الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Algiers 1 Benyoucef BENKHEDDA

Faculté des Sciences

Département de Mathématique et Informatique

Mémoire de fin d'étude pour l'obtention du diplôme de

Master en Informatique

Spécialité : Ingénierie des Systèmes Informatiques Intelligents

Thème

Proposition d'un profil Topical de l'utilisateur en se basant sur l'analyse des activités sociales (tweets et tags)

Présenté par :

Ahmed Dahdouh & Rym Amina Benbaba.

Devant le jury composé de :

Mme. Amel ZIANI	Professeur. Université Alger1	Président
M. Elamine Zemali	Professeur. Université Alger1	Encadrant interne
Mme. Saîda Kichou	Attachée de recherche. CERIST	Encadrant externe
Mme. Khadidja Benmessaoud	Professeur. Université Alger1	Examinateur

Année universitaire : 2019-2020

Remerciements

Avant tout, Nous remercions notre Dieu le tout-puissant de nous avoir donné la force et la patience d'atteindre notre but et d'accomplir notre travail.

Nous tenons à exprimer nos sincères remerciements à :

Mr. ZEMALI Elamine et Mme. KICHOU Saida

Pour l'encadrement qu'ils nous ont assuré et leurs précieux et judicieux conseils qu'ils n'ont cessé de nous prodiguer tout au long de ce projet, leurs confiances témoignées, sans oublier leurs qualités humaines.

Nous remercions également tous les membres de jury pour nous avoir fait l'honneur d'examiner ce mémoire, qu'ils trouvent ici l'expression de notre profond respect.

Nous adressons nos remerciements aux enseignants du département d'informatique, bibliothécaires et administrateurs, à tous ceux qui ont contribué à notre formation de près ou de loin durant notre année d'étude, et à tous ceux qui nous ont accompagnés amis et proches sans oublier tous nos collègues de 2 - ème année Mster Informatique promotion 2015

Enfin, Nous sommes reconnaissants envers nos familles qui ont été une source constante d'encouragement, de soutien et de joie

Le Binôme BENBABA Rym Amina et DAHDOUH Ahmed.

Dédicaces

« L'impossible appartient à ceux qui ne croient pas en eux »

Je dédie ce modeste travail :

À mes chers parents

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours et que dieu vous préserve en bonne santé et longue vie.

À mes deux sœurs

Pour leur encouragement et l'amour dont elle m'a entouré Je leurs exprime ici, toute mon affection et mes souhaits de succès

À mon petit frère Abderrahmane

Que dieu le protège …inchalah

Je te souhaite un avenir radieux plein de bonheur et de succès

À tous mes amis

Khaled (en particulier), Amine, Akram, Yani, Nounou …, Pour leur encouragement et leur présence à mes côtés dans les moments difficiles Je l'exprime ici, Toute ma reconnaissance et mes souhaits de réussite et de succès.

Mes enseignants de Primaire, Secondaire et Universitaire

Qu'ils trouvent dans ce mémoire ma reconnaissance et ma gratitude

Ma famille DAHDOUH

Que la paix d'Allah soit avec tous …

Ahmed

Dédicaces

De l'union « si » avec « mais » naquit enfant nommé « jamais »

« Il n'y a pas de « si » ni de « mais », il faut réussir »

À moi-même

Et parceque nulle personne ne m'a soutenu dans mes moments difficiles comme je l'ai fait, et que nulle personne ne m'a poussé en avant comme je l'ai fait. Parceque je me soutenais, je vais commencer par dédier ce travail à mon âme persévérante et mon esprit courageux.

À mes parents

Je vous remercie infiniment pour me rendre une fille aussi forte, une fille qui ne dépend sur personne. Je vous dédie mon succès.

À ma sœur Nour-El-Houda

Que nulle dédicace ne puisse exprimer ce que je la dois, l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour toi... Trésors de bonté, de générosité et de tendresse, en témoignage de mon profond amour et mes grandes reconnaissances « Que Dieu te garde ».

À ma sœur Djazia Manel

Aucun mot n'est à la hauteur pour décrire tout l'amour que je te porte. Ma petite sœur et ma source de bonheur, je me sens chanceuse d'avoir un trésor aussi précieux que toi. Je t'aime.

À mes chers frères et ma soeur

À vous Imad, Sifou, Samy et Maria mes amours, je vous dédie ce projet en signe de gratitude.

À toutes mes amies

Amira, Hadjer, Nihel, Noussaiba, Imene, Ikram, Sarah, Soussene, Raouia, Lilia, pour leur aide et leur soutien moral durant l'élaboration du travail de fin d'études.

À tous ceux dont l'oubli du nom n'est guère celui du cœur...

Rym Amina

Resumé

De nos jours, les réseaux sociaux sont devenus l'un des plus grands rassemblements de personnes en ligne, ces personnes peuvent interagir et partager des publications pour exprimer leurs airs personnels, leurs activités, ou/et leurs expériences sur plusieurs sujets sous forme des tags, des tweets, des commentaires…etc. Toutes ces données émises par chaque utilisateur, ne peuvent qu'enrichir notre connaissance sur les intérêts de celui-ci.

L'extraction du profil utilisateur à partir des tags et des tweets partagés sur les réseaux sociaux de l'utilisateur, est un domaine de recherche intéressant, qui peut servir à des fins différentes telles que dans les systèmes de recommandation dans le e-commerce, la personnalisation de moteurs de recherche, les systèmes adaptatifs dans les environnements e-Learning, etc.

