

Apports des folksonomies dans le cadre d'un processus de veille : vers la prise en compte des spécificités informationnelles

Fabrice PIROLLI

fabrice.pirolli@iut-dijon.u-bourgogne.fr

Université de Bourgogne, laboratoire CIMEOS / LIMSIC - EA 4177, IUT Département Information-Communication, Boulevard du Dr Petitjean - BP 1786, 21078 Dijon cedex

RÉSUMÉ. *L'utilisation de folksonomies dans le cadre d'un processus de veille est un phénomène qui peu à peu est intégré aux pratiques de la veille sur l'Internet. Les folksonomies lorsqu'elles sont considérées comme sources d'information présentent de nombreux attraits. Leur dimension collaborative et sociale facilite l'identification de cibles et d'experts. La diversité des ressources décrites permet d'élargir le champ des recherches d'information à des domaines parfois mal couverts par d'autres outils. Cependant les informations recueillies doivent être analysées, traitées, organisées et stockées de façon spécifique. Nos propositions concernent particulièrement les aspects d'organisation et de mémorisation des informations en présentant une approche hybride mêlant folksonomies et taxonomies fondée notamment sur la notion « d'association folksonomique ».*

MOTS-CLÉS : *veille, folksonomies, taxonomies, organisation de l'information, associations folksonomiques.*

ABSTRACT. *The use of folksonomy in a strategic intelligence process is progressively included into intelligence practices related to the Internet. Folksonomies considered as information sources offer many advantages. Their social aspect help identifying targets and experts. The variety of indexed data help to widen information retrieval to documents sometimes poorly identified by other searching tools. Yet, collected informations need to be analysed, processed, organised and stored in a specific way. Our propositions are focused on information organisation and storage presenting an hybrid approach mixing folksonomies and taxonomies using the specific concept of "folksonomic association".*

KEYWORDS: *strategic intelligence, folksonomy, taxonomy, information organisation, folksonomic related terms.*

1. Introduction

Les pratiques dites d'indexation sociale accompagnent le développement des différents outils du web participatif orientés sur le partage et la diffusion d'informations. Les systèmes utilisant des folksonomies modifient peu à peu les rapports entre les internautes et les contenus disponibles sur les réseaux : ils déplacent les tâches de description, d'analyse et d'indexation précédemment dévolues aux professionnels de l'information vers l'utilisateur final qui mobilise ses propres schémas de représentations mentale et ses propres pratiques afin de mener à bien ces opérations. Les praticiens de la documentation produisent et utilisent des objets sémiotiques – index, langages documentaires – qui sont interprétés dans un mouvement de circulation entre les impératifs techniques de gestion d'un fonds et les conditions sociales de leur lisibilité [Jeanneret 06]. Or dans le cas des folksonomies ces deux aspects ne sont pas abordés selon une approche normative mais dans une démarche de mutualisation d'objectifs et de contextes d'utilisation individuels. Les folksonomies induisent également l'apparition de nouvelles modalités d'accès aux ressources informationnelles [Francis *et al.* 07]. Dans le cadre de nos travaux, nous questionnons ces nouvelles pratiques d'indexation et de recherche d'information dans le champ spécifique de la veille. Dans un premier temps, considérant les

systèmes à base de folksonomies comme des sources d'information, nous nous interrogeons sur les modalités d'intégration de ces outils dans un processus de veille. Nous proposons dans un deuxième temps un modèle d'organisation et d'exploitation des informations collectées. L'approche présentée est diachronique, afin de prendre en compte la dimension temporelle et mobilise simultanément l'usage de folksonomies, dont le lexique est composé de *tags*, et de taxonomies, exprimées au moyen de descripteurs et de liens sémantiques. Cette approche hybride nous permet de définir une relation « d'association folksonomique » caractérisant les relations significantes entre *tags* et descripteurs.

2. Les apports des folksonomies pour la veille

Le terme de folksonomie, néologisme issu de la contraction des termes anglais « folk » et « taxonomy », initialement proposé par [Vander Wal 07] désigne des pratiques d'indexation de contenu à l'aide de mots-clés, les *tags*, choisis librement par les usagers du système. Les folksonomies sont fondées sur un triptyque composé par l'objet étiqueté, l'étiquetage, et l'identité de l'étiqueteur. Celui-ci s'établit dans un environnement partagé et ouvert : l'information ajoutée par un individu est mise à la disposition de tous les membres appartenant au même réseau social. A ce titre les systèmes à base de folksonomies s'appuient sur les principes du web 2.0 et plus particulièrement sur les notions de partage, de développement collaboratif et d'intelligence collective.

