

Folksonomies

Les usagers indexent le web

Le terme de *folksonomie* est apparu récemment sur le web pour désigner le phénomène d'indexation des documents numériques par l'utilisateur. On rencontre également fréquemment le mot *tag* qui désigne en quelque sorte un mot-clé.

Le terme de *folkologie* est aussi employé, mais plus rarement. L'usage du mot *folksonomie* semble donc plus opportun. L'architecte de l'information Thomas Vander Wal a forgé ce terme en combinant la *taxinomie* (règles de classification, *taxonomy* en anglais) et les usagers (*folk*).

Olivier Le Deuff

Laboratoire Cersic
oledeuff@gmail.com

Ce phénomène ne cesse de prendre de l'ampleur avec l'avènement des nouvelles technologies du web, dites « web 2.0 », qui donnent plus de possibilités d'expression à l'internaute. Les articles et les blogs sur le sujet ne cessent de croître mais l'essentiel de ce mouvement s'exprime en langue anglaise. Cependant, les usagers français pratiquent déjà les techniques de tags via le site *Technorati* qui recense les blogs par mots-clés. Ces pratiques suscitent plusieurs questions.

Le phénomène est-il durable ou n'est-ce qu'un effet de mode ? Les systèmes de classification et les thésaurus sont-ils menacés par ces systèmes ? Débats et interrogations sur ce thème sont actuellement en cours sur le web, notamment depuis l'article un peu provocateur du consultant et spécialiste en réseaux sociaux Clay Shirky [6]. Il est intéressant de noter que ce dernier et Thomas Vander Wal n'hésitent pas à répondre aux critiques émises sur les folksonomies, ce qui crée une dynamique intéressante à suivre.

Caractéristiques

Les folksonomies constituent la possibilité pour l'utilisateur d'indexer des documents afin qu'il puisse plus

aisément les retrouver grâce à un système de mots-clés. Le concept est lié à l'accroissement et à l'accélération de la production d'informations. Le système diffère donc nettement en cela des systèmes classificatoires classiques, comme la classification décimale universelle ou la classification Dewey, qui s'inscrivent dans des processus plus longs et dont le but est d'obtenir un classement cohérent de documents physiques dont le contenu est inscrit dans la durée. Les folksonomies, au contraire, ne reposent sur aucun thésaurus, ce qui confère à l'utilisateur une liberté totale quant au choix des mots-clés. Les folksonomies sont donc initialement centrées sur l'utilisateur. Ce dernier les utilise dans un but personnel, afin d'organiser son propre système d'information.

Le paradoxe vient sans doute du fait qu'un esprit collaboratif marque le fonctionnement des systèmes dits de tags comme le fameux *Technorati*, l'un des plus importants moteurs d'indexation de blogs. Chaque utilisateur s'organise en fonction de ses besoins propres d'information et n'est plus soumis à une organisation établie par des professionnels, qui s'imposeraient à lui. Tout peut alors se trouver indexé par l'utilisateur, depuis ses favoris jusqu'à

Titulaire d'un DEA en information-communication et du Capes de documentation, **Olivier Le Deuff** est enseignant-documentaliste au collège Léon Jozeau Marigné d'Isigny-le-Buat dans la Manche. Webmaster du Guide des égarés, il a contribué aux actes du colloque Le collège unique, outil de promotion collective (IUT de Saint-Denis, 2003) et a publié un article sur l'information literacy dans L'Œil de l'ADBS (n° 6, mars 2006).

ses photos, en passant par ses messages sur son blog. L'esprit collaboratif permet également d'indexer les documents produits par les autres selon ses besoins.

L'ouverture d'esprit caractérise la folksonomie, comme le web 2.0. L'un et l'autre se rapprochent des idéaux premiers d'Internet. Les folksonomies sont donc de ce fait plus proches du monde des logiciels libres que des systèmes propriétaires. C'est pourquoi la notion de partage est souvent évoquée. Les services de photos en ligne comme Flickr, Ryia, Yahoo Photos permettent de partager ses photos, qui peuvent être indexées par d'autres usagers et réutilisées pour d'autres fonctionnalités. Les sites de stockage en ligne comme Box.net ou Overdrive vont également dans cette direction en permettant de stocker des documents personnels qui seront indexés mais qui peuvent également être partagés.

La force des folksonomies réside aussi dans le fait qu'elles ne nécessitent aucun consensus, à l'inverse des taxinomies. Il ne peut leur être reproché a priori des visions politiques ou idéologiques. Un blogueur prétend même¹ : « *Il n'y a pas de politique dans les folksonomies car le système ne comprend pas de niveau méta qui autorise les communautés d'usagers de tags à discuter de la pertinence ou non de taxinomies émergentes. Il y a seulement l'acte de taguer et le résultat cumulé et amplifié de ces tags.* »

Cette affirmation est discutable : l'emploi de mots-clés n'est pas neu-

tre, il est au contraire révélateur. En outre, il faudrait également se demander à qui bénéficie le développement de l'usage des tags.

Deux types de folksonomies

Thomas Vander Wal [8] distingue en effet deux types de folksonomies, les « étroites » (*narrow folksonomies*) et les « générales » (*broad folksonomies*). Il est possible de trouver le terme de *personomie*, qu'il faut alors rapprocher des *narrow folksonomies*.

