



HAL
open science

Profil utilisateur dans les réseaux sociaux : État de l'art

Faneva Ramiandrisoa, Josiane Mothe

► To cite this version:

Faneva Ramiandrisoa, Josiane Mothe. Profil utilisateur dans les réseaux sociaux : État de l'art. Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI 2017) in CORIA 2017, Mar 2017, Marseille, France. pp. 395-404. hal-01787428

HAL Id: hal-01787428

<https://hal.science/hal-01787428>

Submitted on 7 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 18992

The contribution was presented at CORIA 2017 :

<http://www.lsis.org/coria2017/>

To link to this article URL : http://dx.doi.org/10.24348/coria.2017.RJCRI_6

To cite this version : Ramiandrisoa, Faneva and Mothe, Josiane *Profil utilisateur dans les réseaux sociaux : État de l'art.* (2017) In: Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI 2017) in CORIA 2017, 29 March 2017 - 31 March 2017 (Marseille, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Profil utilisateur dans les réseaux sociaux: Etat de l'art

Faneva Ramiandrisoa* — Josiane Mothe**

* Université d'Antananarivo, r.faneva.mahery@gmail.com

** Institut de Recherche en Informatique de Toulouse, UMR5505 CNRS

ESPE, Université Jean Jaurès, Université de Toulouse, Josiane.Mothe@irit.fr

RÉSUMÉ. Les réseaux sociaux sont de plus en plus utilisés; les utilisateurs y échangent des informations et fournissent des éléments sur leur profil. Ces données peuvent être utilisées pour modéliser un individu selon les activités qu'il réalise sur le réseau social: il s'agit du profil de l'utilisateur. Ces profils peuvent ensuite être analysés et exploités selon le domaine d'application. Par exemple, les publicités proposées sur les réseaux sociaux sont différentes pour chaque utilisateur selon les préférences de ces derniers. Cet article correspond à un état de l'art sur les informations constituant un profil utilisateur ainsi que sur la représentation des profils. Nous indiquerons également les pistes de recherche pour nos travaux dans ce domaine.

ABSTRACT. Social networks are widely used and provide more and more data describing users. This data can be used to model an individual according to his activities on the social network through his profile. Users' profiles can then be analyzed and exploited according to the application. For example, the advertisements offered on social networks are different for each user depending on his preferences. This article corresponds to a state of the art related to the information that constitutes user profiles, how user profiles are represented. We also indicate the topics of research for our future research work in this field.

MOTS-CLÉS: Réseaux sociaux₁, profil utilisateur₂, modélisation de l'utilisateur₃.

KEYWORDS: Social network₁, user profile₂, user modeling₃.

1. Introduction

La notion de "réseau social" a été introduite dans les années 1950 en sciences sociales par l'anthropologue britannique Barnes (1954) (Girard, 2012). Elle désigne une structure sociale composée d'un ensemble d'acteurs (individus, organisations, objets, etc...), des ensembles de liens et d'autres interactions sociales entre acteurs. Depuis quelques années, le terme "réseaux sociaux" évoque souvent les réseaux sociaux numériques tels que Facebook, Twitter, ou YouTube, etc. Notre travail se situe dans ce cadre. Girard (Girard, 2012) adopte le terme "média sociaux" pour qualifier globalement ces sites dont elle distingue trois groupes : les outils de publication, les outils de discussion et les réseaux sociaux numériques (RSN). Dans la suite de cet article nous utiliserons de façon indifférenciée le terme de "réseaux sociaux" et de "RSN" qui sont au coeur de nos préoccupations de recherche.

L'utilisation des réseaux sociaux a vu récemment une forte croissance. Plusieurs chercheurs se sont donc intéressés à l'analyse des données qu'ils contiennent et à leur exploitation dans différents domaines tels que la santé (par exemple pour la détection de la dépression (Kang *et al.*, 2016)) ou le marketing (par exemple la proposition de publicités par rapport aux intérêts de l'utilisateur (Thackeray *et al.*, 2008)). Les analyses (prédiction ou détection) sur les individus ou communautés se font en fonction des informations que les utilisateurs ont généré auparavant. Dans ce type d'étude, la mise en place de profils est un moyen efficace pour représenter un individu ou une communauté. Cet article est centré sur cette notion de profil dans les réseaux sociaux.