2.1 Richesse des ressources et modalités d'accès à l'information

Dans les systèmes utilisant les folksonomies l'ensemble des *tags* constitue un mode de représentation du corpus considéré, indépendamment de la nature des ressources qui le composent. Ce corpus peut être composé des ressources traditionnellement mal indexées par les Systèmes de Recherche d'Information en raison de leur nature, de leur format, de leur structure : photos, vidéos, présentations assistées par ordinateurs, etc. Face à ce type de ressources, les moteurs de recherche ne fournissent pas de résultats réellement satisfaisants car les procédés d'indexation automatique ne peuvent s'appliquer. Le recours à une médiation humaine effectuée par des spécialistes (annuaires, portails,...) donne des résultats satisfaisants mais les performances restent en inadéquation avec le volume et la rapidité d'évolution des ressources à traiter. Le recours aux folksonomies constitue dans de nombreux cas le moyen le plus rapide de proposer un accès à de telles ressources. Ce type de représentation peut constituer une source d'information particulièrement riche et pertinente dans le cadre d'un processus de veille. Nous allons présenter les apports d'une telle source d'information selon trois approches complémentaires.

2.2 La dimension sociale

Les folksonomies ne peuvent être étudiées indépendamment de leurs contextes d'utilisation. Elles proposent une organisation et une représentation d'un ensemble de connaissances fondées sur la mise en commun et l'agglomération des représentations individuelles au moyen des réseaux d'échange formés par la communauté des utilisateurs. Chaque nœud, ou chaque sous-ensemble de nœuds, identifié sur ces réseaux peut potentiellement être considéré comme une source d'information ou une cible à surveiller lors d'une démarche de veille [Samier *et al* 02]. Cette notion « d'identification d'experts » constitue l'un des apports clairement identifiés des folksonomies et plus largement des réseaux sociaux et des technologies du web 2.0 en terme de veille [Poiraud-Lambert 09]. Les rapports établis entre individus sont régis par un ensemble de jeux d'influences inter-personnelles et intra-personnelles qui doivent être pris en compte pour intégrer la dimension sociale dans une démarche de veille. Dans la pratique un utilisateur subit, dans l'utilisation qu'il fait d'un système, un ensemble d'influences endogènes ou exogènes qui conditionnent son comportement. Parmi ces influences nous identifions particulièrement :

- la perception de l'outil technique et ses représentations : l'utilisateur d'un système à base de folksonomies en fera un usage correspondant à sa représentation propre des fonctionnalités de l'outil. Les tâches de description de ressources qui nécessitent un travail d'analyse et de synthèse ne seront pas menées de façon identique selon que l'utilisateur les considère comme une aide au classement, un procédé d'annotation, un mode de diffusion, etc. Les folksonomies ne peuvent donc être uniquement considérées comme un procédé d'indexation mais plutôt comme une forme particulière de métadonnées traduisant une représentation individuelle.

- l'influence du groupe : L'utilisateur de folksonomies s'inscrit consciemment dans le cadre d'une démarche collaborative. Son activité au sein de la communauté constituée par l'ensemble des utilisateurs du système façonne peu à peu son image vis à vis de ses pairs. Le comportement de chaque individu est conditionné par un ensemble de facteurs psychologiques émergeants de ses rapports au groupe : jeux d'influences (majoritaire ou minoritaire), attitude face au groupe, tendance à la normalisation, au conformisme, etc. [Blanchet 08]

- le rôle de l'interface : l'outil technique peut également directement impacter le mode de construction des représentations individuelles. Particulièrement s'il suggère des termes aux utilisateurs (*tags* utilisés les plus fréquemment par exemple) favorisant de la sorte une convergence des représentations traduisant un processus d'uniformisation.

Dans une approche centrée sur un travail de veille, la prise en compte de ces éléments nous paraît indispensable dans la mesure où ils conditionnent aussi bien le niveau de vocabulaire utilisé que le type et le mode de représentation de concepts qu'ils traduisent.

