine traité par le texte d'autres entités peuvent être ajoutées à la liste précédente, par exemple dans un texte médical il est important de reconnaître les noms de maladies ou de symptômes.

Plusieurs études sont faites sur les entités nommées, sur la façon de les identifier et de les reconnaître dans un texte. Pour identifier un terme comme étant une entité nommée, on peut soit s'intéresser au terme lui-même (par exemple s'il commence par une lettre majuscule alors il y a une forte chance qu'il représente une entité nommée) soit à son entourage (par exemple s'il est précédé ou suivi d'un mot qui lui donne une qualification). [12]

Les applications de cette tâche ne se limitent pas à la classification de documents, en effet nous la retrouvons dans la veille technologique, dans les interfaces Homme machines, etc...

2.2. Calcul de la distance de similarité sémantique

On parle de mesure de similarité sémantique lorsque la mesure calcule si deux concepts sont similaires sémantiquement, au sens où ils partagent des propriétés et attributs communs. Par exemple (voiture) et (vélo) sont similaires car ils possèdent tous les deux les attributs d'un (véhicule) a un roue.

Dans notre cas on cherche les tweets qui possèdent des mots Similaire on peut calculer la distance sémantique comme suit :

La distance de similarité sémantique est égale à Le nombre de mots en commun fois 2 divisé par la somme de la longueur des deux tweets

2.3. Classification

C'est à cette étape que se fait l'assignation du document à la classe à laquelle il appartient. La détermination de la classe se fait grâce à des algorithmes de classification qui exploitent les données extraites dans l'étape précédente et qui donnent en sortie la classe correspondante.

Ces algorithmes se basent sur leur expérience passée où ils ont appris comment classer les textes, c'est de là que vient leur nom « Algorithmes d'apprentissage ».

Il existe deux types de ces algorithmes : • Les algorithmes d'apprentissage supervisé • Les algorithmes d'apprentissage non supervisé.

3. Les Techniques de classification automatique

3.1. Apprentissage non supervisé

L'apprentissage non supervisé consiste à apprendre à classer sans supervision. Au début de processus nous ne disposons ni de la définition des classes, ni de leurs nombres. C'est l'algorithme de classification qui va déterminer ces informations.

Nous ne disposons pas non plus de données en entrée qui sont déjà classées, c'est aussi à l'algorithme de découvrir par lui-même la structure plus ou moins cachée des données et de former des groupes d'individus dont les caractéristiques sont communes. [13]

L'apprentissage non supervisé est utilisé dans plusieurs domaines tels que : • Médecine : Découverte de classes de patients présentant des caractéristiques physiologiques communes. • Le traitement de la parole : construction de système de reconnaissance de la voix humaine. • Archéologie : regroupement des objets selon leurs époques. • Traitement d'images • Classification de documents.

Exemples :

Dans la littérature il existe plusieurs types d'algorithmes d'apprentissage non supervisé tels que les algorithmes de partitionnements et les algorithmes de classification hiérarchique :

- Le partitionnement : consiste au regroupement des données suivant leur degré de similarité. L'algorithme le plus célèbre appartenant à cette classe est K-means : c'est un algorithme qui permet de partitionner un ensemble de données automatiquement en K clusters.

Il consiste tout d'abord à choisir k points qui représentent les centres des groupes à créer, puis à affecter les autres points aux centres les plus proches.

Cette affectation est faite par le calcul de distance entre les points. Plusieurs distances peuvent être définies telles que la distance euclidienne ou la distance de Manhattan. Par la suite nous procédons à une étape de raffinement des groupes de façon itérative, le raffinement se fait par le recalcule des centres des groupes après chaque itération et par une réaffectation des points aux groupes. L'algorithme s'arrête quand aucun point ne bouge. [13]

•La classification hiérarchique : il existe deux types de classification hiérarchique : Ascendante et descendante. La classification ascendante consiste à utiliser une matrice de similarité afin de partir d'une répartition fine vers un groupe unique.

Donc, il s'agit de fusionner les groupes jusqu'à ce qu'on obtient un seul groupe englobant tous les autres. Cette classification peut être représentée par un arbre hiérarchique ou dendrogramme. La classification descendante se présente comme l'inverse de la classification ascendante. Donc il s'agit de décomposer un cluster unique en sous-groupes jusqu'à l'obtention des singletons. [13]

3.2. Apprentissage supervisé

Contrairement à l'apprentissage non supervisé, nous commençons ici par un ensemble de classes connues et définies à l'avance. Nous disposons aussi d'une sélection initiale de données dont la classification est connue. Ces données sont supposées indépendantes et identiquement distribuées. Elles nous servent pour l'apprentissage de l'algorithme. La classification se fait par l'algorithme selon le modèle qu'il a appris. [14]

Exemples

Il existe plusieurs algorithmes d'apprentissage supervisé, cette section présente quelques-uns des plus connus parmi eux, il s'agit de kNN, LLSF, les réseaux de neurones et Naive Bayes [15].

• kNN : (k nearest neighbor), c'est une approche statistique de classification très connue. Il a été prouvé que c'est une des méthodes les plus performantes après des tests réalisés sur le corpus de données Reuters.

Le principe de l'algorithme kNN est le suivant : étant donné un texte à classer, l'algorithme cherche les k voisins les plus proches parmi les documents utilisés au cours de la phase d'apprentissage, les catégories de ces k voisins les plus proches serviront à donner des poids aux catégories candidates de classification. C'est le degré de similarité entre le document test et le document voisin qui est utilisé comme poids de la catégorie de ce dernier, si plusieurs voisins partagent la même catégorie alors le poids attribué à cette catégorie est égal à la somme des degrés de similarité entre le document test et chacun des voisins appartenant à cette catégorie. Par cette méthode on peut obtenir une liste des poids attribués à

chaque catégorie, le document test est classé dans une catégorie si le poids attribuée à celle-ci est supérieur à un seuil fixé à l'avance.

- Les réseaux de neurones : c'est une structure constituée de suite successive de couches de nœuds et qui permet de définir une fonction de transformation non linéaire des vecteurs d'entrées (composés dans le cas de classification des mots pondérés de leur poids) en vecteur de catégories.

La disposition des neurones dans le réseau ainsi que le nombre de couches utilisées ont une influence sur le résultat de classification. Comparés aux autres méthodes de classification par apprentissage supervisé, les réseaux de neurones ont l'inconvénient que le coût d'apprentissage est assez élevé.

- NB (Naive Bayes) : c'est une méthode de classification probabiliste.

Elle consiste à utiliser les probabilités jointes des mots et des catégories pour estimer la probabilité d'une catégorie sachant un texte à classer. Le caractère « naïf » de cette approche est dû au fait que les mots sont considérés indépendants, c'est-à-dire que la probabilité conditionnelle d'un mot sachant une catégorie est supposée indépendante des probabilités conditionnelles des autres mots sachant la même catégorie, cette assumption rend NB très efficace par rapport aux autres approches bayésiennes.

Plusieurs versions de NB sont proposées dans la littérature, le model mixte multinomial par exemple a permis d'avoir de bonnes performances. [16]

II. La recommandation

1. Les systèmes traditionnels de recommandation

Les systèmes de recommandation effectuer un filtrage de l'information en proposant à un utilisateur un peu nouveaux éléments (par exemple, des chansons, des livres ou des films) pour l'achat ou écoute [A].