Le profil d'un utilisateur est un ensemble d'informations qui décrit les différentes caractéristiques de l'utilisateur telles que ses intérêts, ses habitudes, etc. La construction d'un profil s'appuie sur un processus d'extraction des informations utiles et fournies par les utilisateurs (Hasan *et al.*, 2013). Avec ces profils, les applications peuvent distinguer un individu d'un autre ou au contraire regrouper les individus similaires.

Les informations contenues dans le profil de l'utilisateur varient en fonction des objectifs et domaines d'application (Schiaffino et Amandi, 2009). Outre les variations des informations constituant les profils utilisateurs, les moyens d'extraction de ces informations varient aussi en fonction du domaine d'application mais surtout du type d'information. Ces informations peuvent être acquises de manière implicite ou explicite (Gauch *et al.*, 2007), (Schiaffino et Amandi, 2009). Les informations explicites sont celles qui sont fournies par les utilisateurs eux-mêmes (tels que les noms, âge, etc...); elles sont parfois non fiables (en particulier lorsque les utilisateurs fournissent de fausses données). Les informations implicites sont celles qui sont déduites à partir des données observables à propos des utilisateurs par exemple leurs actions, leurs sentiments par rapport à un produit, etc...

Dans cet article, nous dressons un état de l'art sur les profils d'utilisateur dans les réseaux sociaux. Nous étudions les contenus, représentations et utilisations les plus courants pour les profils utilisateurs dans les réseaux sociaux. Nous indiquerons également les pistes de recherche que nous envisageons de développer.

Tableau 1. Proposition de classement des principales catégories de médias sociaux (Girard, 2012)

Outils de discussion	- Forums : PhpBB - Chat (texte, audio, vidéo) : Messenger, Skype
Outils de publication	- Wikis : Wikipedia - Blogs : Typepad, Blogger
RSN ¹ de contact	- RSN ¹ généralistes : Facebook, Myspace, Twitter - RSN ¹ professionnels : LinkedIn, Viadeo, Xing
RSN ¹ de contenu	- Micro-publication : Twitter - Partage de vidéos : YouTube, Daily Motion - Partage de photos : Flickr - Partage de diapositives : Slideshare - Partage de liens, veille : Delicious - Partage de musique : Deezer

2. Contenu d'un profil utilisateur

Le profil d'un utilisateur est la représentation d'un ensemble d'informations essentielles (différent selon le domaine d'application) décrivant l'utilisateur.

2.1. Intérêts de l'utilisateur

Les intérêts d'un utilisateur correspondent à une des caractéristiques les plus fréquentes dans un profil utilisateur surtout dans les systèmes de recommandation. Les intérêts de l'utilisateur peuvent être classés en deux types : les intérêts à court terme et à long terme (Benammar *et al.*, 2002), (Tchuente *et al.*, 2013). Le profil est à long terme si l'utilisateur est toujours intéressé par le sujet et que cet intérêt ne change que très rarement (par exemple, l'utilisateur est intéressé par l'informatique). Il est à court terme si l'utilisateur est intéressé par le sujet durant une période limitée, que cet intérêt soit éphémère ou change plus fréquemment (par exemple, l'utilisateur est intéressé par le football pendant une coupe du monde).

Il est très important de trouver la méthode adéquate pour extraire les intérêts de l'utilisateur (cela peut dépendre aussi de l'objectif de l'application) à partir des réseaux sociaux dans lesquels beaucoup de données peuvent être utilisées telles que le commentaire, la notation, le partage, etc.

La base de l'extraction des intérêts est d'identifier, à partir des données des utilisateurs (cela peut être les messages, le nombre de clics, le nombre de visites, etc.), ce que l'on appelle les "concepts". Un concept peut être une entité (une organisation, une personne, un produit, etc.), un sujet (comme par exemples "musique", "festival") ou un mot-clé. Autrement dit, le concept est l'objet qui suscite l'intérêt de l'utilisateur. Une fois les concepts trouvés, l'étape suivante est de calculer leur importance en leur affectant un score. Dans la littérature, plusieurs chercheurs ont proposé différents moyens

d'extraire les intérêts des utilisateurs, et surtout de mesurer leur importance.