2.3 Dimension temporelle

Le temps est une variable essentielle dans une démarche de veille pour laquelle la capacité à identifier et surveiller innovations et nouveautés est souvent stratégique et capitale. Or, cette variable est fortement impactée par les usages folksonomiques. En effet, l'enchaînement des différents éléments composant la « chaîne documentaire » telle qu'elle est définie par les praticiens de la documentation est profondément modifié. Dans de nombreux cas les auteurs des ressources informationnelles représentées par les folksonomies sont directement responsables de l'indexation de leurs contenus et du choix des termes les représentant. De ce fait les folksonomies facilitent l'expression de nouveaux concepts dans un environnement donné car elles ont une très faible force

d'inertie face à la nouveauté. Dans la mesure où les folksonomies correspondent à une technologie dite *bottom-up* (dont les concepts émergent de la pratique) l'observation de l'évolution dans le temps d'un corpus de *tags* reflète l'évolution de la production documentaire et des termes utilisés pour la décrire dans un environnement donné. Cette évolution peut se traduire par l'engouement ou *a contrario* la désaffection des utilisateurs pour certains termes dont la fréquence d'utilisation varie fortement au cours du temps. Ainsi les modifications du corpus folksonomique peuvent refléter une modification de l'environnement (technique, juridique, économique).

2.4 Dimension linguistique

La critique la plus couramment formulée à l'encontre des folksonomies concerne l'absence de contrôle du vocabulaire. Les systèmes à base de folksonomies n'imposent aucune contrainte formelle, tant lors des phases de saisies que lors des phases de recherche d'information. Cette absence de contrôle est l'un des éléments qui expliquent la simplicité de l'utilisation des folksonomies. Ils constituent un terrain favorable à l'émergence et à la propagation de nouvelles acceptions, généralement liées à l'usage de termes existants ou encore à l'apparition de néologismes. Néanmoins ils exposent leurs utilisateurs à tous les problèmes sémantiques et syntaxiques rencontrés dans la langue naturelle. Dans une démarche de veille il faut donc bien prendre en compte le fait qu'un système à base de folksonomies nourrit de nombreuses ambiguïtés langagières, d'autant plus présentes s'il s'agit d'un système multilingue. Cette dimension langagière est à reconsidérer entre la nécessité d'une approche normative et l'expression libre de représentations individuelles traduisant l'innovation et l'expression de nouveaux concepts.

3. Propositions pour l'utilisation de folksonomies dans un processus de veille

Sur la base des éléments théoriques présentées en paragraphe 2 il nous semble intéressant que, dans le cadre d'une démarche de veille exploitant les folksonomies, les éléments suivants soient explicitement pris en compte :

- l'étude temporelle de l'évolution des *tags* (fréquence d'utilisation, cycle de vie,...) ;
- le comportement et le profil des utilisateurs du système ;
- l'identification et l'expression des concepts, ainsi que les relations qu'ils entretiennent, au sein du vocabulaire construit ;
- les ambiguïtés langagières et les particularités liées à la libre formulation des *tags* (polysémie, multilinguisme, néologismes,...).

Pour cela nous proposons une approche, que nous qualifions d'hybride et de diachronique, mobilisant simultanément l'usage de folksonomies et de taxonomies tout en accordant une place centrale à la variable temporelle. Nous proposons ainsi de réintroduire une démarche de médiation documentaire afin de pallier certains inconvénients inhérents au principe même de folksonomie et ainsi de proposer un mode d'organisation des informations plus propice à leur exploitation dans une démarche de veille. Les taxonomies, classifications se fondant sur des systèmes préétablis, sont définies comme « un cadre d'organisation pour des ressources numériques de toutes natures destiné à en permettre une présentation ordonnée et y donnant accès par navigation hypertextuelle » [Menon 07]. Le choix de l'utilisation de taxonomies composées de descripteurs permet d'envisager l'emploi de relations hiérarchiques et de relations d'associations, tout en laissant des possibilités d'adaptation de l'outil aux besoins spécifiques de la veille.