Il y a deux types de systèmes de recommandation - *personnalisés* et *non personnalisée*. Les systèmes de recommandation personnalisés considèrent les préférences des utilisateurs à recommander. Les systèmes de recommandation non personnalisés toutefois, ne pas faire usage de préférences de l'utilisateur.

Un exemple de non-personnalisée méthode de recommandation est de retourner les dix meilleures chansons du mois en cours. Les méthodes de recommandation qui seront étudiés dans le présent document sont personnalisées.

Systèmes de recommandations personnalisées utilisent les caractéristiques des articles, des profils d'utilisateurs et les interactions ou transactions entre les utilisateurs et les articles pour prédire les utilisateurs futures adoptant un article. Le filtrage collaboratif et les approches basées sur le contenu sont souvent utilisés dans la recommandation personnalisée

2. Le filtrage collaboratif et la recommandation basée sur le contenu

L'hypothèse du *filtrage collaboratif basé utilisateur à utilisateur* est que si une personne X a la même opinion comme une personne Y sur une question A , X est plus susceptible d'adopter l'avis de Y sur une question différente B qu'un choisie au hasard personne.

Le système de recommandation trouve des gens avec des goûts ou des préférences similaires, selon leurs évaluations passées ou interactions implicites.

Ensuite, le système prédits la préférence d'un utilisateur sur un élément classé en utilisant les préférences de similaire utilisateurs [B].

Une autre approche de recommandation personnalisée est *filtrage collaboratif basé article à article*. Dans cette méthode de recommandation deux articles A et B sont très similaires si une partie relativement importante des utilisateurs qui achètent l'article A achètent aussi l'article B . Ensuite, la préférence d'un utilisateur sur un élément classé B est prédite en se basant sur *l'article* classé A de l'utilisateur.

Le système de recommandation basé sur le contenu classe des articles similaires en comparant leurs particularités et caractéristiques communes. Ensuite, la recommandation d'un article est faite pour l'utilisateur qui aime ou achète des articles similaires avant.

En d'autres termes, divers articles de candidats sont comparés à des éléments précédemment notés par l'utilisateur et les articles les plus correspondants sont recommandés. [C]

3. Autres approches dans les médias sociaux

Les systèmes de recommandation sociaux recommandent les articles sur la base des préférences des amis de l'utilisateur ou d'autres informations sur les médias sociaux, tels que des étiquettes et des commentaires.

Les articles recommandés pourraient ne pas être nécessairement composants de réseau social. Par exemple, dans le cas de Twitter, on peut recommander des articles de presse ou des photos. [D] et voilà quelque exemple de recommandation sur Twitter

3.1. Recommandation des (Follower)

Dans Twitter, les utilisateurs sont intéressés à trouver non seulement leurs amis proches, mais aussi de nouveaux contacts pertinents.

Un utilisateur peut suivre d'autres utilisateurs qui partagent sujets tendances intéressantes.

Ces utilisateurs peuvent être traités comme des sources d'information pour l'utilisateur. Différents algorithmes de recommandation follower peuvent être utilisés. [E]

3.2. Recommandation des hashtags

Il y a plusieurs fins de l'utilisation des hashtags. Certaines personnes les utilisent pour Catégoriser leurs tweets. Certains les utilisent comme support de diffusion de masse en cas de catastrophes ou des événements spéciale comme les élections. Hashtags sont également utilisés pour la promotion de la marque [F].

Les hashtags ne sont ni enregistrés ni contrôlés par n'importe quel utilisateur ou d'un groupe, il peut être difficile pour certains utilisateurs de trouver hashtags appropriées pour leur tweets. Par conséquent, les systèmes de recommandation pour suggérer hashtags appropriées aux utilisateurs sont proposés.

3.3. Recommandation des tweets

Tous les tweets des amis que tu suis sont affichées dans la page d'accueil de l'utilisateur.

Lorsque l'utilisateur suit de nombreux utilisateurs actifs, il y a des chances que l'utilisateur pourrait manquer la lecture de certains tweets intéressants.

En fonction des préférences de l'utilisateur, les tweets importants peut être choisi et a souligné (recommandés) [J]

Conclusion

Dans ce chapitre, nous avons présenté les notions classification et la recommandation (définitions, principe, étapes).

Et dans le chapitre suivant nous allons passer à l'étape de la modélisation de notre système.

Chapitre3 : la modélisation

Introduction

Dans ce chapitre nous présentons la modélisation de notre système de classification et de recommandation.

Notre travail consiste à faire des recommandations que ce soit des tweets ou des amis à la base des résultats de la phase de Clustering, en analysant les clusters pour trouver des utilisateurs de Twitter qui ont des intérêts en commun (des sujets de discussion, des loisirs, des opinions politiques ...)

Notre système se compose de trois parties essentielles,

1. Le module de téléchargement des tweets
2. Le module de Clustering de tweets
3. La recommandation des tweets à lire et/ou des amis à suivre

I. La description de notre système :

Comme déjà mentionné, notre système se compose de 3 phases :

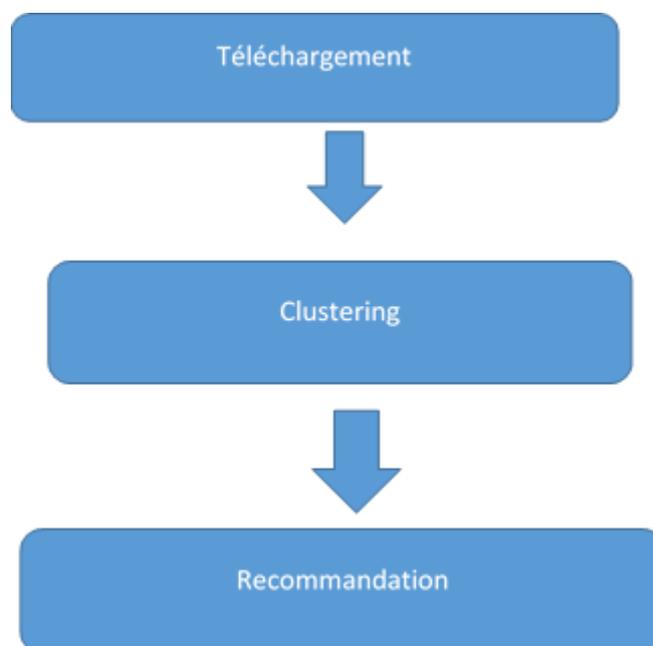


Figure 3 : les trois phases de notre système.

Par la suite, nous détaillons chaque phase :

1. Téléchargement :

Cette phase consiste à télécharger les tweets et les met dans une base de données.

Avant de stocker les tweets, il faut procéder à un prétraitement pour raffiner les données brutes.

Notre prétraitement consiste à éliminer les « stop Word »

1. 2 Elimination de stop words :

Cette étape consiste à supprimer tous les mots standards dans un texte, ce sont des mots très communs et utilisés dans pratiquement tous les textes.

Leur présence peut dégrader la performance de l'algorithme de classification en termes de coût et en termes de précision de la classification.

Nous pouvons prendre comme exemple pour l'anglais I, am, with, what, you ... Et pour le français il s'agit d'éliminer par exemple :

- Les pronoms pronominaux : il, elle, je, lui, ceux....
- Les auxiliaires : avoir, être et toutes leurs conjugaisons possibles • Les opérateurs de conjonction : avec, et, ou, aux... • L', c' ...

2. Le Clustering :

Nous avons adopté la classification selon le contenu et nous avons choisi une méthode naïve parce que c'est une méthode simple et rapide à implémenter.