Selon les approches existantes dans la littérature, l'analyse des sentiments des tweets a été effectuée dans certains travaux. Un autre axe de travail est l'analyse temporelle des tags et tweets. Cependant, rares sont les travaux qui ont considéré à la fois l'analyse des sentiments et le poids temporel pour découvrir les intérêts d'un utilisateur.

Dans ce travail, nous proposons une approche qui combine à la fois l'analyse des sentiments, le poids temporel et la classification sur des tweets enrichis par des commentaires ou des tags de questions enrichis par des tags de réponses. Ainsi, nous sommes en mesure d'extraire les sujets qui intéressent un utilisateur. Nous avons implémenté notre solution en utilisant les algorithmes de l'analyse de sentiments et la classification de texte en différentes catégories, et les techniques de topic modeling pour la génération des topics. Notre approche proposée a été validée par des expérimentations, les résultats sont assez encourageants.

Mots clés : profil utilisateur, topics modeling, tag, tweet, intérêts, les commentaires.

Abstract

Nowadays, social networks have become one of the largest gatherings of people online, these people can interact and share publications to express their personal tunes, activities, and / or experiences on several topics in the form of tags, tweets, comments ... etc. All data transmitted by each user, can only enrich our knowledge on the interests of the latter.

The extraction of the user profile from the tags and tweets shared on the user's social networks is an interesting area of research that can be used for different purposes such as in recommendation systems in e-commerce, customizing search engines, adaptive systems in e-Learning environments, etc.

According to existing approaches in the literature, the sentiment analysis of tweets was conducted in some work. Another working axis is the temporal analysis of tags and tweets. However, few works have considered both sentiment analysis and the temporal weight to discover a user's interests.

In this work, we propose an approach that combines sentiment analysis, the temporal weight and classification of tweets enriched by comments, or tags of questions enriched by tags of responses. Thus, we are able to extract the user's interests' topics. We implemented our solution using algorithms of sentiment analysis and text classification, and topic modeling techniques for generating topics. Our proposed approach was validated by experiments, the results are quite encouraging.

Keywords: user profile, topics modeling, tag, tweet, interest, comments.

ملخص

في الوقت الحاضر، أصبحت مواقع التواصل الاجتماعي واحدة من أكبر التجمعات للأشخاص عبر الإنترنت، ويمكن لهؤلاء الأشخاص التفاعل ومشاركة المنشورات للتعبير عن إيقاعاتهم الشخصية وأنشطتهم و / أو تجاربهم حول العديد من الموضوعات في شكل علامات، تغريدات، تعليقات ... إلخ. كل هذه البيانات الصادرة عن كل مستخدم يمكنها أن تثري معرفتنا على مصالح هذا الأخير.

يعد استخراج ملف تعريف المستخدم من العلامات والتغريدات التي تتم مشاركتها على الشبكات الاجتماعية للمستخدم مجال بحث مثير للاهتمام، والذي يمكن استخدامه لأغراض مختلفة مثل أنظمة توصية التجارة الإلكترونية، تخصيص محركات البحث والأنظمة التكيفية في بيئات التعلم الإلكتروني، إلخ.

وفقًا للمقاربات الموجودة في الأدبيات، تم إجراء تحليل لمشاعر التغريدات في بعض الأعمال والتحليل الزمني للعلامات والتغريدات في أخرى. ومع ذلك، فان الأعمال التي نظرت في كلا التحليلين نادرة جدا.

في هذا العمل، نقترح نهجًا يجمع بين تحليل المشاعر والوزن الزمني والتصنيف على التغريدات المثرية بالتعليقات وعلامات الأسئلة التي تثريها علامات الاستجابة. وبالتالي، نحن قادرون على استخراج الموضوعات التي تهم المستخدم. قمنا بتنفيذ حلنا باستخدام خوارزميات تحليل المشاعر وتصنيف النص إلى فئات مختلفة، وتقنيات نمذجة الموضوع لتوليد الموضوعات. تم التحقق من صحة نهجنا المقترح من خلال التجارب، وكانت النتائج مشجعة للغاية.

الكلمات المفتاحية: ملف تعريف المستخدم، نمذجة الموضوعات، الوسم، التغريدات، الاهتمامات، التعليقات.

Sommaire

Liste des figures

Liste des tableaux

Liste des abréviations

AED	Analyse et l'Exploration de Données.
AG	Algorithmes Génétiques.
ANOVA	ANalysis Of Variance.
API	Interface de Programmation dApplication.
ARS	Analyse des Réseaux Sociaux.
CTM	Correlated Topic Model.
DTM	Dynamic Topic Modeling.
EM	Expectation-Maximisation.
FC	Filtrages Collaboratives.
HDP	Hierarchical Dirichlet Process.
HLDA	Hierarchical Latent Dirichlet Allocation.
IDF	Inverse-document-frequency.
IHM	Interface Homme Machine.
LDA	Latent Dirichlet Allocation.
LLDA	Labeled LDA.
LSI	Latent semantic Indexation.
LTN	Latent Theme Network.
MG-LDA	Multi-Grains Latent Dirichlet Allocation
NLP	Natural Language Processing
OSN	Online Social Network.
PLDA	Partially labeled Latent Dirichlet Allocation.
PLSI	Probabilistic Latent Semantic Indexing.
RI	Recherche d'Informations.
SVD	Singular Value Décomposition.
Tags	Étiquetage social.
TF	term-frequency.
TFIDF	term-frequency inverse-document-frequency.
VB	Méthode Bayésienne Variationnelle.