3.1 Les informations collectées

L'approche présentée s'appuie sur la collecte d'informations caractérisant des ensembles de ressources identifiées sur le web au moyen d'outils de partage de signets. Les informations traitées concernent à la fois les ressources en tant qu'unités documentaires, le vocabulaire constitué par l'ensemble des *tags* et les personnes partageant ces ressources (nœuds des réseaux sociaux). Les ressources informationnelles jugées pertinentes ou potentiellement pertinentes par rapports aux besoins exprimés en terme de veille sont caractérisées par un ensemble de données bibliographiques (auteurs du document, titre, adresse...) auxquelles s'ajoutent le contenu informationnel des *tags*. Afin de tenir compte des différents types d'informations que peuvent représenter les *tags* nous proposons de les qualifier contextuellement. Pour cela nous nous appuyons sur la typologie, présentée par [Roulleaux *et al.* 07] pour caractériser différents types de métadonnées. Nous distinguons ainsi informations de sujet, informations d'usage, information d'actif et informations relationnelles. Enfin chaque ressource est qualifiée par l'identifiant de la personne à l'origine de son partage. En plus des informations issues de la recherche et de la collecte de ressources sur le web, l'approche présentée implique l'utilisation de taxonomies. Le processus de veille s'appuiera à chaque étape de son déroulement sur des taxonomies organisées et contrôlées par les experts responsables du travail de veille. Lors de la phase d'initiation, les responsables de la veille constituent et organisent un ensemble de taxonomies qui constitueront le point de départ du processus de collecte. Ces taxonomies permettent de préciser l'environnement sémantique et les champs de connaissances dans lesquels s'inscriront les démarches de recherche d'information. Celles-ci sont formulées à l'aide d'outils terminologiques existants : dans la démarche de prototypage menée actuellement les taxonomies sont extraites de TermSciences, portail terminologique de l'INIST¹. Ainsi, ces taxonomies traduisent l'ensemble des représentations admises et partagées par la communauté des veilleurs. Ce cadre d'organisation sera progressivement enrichi par les folksonomies liées aux ressources collectées. La particularité de cette approche est donc de combiner dans un même système de représentations descripteurs et *tags*. Le lexique ainsi constitué est composé de termes normalisés et d'expressions non contrôlées.

3.2 La relation d'association folksonomique

Au cours des phases de recherche et de sélection de ressources, toutes ces informations sont intégrées au système de veille. Les *tags* subissent un traitement spécifique dans la mesure où ils seront qualifiés en fonction de leur nature avant d'être mémorisés. Deux situations types sont envisagées :

- le *tag* correspond à un descripteur contenu dans l'une des taxonomies : le concept qu'il représente est déjà connu ;
- le *tag* ne correspond pas à un descripteur déjà présent dans le système : le *tag* traduit un concept nouveau qui n'est pas encore représenté par les taxonomies ou bien il représente une variante orthographique ou syntaxique d'un concept déjà identifié.

¹ Portail terminologique de l'Institut National de l'Information Scientifique et Technique.
<http://www.termsciences.fr/>

Dans cette deuxième situation, et afin d'inclure le nouveau terme dans un système relationnel, nous définissons la notion de relation d'association folksonomique présentée ci-dessous en figure I. Une telle relation sera établie entre *tags* - traduisant de la sorte la co-occurrence de termes en rapport à une ressource identifiée - ou entre *tag* et descripteur. Dans les deux cas, cette association folksonomique n'est pas nécessairement de nature sémantique mais traduit l'existence d'un lien entre différents