L'algorithme de cette méthode est un algorithme itératif, son principe se base sur la comparaison sémantique entre un tweets donné et les autres tweets pour trouver les tweets le plus proche sémantiquement.

Nous donnons par la suite un pseudo code de l'algorithme utilisé.

2. 1 L'algorithme de Clustering :

Initialisation :

Indice_max =0 ;

Dis_max =0 ;

Dis =0 ;
Ci =0 ;
Seuil =0.5 ;

Traitement

Pour i :=1 à nb_tweets faire

Début

Dis_max=-1 ;
Indice_max=-1 ;

Pour j :=1 à nb_tweets faire

Si (i<>j) alors

Début

Dis = distance _sémantique (tweets[i], tweets[j])

Si (d > dis_max) alors

Début

Indice_max :=j ;

Dis_max :=dis ;

Fin

Fin ;

Fin ;

i (dis_max > seuil) alors

Début

Si (Clusters [indice_max] <> 0) alors

Clusters[i] := Clusters [indice_max] ;

Sinon

Début

ci := ci+1 ;

Clusters [indice_max] :=ci ;

Clusters[i] :=ci ;

```
    Fin ;  
  Fin ;  
Sinon  
  Début  
  
  Ci := ci+1 ;  
  Clusters[i] :=ci ;  
  Fin ;  
  
Fin.
```

2. 2 La procédure de distance de similarité sémantique

Procédure distance sémantique (tweet1, tweet2, var dis_sem)

```
  Début  
  Nb_cara_com :=0 ;  
  Dis_sem :=0 ;  
  taille_tweets1= longueur (tweet1) ;  
  taille_tweets2= longueur (tweet2) ;  
  Pour i := 1 à taille_tweets1 faire  
    Pour j := 1 à taille_tweets2 faire  
      Si tweet[i] =tweet[j] alors  
        Nb_cara_com = Nb_cara_com +1 ;  
      Fsi ;  
  
  Dis_sem := 2* Nb_cara_com/ (taille_tweets2 + taille_tweets2) ;  
Fin procédure ;
```

Remarque :

Seuil est une valeur entrée.

2. 3 L'explication de l'algorithme de Clustering

Avant l'exécution, tous le tweets ne sont pas affectés à des clusters

Les résultats de l'exécution se séparent en deux étapes :

1. 1^{er} étape :

L'algorithme fixe un tweet et itère sur les autres. Pour trouver le tweet le plus proche à l'aide d'un algorithme de max et une procédure de mesure de similarité sémantique entre eux.

2. 1^{er} étape :

Après la comparaison si on trouve un tweet le plus proche à le premier tweet on aura 2 possibilités

- a. Le tweet est affecté au cluster de tweet le plus proche si ce dernier est déjà affecté à un cluster.
- b. Sinon les deux tweets (tweet et son plus proche tweet) sont affectés à un nouveau cluster, l'algorithme incrémente le nombre de clusters créés.

Si, après la comparais, l'algorithme ne trouve pas un tweet le plus proche, notre tweet est affecté à un nouveau cluster, l'algorithme incrémente le nombre de clusters créés.

2. 4 L'explication de la procédure de similarité sémantique

Cette procédure consiste à calculer :

La taille de chaque tweet (le nombre de mots dans le tweet)

Ensuite, elle calcule le nombre de mots répétés (c'est-à-dire les mots en commun) entre les deux tweets.

Et puis on fait le calcul la distance sémantique comme suit :

La distance de similarité sémantique est égale à Le nombre de mots en commun fois 2 divisé par la somme de la longueur des deux tweets.

3. La recommandation :

Dans cette partie nous présentons le dernier module qui concerne la recommandation des amis à suivre et/ou des tweets à lire.

Nous choisissons une recommandation à base de contenu à l'aide des résultats de l'algorithme de Clustering de la phase précédente.

Le principe est d'exploiter les Clusters comme des groupes des intérêts en commun c'est-à-dire si nous trouvons deux tweets différents postés par deux utilisateurs différents, nous considérons la possibilité de recommander l'un à l'autre.

3. 1 L'algorithme de recommandation :

Initialisation

```
User="" ;  
Cluster=" " ;
```

Traitement

```
    Pour i :=1 à nb_tweets faire  
Début  
Si (user_traité= faux) alors  
    Début  
        User := resultat_cluster[i] [3] ;  
        Cu := nombre_tweets ;  
        Ecrire ('user+1 recommandation') ;  
    Fsi ;  
Fpour ;  
    Pour i := 1 à longueur (tweet) faire  
        Si (nombre_tweets[i] [2] ==cluster) alors  
            Si (nombre_tweets[i] [0] ==user) alors  
                Nombre_tweets[i] [3] ==x ;  
            Sinon  
                Fsi ;  
    Fpour ;  
Fin.
```

Remarque :

Le mode est une valeur entrée.

Si on a choisi le mode 1 cet algorithme va recommander les amis à suivre.

Si on a choisi le mode 2 cet algorithme va recommander les tweets.

3. 2 L'explication de l'algorithme de recommandation :

Dans cette phase l'exécution va baser sur les Cluster.

Commençant par le premier tweet on vérifie la 4^{ème} colonne et on cherche tous les tweets qui ont le même cluster que le premier tweet et n les recommander chacun a l'autres et ainsi de suite avec les autres tweets.

Exemples de résultats :

➤ **Clustering**

Le résultat de la phase Clustering sera un tableau contient 4 colonnes le numéro, User, Tweet et Cluster comme celui :

Num	User	Tweet	Cluster
1	Mohamed	J'aime beaucoup le football.	1
2	Imane	Il reste que 1 mois pour le meilleur mois de l'année (ramadan).	2
3	Nouria	J'ai un petit problème dans mon application qui était programmé par le langage JavaScript.	3
4	Ahmed	J'ai un match de football à bientôt (j'aime).	1
5	Hala	JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives mais aussi pour les serveurs	3
6	Islam	Visca el Barça, le meilleur club de football dans le monde j'aime beaucoup	1

Tableau1 : les résultats de Clustering

➤ **Recommandation des Amis**

User	Amis recommandés
Mohamed	- Ahmed - Islam

Nouria	- Hala
Ahmed	- Mohamed - Islam
Imane	
Hala	- Nouria
Islam	- Mohamed - Ahmed

Tableau 2 : les résultats de Recommandation des Amis

➤ **Recommandation des Tweet**

User	Tweet recommandés
Mohamed	- Visca el Barça, le meilleur club de football dans le monde j'aime beaucoup. - J'ai un match de football à bientôt (j'aime).
Nouria	- JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives mais aussi pour les serveurs.
Ahmed	- J'aime beaucoup le football. - Visca el Barça, le meilleur club de football dans le monde j'aime.

	<p>beaucoup.</p> <ul style="list-style-type: none"> - J'ai un match de footballà bientôt (j'aime).
Imane	
Hala	<ul style="list-style-type: none"> - J'ai un petit problème dans mon application qui était programmé par le langage JavaScript.
Islam	<ul style="list-style-type: none"> - J'aime beaucoup le football. - J'ai un match de footballà bientôt (j'aime).

Tableau3 : les résultats de Recommandation des Tweets

Conclusion :

Ce chapitre a été consacré à la modélisation. Après avoir effectué une étude préliminaire, passant par les étapes du processus de développement de notre projet.

Et dans le chapitre suivant nous allons passer à l'étape d'implémentation de notre système.