Dans leur travail, Abel *et al.* (Abel *et al.*, 2011b) proposent deux types de profils pour modéliser les intérêts des utilisateurs de Twitter, des profils d'utilisateurs basés sur des entités et des profils basés des sujets, tous étant construits à partir des tweets des utilisateurs. Le profil basé sur l'entité est un ensemble d'entités pondérées où le poids de chaque entité est calculé soit en fonction du nombre de tweets de l'utilisateur contenant ou se référant à l'entité donnée, soit en fonction du nombre d'occurrences d'entités dans les tweets de l'utilisateur combinés avec des articles externes. Le profil basé sur les sujets suit le même principe que celui basé sur les entités en changeant juste les entités par les sujets. Dans leurs études, ils ont utilisé OpenCalais² pour extraire les entités et les sujets. Dans leur autre travail (Abel *et al.*, 2011a), les mêmes auteurs modélisent l'intérêt des utilisateurs (c'est-à-dire le profil utilisateur) en utilisant des sujets mais également en prenant en compte les changements au fil du temps.

De même, Kapanipathi *et al.* (Kapanipathi *et al.*, 2014) créent un profil d'un utilisateur basé sur les entités en utilisant des bases de connaissances publiques (exemple : Wikipédia) pour localiser des entités. Ils tirent parti des relations hiérarchiques dans les bases de connaissances pour déduire les entités qui seront ensuite présentées sous forme de graphe d'intérêt hiérarchique.

Tchunte *et al.* (Tchunte *et al.*, 2013) proposent une méthode pour déduire les intérêts de l'utilisateur en utilisant les réseaux égocentriques et appliquant un algorithme basé sur la communauté. Ils calculent l'importance (scores ou poids) des intérêts en tenant compte des mesures sémantiques et structurelles du réseau égocentrique. L'inconvénient de cette approche est qu'un réseau éparpillé ou petit peut conduire à des interprétations erronées dans le processus de modélisation du profil de l'utilisateur. Ainsi dans leur second travail (On-At *et al.*, 2014), les auteurs améliorent leur précédente méthode en ajoutant plus de relations dans le réseau de l'utilisateur, en utilisant la technique de prédiction de liens et le "snowball sampling". Dans un autre de leurs travaux (Canut *et al.*, 2015), les auteurs intègrent la notion de temps dans leur méthode (Tchunte *et al.*, 2013) afin de prendre en compte les caractéristiques dynamiques des réseaux sociaux. Ils ont obtenu de bonnes performances en appliquant leur méthode d'extraction des intérêts des utilisateurs sur des données extraites de facebook et de DPLB (combiné avec Mendeley). Mais dans leur dernier travail (On-at *et al.*, 2016), ils ont constaté que leur méthode ne produit pas de bons résultats sur Twitter.

2.2. Comportements de l'utilisateur

Le comportement représente ce que les utilisateurs font en ligne et couvre diverses activités sociales. Par exemple, une personne va passer cinq heures par jour sur facebook et l'activité qu'elle y réalise le plus est de discuter avec ses amis.

Pour mieux comprendre les comportements sur les réseaux sociaux, nous allons

2. <http://www.opencalais.com>. OpenCalais est un outil d'extraction d'entités, sujets, événements, etc. à partir de textes.

suivre l'exemple de Jin *et al.* (Jin *et al.*, 2013) en parlant de deux aspects complémentaires : la *connectivité ou interaction*, et le *trafic d'activité*.

Connectivité et interaction . Habituellement, comme dit dans (Jin *et al.*, 2013), un réseau social est modélisé soit sous forme de graphe non orienté (exemple : graphe d'amitié, graphe d'interaction), soit sous forme de graphe orienté (exemple : graphe de followers). Cette technique de représentation est surtout utilisée sur les analyses des interactions entre les utilisateurs parce qu'il est plus facile d'analyser les interactions sociales et de caractériser ce type de comportement (connection avec d'autres, etc.). Dans le cas de la création du profil d'un utilisateur, le réseau égo-centré³ est utilisé pour représenter le comportement de l'utilisateur. Ainsi, au lieu de simplement lister tous les amis ou relations dans un profil, la représentation sous forme de graphe est mieux adaptée.

Il existe différentes façons de modéliser un réseau social en graphe :

– Graphe non orienté : Le meilleur exemple de ce modèle est le graphe d'amitié (par exemple sur Facebook), dans lequel chaque noeud représente un utilisateur et une arête représente l'amitié entre deux utilisateurs.