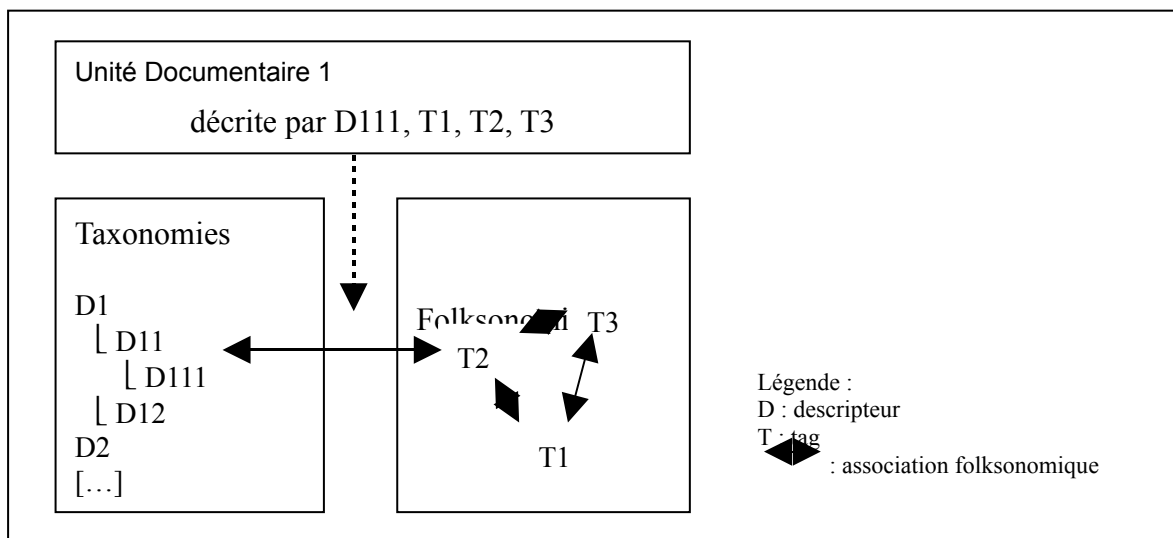


Figure I. La notion d'association folksonomique

concepts dont l'origine est fondée à la fois par le contenu et la nature de l'unité informationnelle et par les caractéristiques propres de l'auteur de la folksonomie. Toutes les relations d'association folksonomiques établies sont considérées d'égale importance. Elles expriment la part de subjectivité induite par la pratique d'une indexation sociale sur la base du triptyque fondateur des folksonomies : étiquetage, objet étiqueté et identité de l'étiqueteur.

3.3 Exploitation des informations

Les fondements techniques du système d'aide à la veille présentés ci-dessus permettent d'envisager diverses modalités de traitement pour l'exploitation des données recueillies afin de répondre aux objectifs présentés dans le paragraphe précédent : étude diachronique, prise en compte du comportement et du profil des utilisateurs, identification et représentation de concepts et enfin levée des ambiguïtés langagières.

3.3.1 Etude diachronique

L'étude diachronique de l'évolution du corpus des *tags* et des descripteurs est rendue possible par la prise en compte de la date d'indexation des ressources par les utilisateurs. Cette étude permet notamment la surveillance de :

- l'apparition de nouveaux termes qui peut traduire l'émergence de nouveaux concepts ou de nouvelles pistes de recherche ;
- la fréquence d'utilisation des *tags* ou des descripteurs traduisant les phénomènes d'engouement ou de désaffection vis à vis des concepts représentés ;
- l'apparition de nouvelles associations folksonomiques et leurs évolutions qui permettent de situer *tags* et descripteurs dans un univers relationnel enrichi de relations autres que sémantiques ;

- l'ajout de *tags* pour décrire une ressource informationnelle donnée.

Dans le cas du prototype actuellement développé, ces différents indicateurs sont présentés aux utilisateurs du système lors des phases de consultation à l'écran.

3.3.2 Comportement et profil des utilisateurs

Le comportement et le profil des utilisateurs peuvent être pris en compte du fait que chaque *tag* est systématiquement lié à son auteur. Cette approche permet l'identification de cibles ou d'experts, non seulement sur la base des unités documentaires indexées mais également en fonction de leur utilisation du système de partage de ressources. Il devient possible d'identifier les personnes à l'origine de l'énonciation de nouveaux concepts, de mener des études quantitatives sur l'ensemble des *tags* partagés ainsi qu'en fonction du contenu informationnel de ces *tags* (informations de sujet, informations d'usage, information d'actif et informations relationnelles). Ce dernier élément nous semble particulièrement pertinent dans la mesure où il traduit les pratiques individuelles des utilisateurs permettant ainsi une surveillance fondée sur des catégories d'utilisateurs. [Guy *et al.* 06] montrent que la distribution des *tags* connaît un effet de « longue traîne » traduisant le fait qu'une majorité d'utilisateurs mobilisent une faible quantité de *tags* différents alors qu'une partie réduite d'entre eux mobilise un grand nombre de *tags* pour indexer [Crepel 08].

3.3.3 Identification et représentation de concepts

Le modèle de représentation que constitue l'ensemble des descripteurs et des *tags*, ainsi que leurs relations, exprime simultanément des relations sémantiques et des relations contextualisées entre les différents termes présents. Les ressources informationnelles sélectionnées opèrent un rôle de médiation entre les représentations subjectives exprimées par l'utilisateur et l'ensemble du vocabulaire mobilisé pour indexer le corpus sélectionné. Cette contextualisation permet d'envisager différentes modalités d'accès à l'information laissant au veilleur la possibilité d'exploiter ou non cet environnement relationnel élargi par rapport à un terme, une ressource ou un individu.