Chapitre4 :L'implémentation

Introduction

A ce stade, les cas d'utilisation sont terminés, le problème a été analysé en profondeur .Nous avons défini une conception mieux appropriée aux besoins de l'application. Nous pouvons alors entreprendre le dernier chapitre qui contient les outils de programmation et quelque capture d'image de notre application.

I. les outils de programmation

1. Qu'est-ce que Java ?

Java est un langage de programmation orienté objet et un environnement d'exécution, développé par Sun Microsystems. Il fut présenté officiellement en 1995.

Le Java était à la base un langage pour Internet, pour pouvoir rendre plus dynamiques les pages (tout comme le JavaScript aujourd'hui). Mais le Java a beaucoup évolué et est devenu un langage de programmation très puissant permettant de presque tout faire, je dis bien presque car nous verrons pourquoi il ne permet pas de tout faire.

Java est aujourd'hui officiellement supportée par Sun, mais certaines entreprises comme IBM font beaucoup pour Java.

Java est complètement orienté objet. Java vous permet et vous pousse même à développer vos applications d'une façon orientée objet et vous permet d'avoir une application bien structurée, modulable, maintenable beaucoup plus facilement et efficace. Cela augmente une fois de plus votre productivité [17]

➤ Java et l'api

Contrairement à la plupart des autres langages (sauf la plateforme .Net), Java met à la disposition du développeur une API très riche lui permettant de faire de très nombreuses choses.

Contrairement à des langages comme le C, où il fallait coder par nous-même des fonctionnalités basiques comme le chargement d'images, ou bien d'avoir recours pour quasiment n'importe quoi à une librairie de code tierce, Java vous propose à peu près tout ce dont vous avez besoin directement dans le JDK.

Ceci est un énorme avantage, qui augmente encore une fois grandement votre productivité de développement.

De plus il existe énormément d'API tierces de très bonnes qualités, pour des fonctionnalités qui viendraient à manquer au JDK. Nous avons donc vu que Java possède de nombreux avantages non négligeables.

➤ Résumé des avantages de Java :

- Portabilité excellente
- Langage puissant
- Langage orienté objet
- Langage de haut niveau
- JDK très riche
- Nombreuses librairies tierces
- Très grande productivité
- Applications plus sûres et stables
- Nombreuses implémentations, JVM et compilateurs, libres ou non
- IDE de très bonne qualité et libres : Eclipse et Netbeans par exemple
- Supporté par de nombreuses entreprises telles que Sun ou encore IBM et des projets comme Apache [17]

2. Netbeans 8

Netbeans passe de la version 7.4 à la version 8 pour s'aligner avec Java 8. À la suite de la publication de la plateforme de développement, Oracle a sorti une nouvelle version de son environnement de développement open source.

Cette nouvelle mouture apporte un nombre important de nouvelles fonctionnalités pour les développeurs Java, C/C++, HTML5, JavaScript et PHP.

NetBeans 8 se présente comme l'outil idéal pour découvrir les nouveautés de Java 8, grâce aux améliorations qui ont été apportées à l'outil et à l'éditeur de code pour travailler convenablement avec les expressions Lambadas, Profiles et Streams. [18]

3. Api Twitter for java

Twitter4J est une bibliothèque Java officieux de l'API Twitter.

Avec Twitter4J, vous pouvez facilement intégrer votre application Java avec le service Twitter. Twitter4J est une bibliothèque officieuse.

- **Twitter4J est :**

✓ 100 % Pure Java - fonctionne sur toutes les versions plate-forme Java 5 ou ultérieure

✓ Plate-forme Android et Google App prêt

✓ Dépendance Zéro : Aucun pot supplémentaire nécessaire

✓ Prise en charge intégrée

Clustering de tweets pour la recommandation

✓ 100 % Twitter API 1.1 compatible

- **Comment Utiliser ?**

Il suffit d'ajouter twitter4j -core- 4.0.3.jar à votre application classpath.

Si vous êtes familier avec le langage Java, en regardant dans la Java DOC devrait être le chemin le plus court pour vous permettre de démarrer. [19]

- **Les package**

```
Package com.ubiteck.sample ;
```

```
Import java.util.logging.Logger ;
```

```
Import twitter4j.Query ;
```

```
Import twitter4j.QueryResult ;
```

```
Import twitter4j.Status ;
```

```
Import twitter4j.Tweet ;
```

```
Import twitter4j.Twitter ;
```

```
Import twitter4j.TwitterException ;
```

```
Import twitter4j.TwitterFactory ;
```

```
Import twitter4j.auth.AccessToken ;
```

```
Import twitter4j.auth.RequestToken ;
```

- **La connexion avec le Twitter**

```
Try {
```

```
ConfigurationBuilder cb = new ConfigurationBuilder () ;
```

```
cb.setDebugEnabled (true)
```

```
.setOAuthConsumerKey ("vWGIV2hNVbF6sEzZ0HQWdA")
```

```
.setOAuthConsumerSecret ("mRBKL4SMFEfonE6v1id4hvt0JwFG1qn0wh19a1LQ")
```

```
.setOAuthAccessToken ("1642606040FZwxKLIOGdzqHW7swEpsIE3wTPScYYKku7762o")
```

```
.setOAuthAccessTokenSecret ("Xxba4EPhlzFFAjDri7qhU5RQakPxfa1aMEw161vRlmt") ;
```

- **Les fonctionnalités du Twitter 4 java**

On peut faire beaucoup de chose avec cet Api, par exemple on peut :

- Consulter notre timeline (le mur).

- Envoyer de tweets.
- La publication.
- Télécharger les tweets de même utilisateur.
- On peut aussi télécharger des tweets avec la date...etc.