– Graphe orienté : Un exemple de ce modèle est le graphe d'interaction latente. Par exemple, si l'utilisateur A visite le profil de l'utilisateur B alors il existe un arc dirigé de A à B.

– Echantillonnage de graphe (*graph sampling*) : Avec la forte croissance d'utilisateurs et de données sur les réseaux sociaux, la taille du graphe permettant de les modéliser devient très grande. Des défis doivent être résolus concernant la capacité de stockage, ou le temps de calcul (par exemple pour le calcul du diamètre du graphe, les mesures de centralité du graphe, etc.). L'échantillonnage de graphe est une technique utilisée pour obtenir un graphe plus petit mais cependant représentatif de l'ensemble du graphe et sans perte de propriétés (comme par exemple la distribution des degrés) (Jin *et al.*, 2013).

Trafic d'activité. Beaucoup d'activités des utilisateurs ne peuvent pas être représentées sous forme de graphe (par exemple le temps passé sur le réseau social). D'autres types d'études s'intéressent ainsi à d'autres types de comportements (Jin *et al.*, 2013).

La *surveillance du trafic* (comme par exemple la fréquence d'accès aux réseaux sociaux, le temps total passé, la durée d'une session, etc.), la *caractérisation de la navigation* (par exemple la comparaison de l'utilisation des réseaux sociaux et des moteurs de recherche), etc.

Comme exemple sur l'analyse des trafics des activités, Benevenuto *et al.* étudient les données clickstream avec HTTP de 37,024 utilisateurs (Benevenuto *et al.*, 2009). Ils analysent les sessions des utilisateurs en considérant trois caractéristiques : *fréquence d'accès aux réseaux sociaux*, *durée totale passée sur les réseaux sociaux*, et

3. Sous-graphe centré sur l'utilisateur du graphe représentant le réseau social en entier. Par exemple le réseau égo-centré d'un utilisateur u est le graphe $G = (V, E)$ où V est l'ensemble des utilisateurs directement connectés avec u et E l'ensemble des liens.

durée d'une session. Ils ont également défini les activités des utilisateurs avec 41 activités classées en 9 groupes.

Schneider *et al.* étudient aussi des données clickstream en considérant comme caractéristiques à analyser : la popularité, les caractéristiques des sessions ainsi que la dynamique de ces sessions (Schneider *et al.*, 2009).

2.3. Connaissances, expériences et compétences

Ces trois caractéristiques sont très importantes dans plusieurs domaines comme par exemple dans la recommandation d'emplois et donc sur les réseaux sociaux qui ont accès aux études ou activités professionnelles comme LinkedIn).

La connaissance est très utilisée sur les systèmes adaptatifs tels que les systèmes éducatifs adaptatifs qui est l'un des plus étudiés dans ce domaine (Sosnovsky et Dicheva, 2010). Par exemple, la connaissance qu'un étudiant a acquise en suivant un cours en ligne pourrait être mesurée et enregistrée puis utilisée pour lui proposer d'autres cours.

Les compétences de l'utilisateur sont essentielles dans des domaines tels que la gestion des connaissances (Schiaffino et Amandi, 2009). Les compétences peuvent également être évaluées. Par exemple, étant donnée une compétence, différentes valeurs peuvent être affectées à cette compétence. Sure *et al.* (Sure *et al.*, 2000) utilisent des entiers comme valeurs (0 : aucune connaissance, 1 : débutant, 2 : intermédiaire et 3 : expert).

Contrairement aux connaissances, les expériences ne changent pas au fil du temps (Sosnovsky et Dicheva, 2010) ou changent très rarement (le plus souvent, d'autres expériences s'ajoutent). Par exemple, dans un système de tutorat, l'expérience peut être la profession de l'utilisateur, etc.

Ces caractéristiques peuvent être utilisées pour découvrir des experts dans un domaine donné ou inversement pour exclure des individus dont les connaissances, les expériences ou les compétences ne correspondent pas à une tâche particulière.