3.3.4 Ambiguïtés langagières

Les problèmes posés par les ambiguïtés langagières sont traités par l'utilisation simultanée de deux modes de représentation de façon conjointe : un vocabulaire libre constitué par les folksonomies et un vocabulaire contrôlé constitué par les taxonomies. Il est ainsi possible d'intégrer les différentes variantes d'un terme : les descripteurs correspondent à des formes normalisées, les *tags* associés traduisent les différentes variantes identifiées sur le corpus traité. Ces variantes peuvent résulter de phénomènes de polysémie ou d'homographie mais également de variations d'usages de termes. L'étude diachronique permet éventuellement à terme en fonction des usages constatés de substituer un terme à un autre, opération se traduisant par la modification du statut d'un terme. Les problèmes posés par les phénomènes de synonymies ou de quasi-synonymies ne sont envisagés qu'au moyen de liens entre descripteurs et *tags*. L'évolution du modèle des taxonomies vers une organisation théaurale est envisagée dans la mesure où l'utilisation d'un thésaurus, constitué de descripteurs et de non-descripteurs, permet un traitement plus élaboré des problèmes liés à la clarification du sens des termes.

4. Conclusion

Les informations traitées par les systèmes à base de folksonomies présentent de nombreuses

caractéristiques qui répondent aux exigences d'un travail de veille : surveillance d'un milieu et de pratiques, identification de nouveaux concepts, identification d'experts, constitution de réseaux. Cependant, le recours aux folksonomies ne peut constituer qu'un élément d'une démarche plus générale. Un processus de veille sur le web ne peut s'appuyer uniquement sur l'utilisation de folksonomies sans se heurter à certaines limites inhérentes au principe d'indexation sociale. La première limite est de nature technique et concerne le taux de couverture des ressources disponibles sur le web. Les systèmes folksonomiques ne couvrent qu'une infime partie des ressources disponibles. Ils ne peuvent en aucun cas être considérés comme unique source d'information. La deuxième limite concerne la nature du travail d'indexation collaborative : celui-ci résulte d'une démarche volontaire, les ressources et les *tags* mis à la disposition des internautes correspondent à ce que l'auteur accepte de partager avec son réseau de relations. Cette notion de sélection et d'arbitrage peut sembler contradictoire à la notion d'exhaustivité qui demeure un élément essentiel dans une démarche de recherche d'information. L'approche que nous présentons dans cette communication regroupe un ensemble d'éléments visant à faciliter l'exploitation des folksonomies pour la veille. Le modèle d'organisation et de traitement des informations collectées est actuellement en phase de test au sein de notre laboratoire dans le cadre de travaux portant sur le travail collaboratif, les réseaux sociaux et le web 2.0. Les retours d'expériences qui seront collectés à l'issue de la phase de test permettront de valider ou d'invalider certaines de nos propositions.

5. Bibliographie

- Blanchet A., « La psychologie des groupes », 2ème édition, Paris : Armand Colin, Collection universitaire de poche, , 2008, 128p
- Crepel M., « Les folksonomies comme support émergent de navigation sociale et de structuration de l'information sur le web », *Réseaux* 2008/6, n°152, pp. 169-204
- Ertzcheild O., Gallezot G., « Étude exploratoire des pratiques d'indexation sociale comme une renégociation des espaces documentaires. Vers un nouveau big bang documentaire ? », *document numérique et société*, ADBS éditions, collection sciences et techniques de l'information, 2006, 344p
- Francis E. Et Quesnel O., « Indexation Collaborative et folksonomies », *Documentaliste-Sciences de l'information* 2007/1, volume 44, pp. 58-63
- Guy M., Tonkin E., « Folksonomies tidying up tags ? », *D-Lib Magazine*, Vol 12, n°1, 2006
- Jeanneret Y., « Désigner, entre sémiotique et logistique, Indice, index, indexation, sciences et techniques de l'information », édition ADBS, 2006, pp. 17-36
- Menon B., « Les langages documentaires – un panorama , quelques remarques critiques et un essai de bilan », *Documentaliste-Sciences de l'information* 2007/1, volume 44, pp. 18-27
- Morrison J., « Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web », *Information Processing and Management*, 44, 2008, pp. 1562–1579
- Poiraud-Lambert F., « De l'usage des Tags (1) - Étude de l'usage de 131 veilleurs », <http://www.tecoman.info/article-25766414.html>, 2009
- Roulleaux Dugage M., « Organisation 2.0, le Knowledge Management nouvelle génération », Eyrolles, éditions d'organisation, 2007, Paris, 258p
- Samier H., Sandoval V., « La veille stratégique sur l'Internet », *Hermes Science*, Paris : Lavoisier, 2002, 191p
- Vander Wal T., « Folksonomy Coinage and Definition », <http://vanderwal.net/folksonomy.html>, 2008