II. Description de l'application

1. Page d'accueil

Elle contient la barre de menu et le bouton quitter

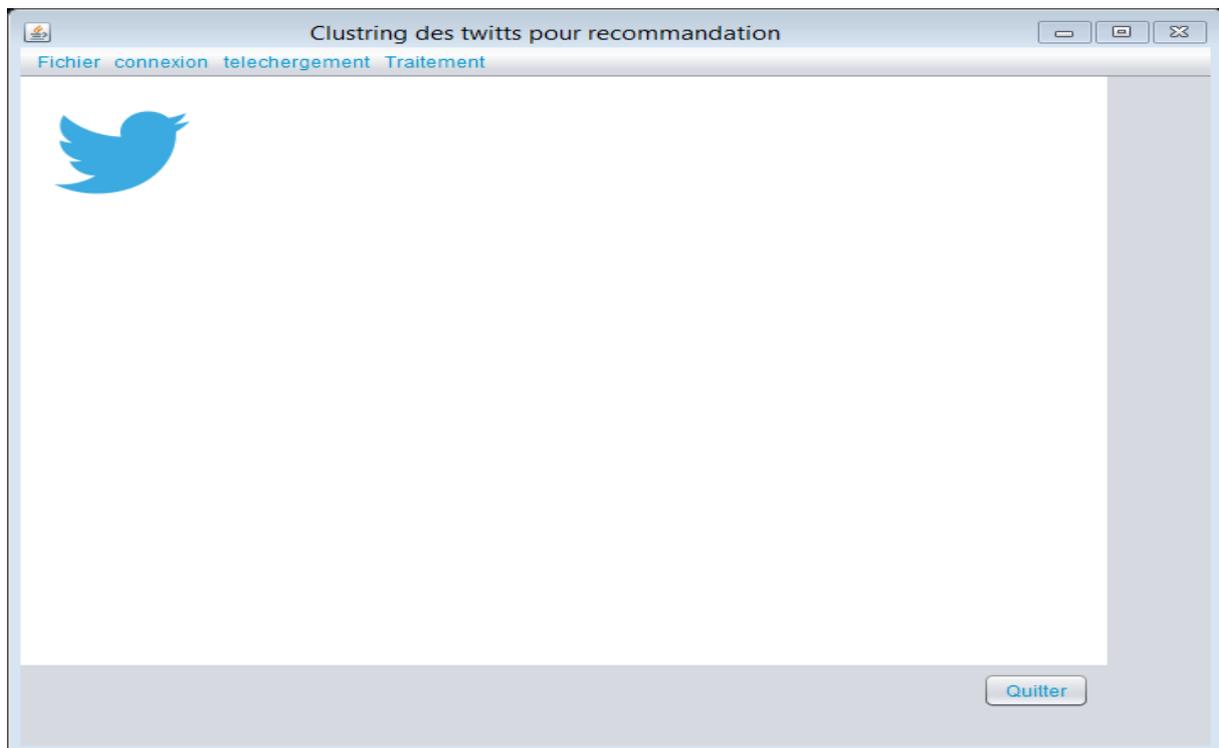


Figure 4 : Page d'accueil.

2. La barre de menu

La barre de menu contient des boutons :

2.1. Le bouton fichier



Figure 5 : le bouton fichier.

2.2. Le bouton connexion

Elle contient le bouton **connecter** qui fait la connexion avec Twitter

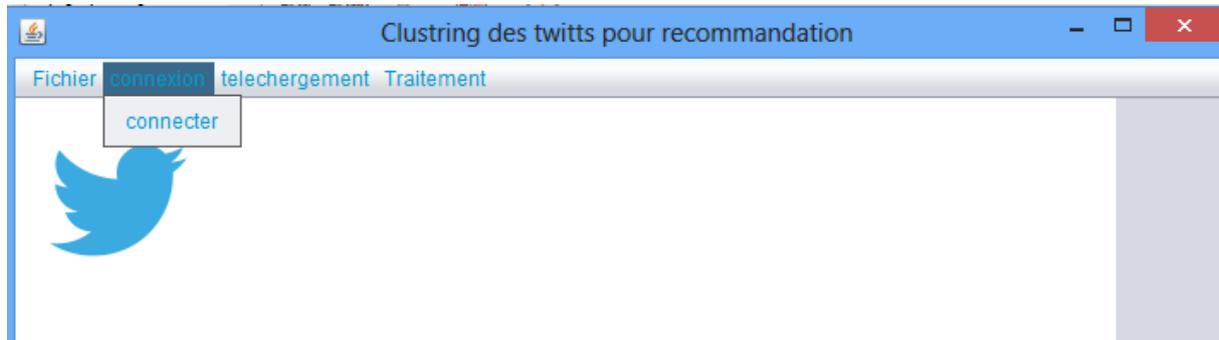


Figure 6 : Le bouton connexion.

2.3. Le bouton téléchargement

Elle contient le bouton **charger** qui fait le téléchargement du tweets

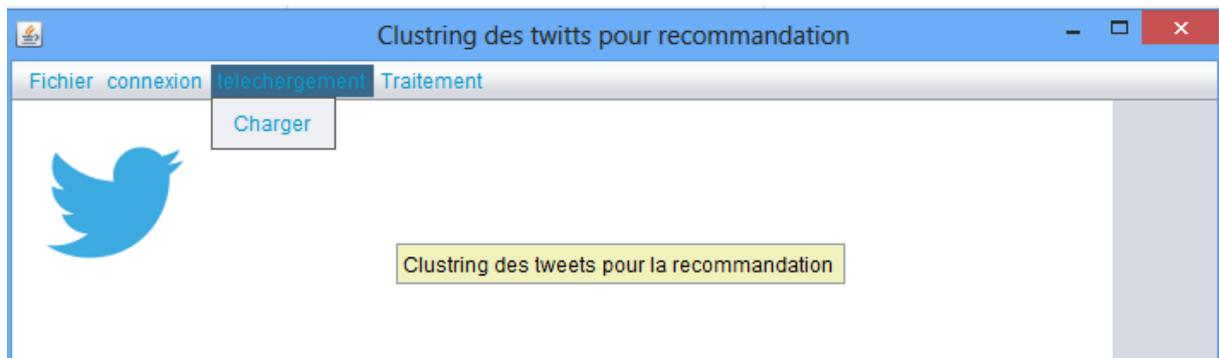


Figure 7 : Le bouton téléchargement.

2.4. Le bouton traitement

Elle contient les boutons :

- **Clustering** qui fait le Clustering.
- **Recommandation** qui contient de bouton **AMIS** et **TWEET**



Figure 8 : Le bouton traitement.

3. La fenêtre de connexion

Elle contient la connexion avec Twitter.

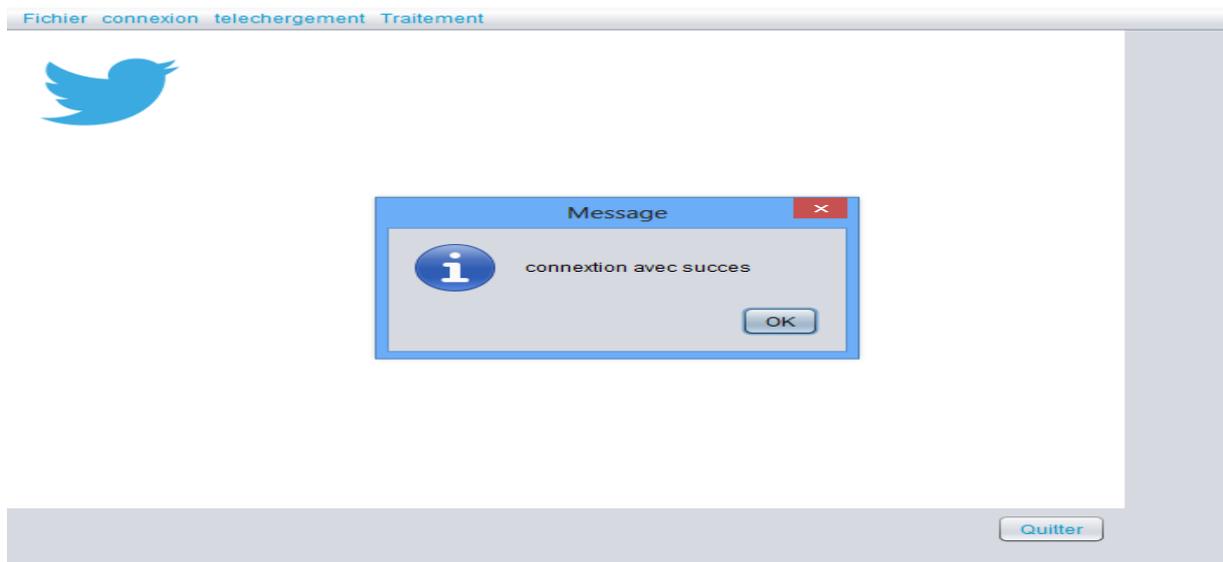


Figure 9 : La fenêtre de connexion.

4. La fenêtre de chargement

Cette fenêtre contient :

- le variable à saisir **mot clé**.
- le mode de chargement **hors linge** et **En linge**.
- Le bouton **ok** pour le chargement.

- Le bouton **afficher** pour afficher les tweets téléchargé.