2.4. Autres éléments

Objectifs de l'utilisateur. Cette caractéristique représente ce que les utilisateurs souhaitent réaliser dans un contexte donné, il s'agit de la tâche au centre de son attention. Par exemple sur Youtube, connaître l'objectif de l'utilisateur permettrait de lui proposer des vidéos pertinentes par rapport à ses recherches. Obtenir le but ou l'objectif de l'utilisateur n'est pas une tâche triviale. L'une des techniques les plus utilisées qui permettent d'identifier l'intention d'un utilisateur à partir des tâches qu'il effectue est le "Plan recognition" (Schiaffino et Amandi, 2009). Cette technique consiste à considérer toutes les actions de l'utilisateur sur le réseau social comme une tâche, puis l'objectif est défini comme la tâche ayant l'intention calculée comme la plus élevée de l'utilisateur. Autrement dit, l'objectif est la prochaine tâche susceptible d'être réalisée, en se basant sur les tâches précédentes effectuées.

Contexte. Il existe plusieurs types de contexte selon le domaine d'application. Les différents types incluent les contextes environnementaux, les contextes personnels, les contextes sociaux et les contextes spatio-temporels (Göker et Myrhaug, 2002). Le contexte environnemental capture les entités situées à proximité d'un utilisateur telles que les objets, la température, les personnes, la lumière, etc. Le contexte personnel comprend le contexte physiologique (comme le poids, la couleur des cheveux, etc.) et le contexte mental (comme l'humeur, le niveau d'angoisse ou de stress, etc.). Le contexte social peut contenir des informations telles que les amis, les voisins, les collègues, etc., des informations qui décrivent les aspects sociaux de l'utilisateur. Enfin, le contexte spatio-temporel est une combinaison des attributs suivants : temps, emplacement, ou direction.

Emotion. Les émotions ont été récemment intégrées dans le profil des utilisateurs. La modélisation de certaines émotions des utilisateurs permet d'obtenir plus de précision sur certaines études. Par exemple, Rodrigo *et al.* (Rodrigo *et al.*, 2007) ont découvert que des émotions telles que l'ennui et la confusion amènent les élèves à changer de comportements dans les travaux (ils sont plus attirés à jouer). Cette caractéristique est largement utilisée sur les analyses psychologiques telles que dans (De Choudhury *et al.*, 2013a), (Kang *et al.*, 2016) et (De Choudhury *et al.*, 2013b), où les auteurs essayent de détecter la dépression à partir des données récoltées sur les réseaux sociaux.

Caractéristiques individuelles. Cette caractéristique est une information personnelle sur l'utilisateur et comprend principalement des informations comme l'âge, le nom, etc. Connaître ce type d'information est très utile, par exemple, des attributs tels que l'âge, le sexe et l'adresse peuvent avoir un impact sur le type de musique que les gens aiment ou écoutent.

3. Modèles de représentation d'un profil

Représentation sous forme de vecteurs. Il s'agit de l'une des représentations de profil la plus utilisée, en particulier pour sa simplicité. Habituellement, le vecteur profil correspond à un ensemble de caractéristiques avec pour chacune son poids (la valeur de la coordonnée pour une dimension). La façon de calculer le poids varie selon l'application.

Par exemple, un profil basé sur les mots-clés associés à l'utilisateur s'appuie sur une pondération TF-IDF issue du domaine de l'indexation en recherche d'information (Robertson et Jones, 1976).

Représentation sous forme de concepts hiérarchiques. Par rapport à la représentation sous forme de vecteur, l'utilisation de concepts hiérarchiques permet au système de généraliser le profil. Par exemple, un utilisateur intéressé par les jeux olympiques, est probablement également intéressé par le sport. Le niveau de la hiérarchie des concepts peut être fixé (Trajkova et Gauch, 2004) ou être dynamique (Tchunte *et al.*, 2013). Ici le concept peut remplacer les caractéristiques (intérêts, émotions,

compétences, etc.) citées dans la section 2 selon le domaine d'application, mais le plus souvent cette technique est utilisée pour représenter les intérêts.

Représentation sous forme d'ontologies. L'ontologie permet d'avoir une représentation plus sémantique en associant des liens entre les termes ou les items du profil de l'utilisateur (Trajkova et Gauch, 2004), (Hernandez *et al.*, 2007), (Sosnovsky et Dicheva, 2010).

Représentation sous forme de graphes. Cette méthode est très utilisée pour avoir une analyse des relations entre les paires de caractéristiques (intérêts, émotions, etc.) où un nœud représente une instance de caractéristiques et un arc la relation entre deux nœuds. Cette représentation permet d'analyser des relations sémantiques (Valafar *et al.*, 2009), (Jin *et al.*, 2013), et est proche de la représentation sous forme d'ontologies lorsque le graphe est un arbre. La plupart du temps, cette technique est utilisée pour analyser les relations entre les utilisateurs dans un réseau social où un nœud représente un utilisateur mais ce type de représentation peut être utilisée pour un profil d'un utilisateur, en ne considérant que le réseau égo-centré (voir la section 2.2). Il y a d'autres types de représentation d'un réseau comme les réseaux Baesiens, etc.