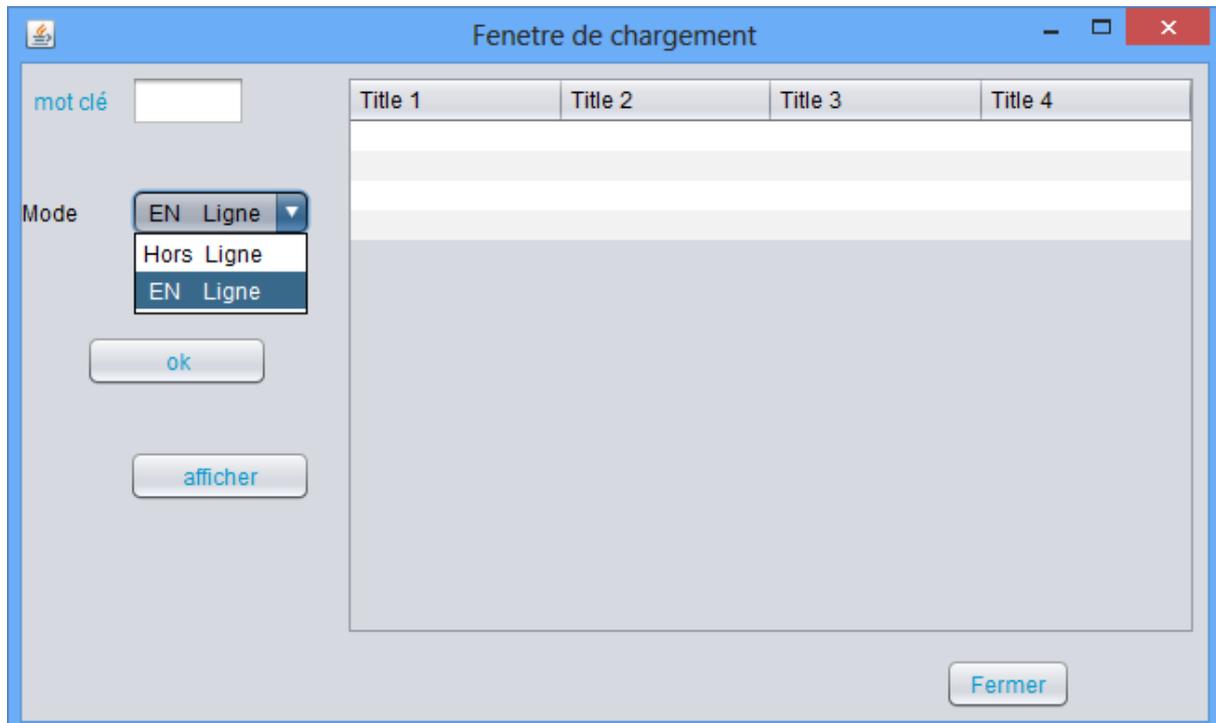


Figure 10 : La fenêtre de chargement.

3.1. Les résultats de chargement

➤ Avec le mode hors linge

Si on a choisi le mode **hors linge** cette fonction va récupérer les tweets à partir de fichier texte

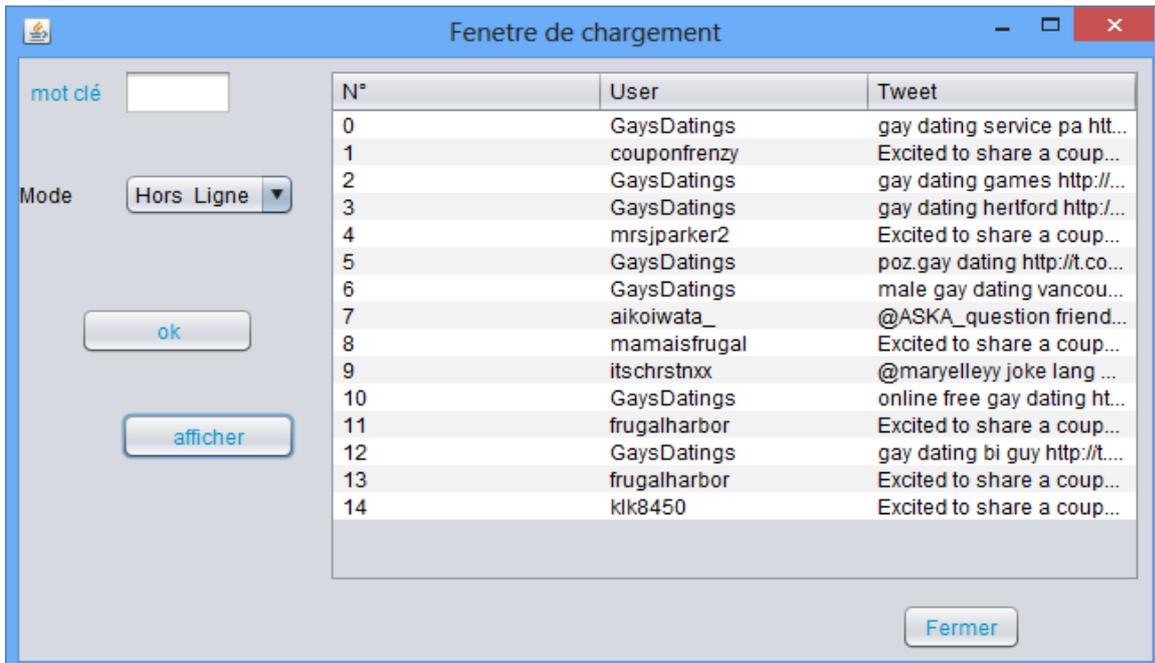


Figure 11 : Les résultats de chargement Avec le mode **hors linge**

➤ **Avec le mode En linge**

Si on a choisi le mode **En linge** cette fonction va récupérer les tweets de Twitter directement

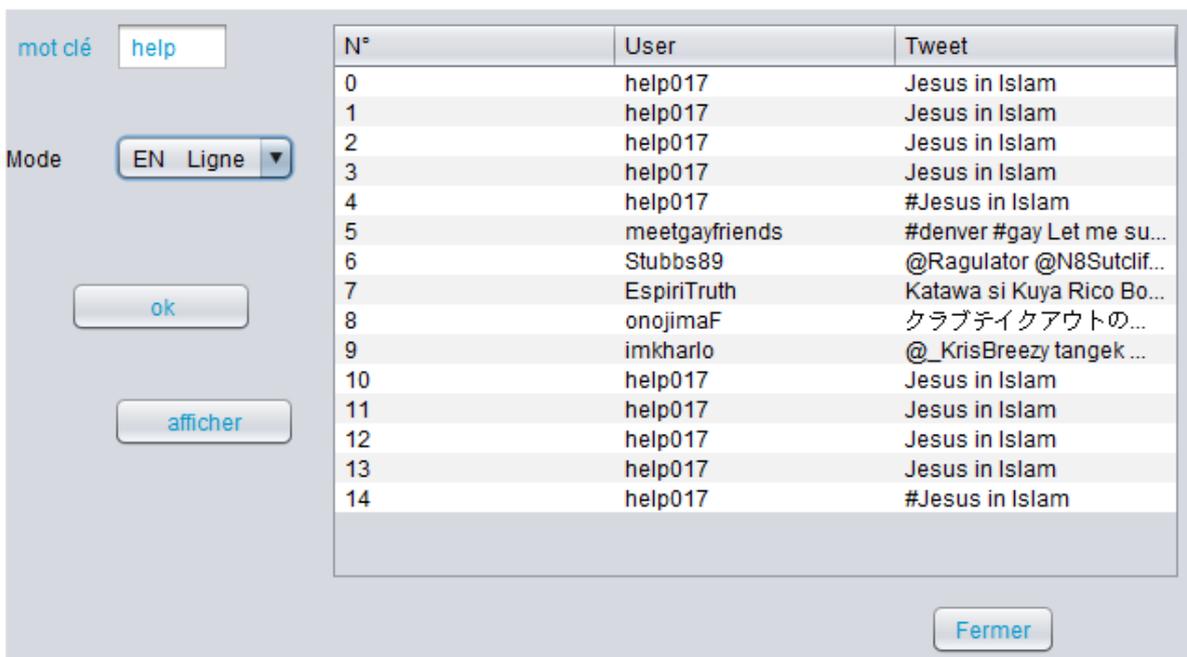


Figure 12 : Les résultats de chargement Avec le mode **En linge**.

5. La fenêtre de Clustering

- le variable à saisir **seuil**.
- Le bouton **ok** pour le Clustering.
- Le bouton **afficher** pour afficher les Cluster.

Et on aura comme résultat :

- Le nombre de tweets téléchargé.
- Le nombre de cluster.

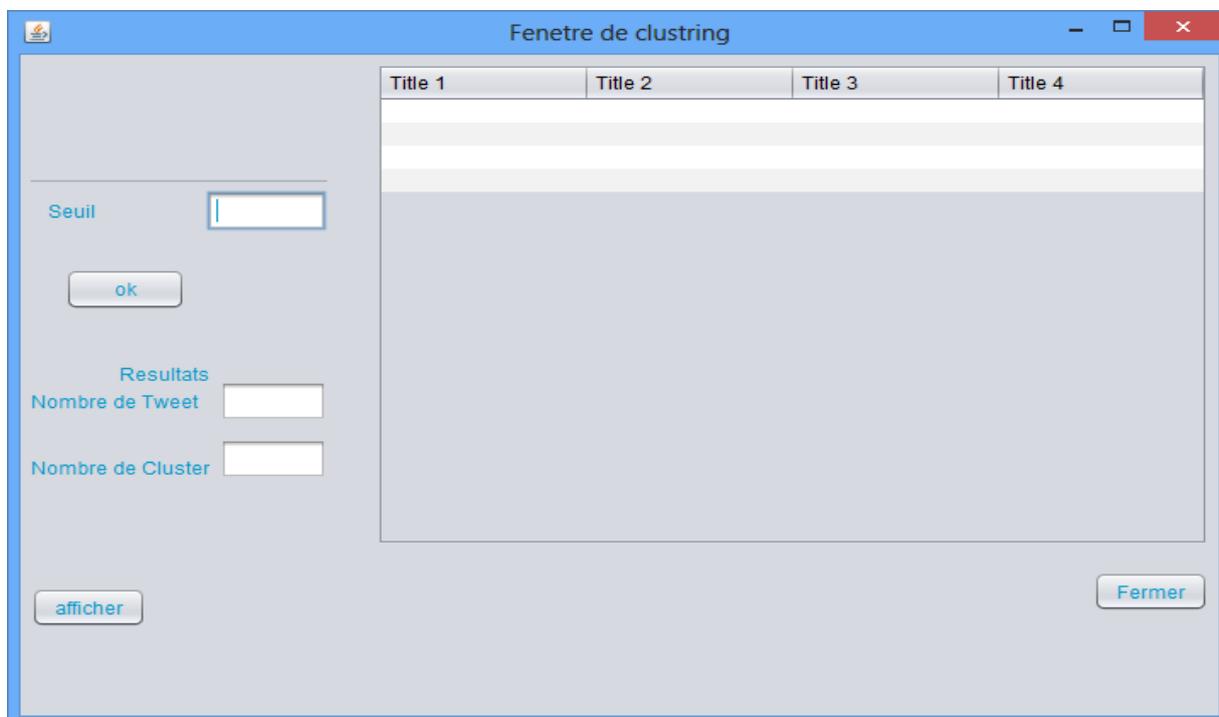


Figure 13 : La fenêtre de Clustering.

5.1. Les résultats de Clustering

On exécute cette fonction plusieurs fois et chaque fois on change la valeur de **seuil** pour trouver la déférence entre les résultats

On exécute cette fonction avec :

- Seuil=0.2
- Seuil=0.9
- Seuil=0.01

- Seuil=0.5
- Avec seuil =0.2

Le nombre des tweets=15

Le nombre de cluster=17

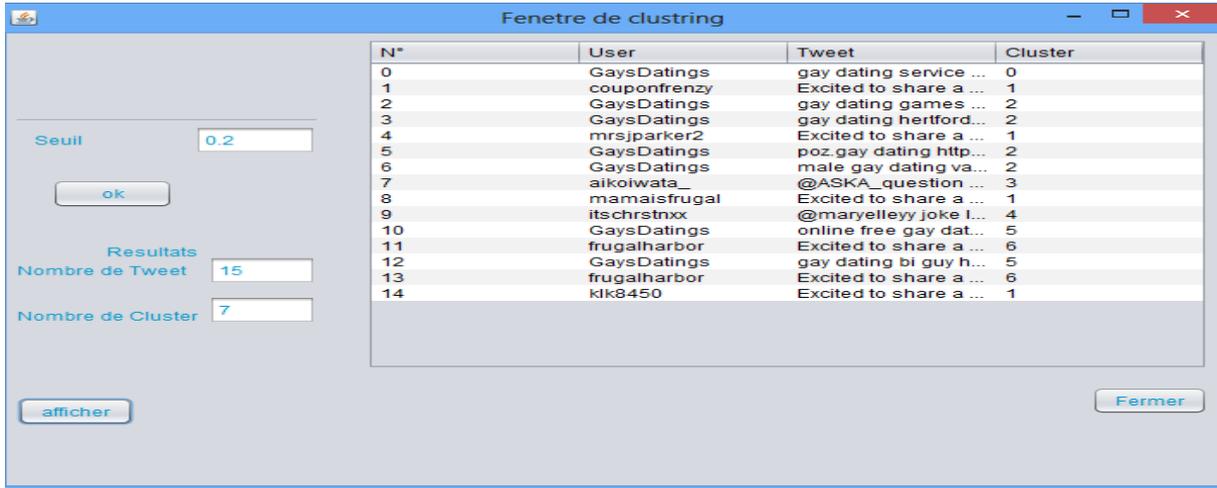


Figure 14 : Les résultats de Clustering avec seuil =0.2.

Avec seuil =0.9

- Le nombre des tweets=15
- Le nombre de cluster=11



Figure 15 : Les résultats de Clustering avec seuil =0.9.

- Avec seuil =0.01
- Le nombre des tweets=15
- Le nombre de cluster=6

N°	User	Tweet	Cluster
0	GaysDatings	gay dating service ...	0
1	couponfrenzy	Excited to share a ...	1
2	GaysDatings	gay dating games ...	2
3	GaysDatings	gay dating herford...	2
4	mrsjparker2	Excited to share a ...	1
5	GaysDatings	poz.gay dating http...	2
6	GaysDatings	male gay dating va...	2
7	aikoiwata_	@ASKA_question ...	3
8	mamaisfrugal	Excited to share a ...	1
9	itschrstnxx	@maryelley joke l...	3
10	GaysDatings	online free gay dat...	4
11	frugalharbor	Excited to share a ...	5
12	GaysDatings	gay dating bi guy h...	4
13	frugalharbor	Excited to share a ...	5
14	klk8450	Excited to share a ...	1

Figure 16 : Les résultats de Clustering avec seuil =0.01.