4. Piste de recherche

Dans nos recherches, nous souhaitons détecter des personnes dépressives à partir des informations échangées sur les réseaux sociaux. Voici quelques pistes qui nous intéressent sur le contenu d'un profil pour cette tâche :

– *Intérêts de l'utilisateur* : c'est une information utile parce que la dépression peut avoir une conséquence sur l'intérêt de la personne (exemple : une personne dépressive s'intéresse à moins de centres d'intérêt).

– *Comportements de l'utilisateur* : la dépression se traduit par des changements du comportement de la personne.

– *Emotion* : L'émotion de la personne fera partie des analyses parce que l'un des signes de la dépression est l'émotion correspondante.

– *Contexte* : on parle ici de contexte mental (comme l'humeur, le niveau d'anxiété ou de stress, etc.). Par exemple une personne dépressive connaît un trouble de l'humeur et de l'anxiété.

5. Conclusion

Dans cet article, nous avons présenté les différents types de contenu d'un profil utilisateur à partir des données provenant des réseaux sociaux ainsi que les techniques de représentation de ces profils. Ces profils peuvent être exploités soit sur les sites des réseaux sociaux eux-mêmes soit dans des applications externes. Nous nous intéresserons aussi à l'exploitation des profils utilisateurs dans nos recherches qui est de détecter les changements (signaux faibles) des utilisateurs sur les réseaux sociaux. En

premier, nous appliquerons notre recherche dans le domaine de la santé concernant la détection des personnes dépressives. Nous envisageons de participer à la tâche eRisk (<http://erisk.irlab.org/>) associée à la conférence CLEF. Dans un second temps, nous tenterons de généraliser notre méthode dans d'autres domaines.

6. Bibliographie

- Abel F., Gao Q., Houben G.-J., Tao K., « Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web », *Proc. of the 3rd International Web Science Conference*, ACM, p. 2, 2011a.
- Abel F., Gao Q., Houben G.-J., Tao K., « Semantic enrichment of twitter posts for user profile construction on the social web », *Extended Semantic Web Conference*, Springer, p. 375-389, 2011b.
- Ahn Y.-Y., Han S., Kwak H., Moon S., Jeong H., « Analysis of topological characteristics of huge online social networking services », *Proc. of the 16th inter. conf. on World Wide Web*, ACM, p. 835-844, 2007.
- Asur S., Huberman B. A., « Predicting the future with social media », *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on, vol. 1, IEEE, p. 492-499, 2010.
- Benammar A., Hubert G., Mothe J., « Automatic profile reformulation using a local document analysis », *European Conference on Information Retrieval*, Springer, p. 124-134, 2002.
- Benevenuto F., Rodrigues T., Cha M., Almeida V., « Characterizing user behavior in online social networks », *Proc. of the 9th ACM SIGCOMM conference on Internet measurement conference*, p. 49-62, 2009.
- Bontcheva K., Rout D., « Making sense of social media streams through semantics : a survey », *Semantic Web*, vol. 5, n° 5, p. 373-403, 2014.
- Canut M.-F., On-At S., Péninou A., Sèdes F., « Time-aware egocentric network-based user profiling », *IEEE/ACM Inter. Conf. on Advances in Social Networks Analysis and Mining (ASO-NAM)*, p. 569-572, 2015.
- De Choudhury M., Counts S., Horvitz E., « Social media as a measurement tool of depression in populations », *Proc. of the 5th Annual ACM Web Science Conference*, p. 47-56, 2013a.
- De Choudhury M., Gamon M., Counts S., Horvitz E., « Predicting Depression via Social Media. », *ICWSM*, p. 2, 2013b.
- Gauch S., Speretta M., Chandramouli A., Micarelli A., « User profiles for personalized information access », *The adaptive web*, Springer, p. 54-89, 2007.
- Girard A., L'intégration des médias sociaux dans les stratégies d'e-GRH : le cas du recrutement., PhD thesis, Montpellier 2, 2012.
- Göker A., Myrhaug H. I., « User context and personalisation », *Workshop proceedings for the 6th European Conference on Case Based Reasoning*, 2002.
- Hasan O., Habegger B., Brunie L., Bennani N., Damiani E., « A discussion of privacy challenges in user profiling with big data techniques : The eexcess use case », *2013 IEEE International Congress on Big Data*, IEEE, p. 25-30, 2013.
- Hernandez N., Mothe J., Chrisment C., Egret D., « Modeling context through domain ontologies », *Information Retrieval*, vol. 10, n° 2, p. 143-172, 2007.