- Avec seuil =0.5
- Le nombre des tweets=15
- Le nombre de cluster=7

N°	User	Tweet	Cluster
0	help017	Jesus Islam	0
1	help017	Jesus Islam	0
2	help017	Jesus Islam	0
3	help017	Jesus Islam	0
4	help017	#Jesus Islam	1
5	meetgayfriends	#denver #gay suck...	2
6	Stubbs89	@Ragulator @N8...	3
7	EspiriTruth	Katawa si Kuya Ri...	4
8	onojimaF	クラブテイクアウ...	5
9	imkharlo	@_KrisBreezy ...	6
10	help017	Jesus Islam	0
11	help017	Jesus Islam	0
12	help017	Jesus Islam	0
13	help017	Jesus Islam	0
14	help017	#Jesus Islam	1

Figure 17 : Les résultats de Clustering avec seuil =0.5

➤ **Remarque**

On peut remarquer qu'il y a une relation positive ente le seuil et le nombre des clusters

Si on augmente la valeur de seuil, le nombre des clusters augment.

6. La fenêtre de recommandation

Elle contient le bouton **Recommander** et le bouton **Fermer**



Figure 18 : La fenêtre de recommandation.

- **Recommandation des amis a suivi**



Figure 19 : La recommandation des amis.

- **Recommandation des tweets a suivi**



Figure 20 : La recommandation des tweets a suivi.

Conclusion

Dans ce chapitre, nous avons décrit brièvement le processus de réalisation de notre application, les outils de programmation, l'implémentation.

En effet, nous avons achevé l'implémentation et les tests de tous les cas d'utilisation.

Conclusion générale

Ce projet de fin d'étude consiste à réaliser une application qui fait la classification de tweets et puis la recommandation des amis et /ou des tweets a suivi

On a commencé par les définitions de base de web social, les réseaux sociaux et en particulier le Twitter comme un réseau essentiel de notre projet.

Nous souhaitons que notre recherche constitue un support d'aide aux votre travaux futurs quand il fera son entrée dans la vie professionnelle d'une façon plus précise et plus large.

En effet, ce travail étant une œuvre humaine, n'est pas un modèle unique et parfait, c'est pourquoi nous restons ouverts à toutes les critiques et nous sommes prêts à recevoir toutes les suggestions et remarques tendant à améliorer d'avantage cette étude. Etant donné que tout travail informatique a été toujours l'œuvre d'une équipe.

Bibliographie

[A] : Armentano, M.G., Godoy, D.L., Amandi, A.A : A Topology-Based Approach for Followees Recommendation in Twitter. In : 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems. Barcelona, Spain (July 2011).

[B] : Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E : Short and Tweet : Experiments on Recommending Content from Information Stérâmes. In : The 28th International Conference on Human Factors in Computing Systems (2010).

[C] : Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., Liao, W.K : Social media evolution of the Egyptian revolution. Communications of ACM 55(5), 74–80 (May 2012).

[D] : Nasir fard, P., Hayes, C : Tadvise : A Twitter Assistant Based on Twitter Lists. In : The 3rd International Conference on Social Informatics. pp. 153–160 (2011).

[E] : De Francisci Morales, G., Gionis, A., Lucchese, C : From Chatter to Headlines : Harnessing the Real-Time Web for Personalized News Recommendation. In : The 5th ACM International Conference on Web Search and Data Mining (2012).

[F] : Garcia, R., Amatriain, X : Weighted Content Based Methods for Recommending Connections in Online Social Networks. In : The 2nd ACM Workshop on Recommendation Systems and the Social Web. Barcelona, Spain (Juin 2010).

[J] : Hannon, J., Bennett, M., Smith, B : Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approches. In : The 4th ACM Conference on Recommender Systems (2010).

Webographie

- [1] http://fr.wikipedia.org/wiki/Web_social consulté le 23 mars 2015
- [2] <http://www.incroyable-webagency.fr/bourg-en-bresse/agence-typo3> consulté le 05 mars 2015
- [3] <https://alabns.wordpress.com/2014/12/06/le-web-social/> consulte le 05 mars 2015
- [4] <http://www.destination-webmarketing.fr/principaux-reseaux-sociaux-description-et-fonction/> (consulté le 26 mars 2015).
- [5] <http://tpenetwork.blogspot.com/2014/01/tpe-les-reseaux-sociaux-presentation.html>
- [6] <https://beneditopedro23.wordpress.com/veille-technologique/>
- [7] <http://realteenagedreamblog.skyrock.com/4.html> consulté le 26 mars 2015
- [8] <http://fr.wikipedia.org/wiki/Twitter> consulté le 26 mars 2015
- [9] <http://fr.wikipedia.org/wiki/Twitter>
- [10] <http://forum.alidhan.net/topic/91048-question-twitter/>
- [11] <https://www.reador.net/2013/03/21/le-data-mining-au-sein-de-zone-project/>
- [12] http://www.zone-project.org/wp-content/uploads/2013/03/ZONE-project_SVM_Ameni_Bouaziz.pdf
- [13] http://bibfac.univ-tlemcen.dz/bibfs/opac_css/doc_num.php?explnum_id=1537
- [14] http://www.zone-project.org/wp-content/uploads/2013/03/ZONE-project_SVM_Ameni_Bouaziz.pdf
- [15] http://bibfac.univ-tlemcen.dz/bibfs/opac_css/doc_num.php?explnum_id=1537
- [16] http://fr.wikipedia.org/wiki/Classification_na%C3%AFve_bay%C3%A9sienne
- [17] <http://info-rital.developpez.com/tutoriel/java/pourquoi/>
- [18] <http://www.developpez.com/actu/69155/NetBeans-s-aligne-sur-Java-8-la-version-8-0-de-l-EDI-open-source-ameliore-ses-outils-pour-supporter-les-expressions-Lambdas-Streams-et-Profiles/>
- [19] <http://twitter4j.org/en/index.html>

Liste du figures

Figure 1 : <i>Le web social de 2000 à 2010</i>	03
Figure 2 : <i>Les réseaux sociaux</i>	06
Figure 3 : les trois phases de notre système.....	20
Figure 4 : Page d'accueil.....	32
Figure 5 : le bouton fichier	32
Figure 6 : Le bouton connexion.	33
Figure 7 : Le bouton téléchargement.	33
Figure 8 : Le bouton traitement.	34
Figure 9 : La fenêtre de connexion.	34
Figure 10 : La fenêtre de chargement.	35
Figure 11 : Les résultats de chargement Avec le mode hors linge.....	36
Figure 12 : Les résultats de chargement Avec le mode En linge.....	36
Figure 13 : La fenêtre de Clustering.....	37
Figure 14 : Les résultats de Clustering avec seuil =0.2.....	38
Figure 15 : Les résultats de Clustering avec seuil =0.9.....	38
Figure 16 : Les résultats de Clustering avec seuil =0.01.....	39
Figure 17 : Les résultats de Clustering avec seuil =0.5.....	40
Figure 18 : La fenêtre de recommandation.....	41
Figure 19 : La recommandation des amis.	41
Figure 20 : La recommandation des tweets a suivi.....	42

Liste des tableaux

Tableau1 : les résultats de Clustering.....	25
Tableau 2 : les résultats de Recommandation des Amis.....	25
Tableau3 : les résultats de Recommandation des Tweets	26