- Jin L., Chen Y., Wang T., Hui P., Vasilakos A. V., « Understanding user behavior in online social networks : A survey », *IEEE Communications Magazine*, vol. 51, n° 9, p. 144-150, 2013.
- Kang K., Yoon C., Kim E. Y., « Identifying depressive users in Twitter using multimodal analysis », *Inter. Conf. on Big Data and Smart Computing, BigComp*, p. 231-238, 2016.
- Kapanipathi P., Jain P., Venkataramani C., Sheth A., « User interests identification on twitter using a hierarchical knowledge base », *European Semantic Web Conference*, Springer, p. 99-113, 2014.
- Mislove A., Koppula H. S., Gummadi K. P., Druschel P., Bhattacharjee B., « Growth of the flickr social network », *Proc. of the 1st workshop on Online social networks*, ACM, p. 25-30, 2008.
- On-At S., Canut M.-F., Péninou A., Sèdes F., « Deriving user's profile from sparse egocentric networks : Using snowball sampling and link prediction », *Inter. Conf. on Digital Information Management (ICDIM)*, IEEE, p. 80-85, 2014.
- On-at S., Quirin A., Péninou A., Baptiste-Jessel N., Canut M.-F., Sèdes F., « Taking into account the evolution of users social profile : Experiments on Twitter and some learned lessons », *IEEE Inter. Conf. on Research Challenges in Information Science (RCIS)*, p. 1-12, 2016.
- Robertson S. E., Jones K. S., « Relevance weighting of search terms », *Journal of the American Society for Information science*, vol. 27, n° 3, p. 129-146, 1976.
- Rodrigo M. M. T., Baker R. S., Lagud M. C., Lim S. A. L., Macapanpan A. F., Pascua S., Santillano J. Q., Sevilla L. R., Sugay J. O., Tep S. *et al.*, « Affect and usage choices in simulation problem solving environments », *Frontiers in Artificial Intelligence and Applications*, vol. 158, p. 145, 2007.
- Schiaffino S., Amandi A., « Intelligent user profiling », *Artificial Intelligence An International Perspective*, Springer, p. 193-216, 2009.
- Schneider F., Feldmann A., Krishnamurthy B., Willinger W., « Understanding online social network usage from a network perspective », *Proc. of the 9th ACM SIGCOMM conference on Internet measurement conference*, p. 35-48, 2009.
- Sosnovsky S., Dicheva D., « Ontological technologies for user modelling », *International Journal of Metadata, Semantics and Ontologies*, vol. 5, n° 1, p. 32-71, 2010.
- Sure Y., Maedche A., Staab S., « Leveraging Corporate Skill Knowledge-From ProPer to Onto-ProPer. », *PAKM*, 2000.
- Tchuente D., Canut M.-F., Jessel N., Peninou A., Sèdes F., « A community-based algorithm for deriving users' profiles from egocentric networks : experiment on Facebook and DBLP », *Social Network Analysis and Mining*, vol. 3, n° 3, p. 667-683, 2013.
- Thackeray R., Neiger B. L., Hanson C. L., McKenzie J. F., « Enhancing promotional strategies within social marketing programs : use of Web 2.0 social media », *Health promotion practice*, vol. 9, n° 4, p. 338-343, 2008.
- Trajkova J., Gauch S., « Improving ontology-based user profiles », *Coupling approaches, coupling media and coupling languages for information retrieval*, p. 380-390, 2004.
- Valafar M., Rejaie R., Willinger W., « Beyond friendship graphs : a study of user interactions in Flickr », *Proc. of the 2nd ACM workshop on Online social networks*, p. 25-30, 2009.
- Viswanath B., Mislove A., Cha M., Gummadi K. P., « On the evolution of user interaction in facebook », *Proc. of the 2nd ACM workshop on Online social networks*, p. 37-42, 2009.