Cette distinction est importante, car les premières sont surtout utilisées dans un objectif individuel tandis que les secondes privilégient l'aspect collectif et collaboratif du partage d'information. Ainsi les sites de partages de favoris, comme *del.icio.us* ou *Connotea*, sont plutôt des folksonomies générales puisqu'un même site peut être partagé par plusieurs utilisateurs et recevoir le même tag. Ce type de folksonomies s'appuyant sur des réseaux sociaux ne fait pas que classer de l'information et la partager. Il met en relation des usagers qui partagent les mêmes centres d'intérêts. L'utilisateur indexeur devient à son tour un peu indexé et mis en relation à la fois avec d'autres mots-clés, d'autres sites et d'autres usagers.

Qu'est-ce qu'un tag ?

Le terme qui revient le plus souvent dans les folksonomies est celui de tag qui désigne en fait un mot-clé, une catégorie ou une métadonnée. Le mot tag signifiant en anglais : étiquette de balisage, étiquetage, fléchage, marquage, voire traçage, il est possible de rencontrer des traductions comme *tagage collaboratif*. Voici la définition qu'en donnent Guy Marieke et Emma Tonkin [2] :

« *Que sont vraiment les tags ? Une définition simple serait de dire que les tags sont des mots-clés, des catégories de noms, ou des méta-*

données. Essentiellement, un tag est simplement un jeu de mots-clés librement choisi. Cependant, du fait que les tags ne sont pas créés par des spécialistes de l'information, ils ne suivent aucune indication formelle. Cela signifie que ces items peuvent être catégorisés avec n'importe quel mot définissant une relation entre la ressource en ligne et un concept issu de l'esprit de l'utilisateur. Un nombre infini de mots peut être choisi, dont quelques-uns sont issus de représentations évidentes tandis que d'autres ont peu de signification en dehors du contexte de l'auteur du tag. »

Le tag peut alors prendre toutes les formes possibles, selon le désir de l'internaute et surtout selon sa culture et sa maîtrise de la langue. Le système ne reposant sur aucun thésaurus, des mots absents du diction-

2005 : chronologie d'un succès

Janvier 2005	Technorati utilise le tracking de tags sur les blogs.
Mars 2005	— Le système de tags est un des thèmes d'actualité au sommet d'architecture de l'information à Montréal. — Article de Clay Shirky : « <i>Ontology is overrated</i> ». Dans cet entretien provocateur, il prédit la fin des hiérarchies et des classifications traditionnelles. — Yahoo achète Flickr qui permet de taguer des photos. C'est une des premières acquisitions par un grand groupe d'un site dit Web 2.0.
Août 2005	25 millions de billets seraient tagués selon Dave Sifry*, le créateur de Technorati.
Novembre 2005	Amazon accepte l'usage des tags et Google sort Google base qui permet également l'indexation d'articles par tags.
Décembre 2005	— Yahoo achète del.icio.us pour près de 30 millions de dollars. — Les folksonomies sont saluées comme une des meilleures idées de l'année 2005 par le <i>New York Times</i> .

* Dave Sifry, « State of the Blogosphere, August 2005, Part 3: Tags and tagging », <http://www.technorati.com/weblog/2005/08/37.html>

1. La plupart des citations étant en anglais, nous avons effectué la traduction en français. La citation originale se trouve ici : <http://www.test.org.uk/archives/002380.html>

naire ou des néologismes peuvent devenir des tags. De même, certains sites permettent de lier le tag à un nom de personne. C'est le cas du site Tagalag.

Le mouvement ne cesse de s'accroître mais il n'est guère aisé de chiffrer le nombre de billets de même qu'il est difficile d'évaluer le nombre de blogs.

Inconvénients et avantages

Adam Mathes [1], dans son article paru en décembre 2004, résume le problème en une phrase : « *Une folksonomie représente en même temps ce qu'il y a de meilleur et de pire dans l'organisation de l'information.* » Mathes montre ainsi que la force des folksonomies constitue également sa faiblesse. Sa souplesse d'utilisation et son adaptabilité infinie font que l'organisation de l'indexation par l'utilisateur est chaotique.

Les folksonomies suscitent par conséquent de nombreuses critiques, notamment de professionnels de l'information qui y voient parfois des similitudes avec l'emploi des métadonnées. Les folksonomies constituent également un moyen pour attirer des visiteurs sur un site par l'emploi de mots-clés détournés.

Elles peuvent dès lors être source d'info-pollution, pour reprendre l'expression d'Éric Sutter [7], vidant de leur sens de nombreux mots-clés. Les photos recevant les mots-clés de vacances sont tellement innombrables, que cela n'a plus grand intérêt si ce n'est pour celui qui les a indexés ainsi. Le fait qu'elles ne reposent sur aucun thésaurus fiable introduit de nombreuses confusions et des polysémies dérangeantes. Il en résulte une Babel sémantique qui peut finalement aboutir à un résultat inefficace et à une grande perte de temps.

Le nouveau danger est que certaines entreprises tentent de faire croire, via les systèmes de recommandation par les internautes, à la fiabilité de leurs produits. Or il n'est pas si dif-

ficile de créer de faux usagers. Cette pratique est qualifiée d'*astroturfing*.

Critiques et débats

Beaucoup de spécialistes de l'information, comme l'architecte de l'information sur le web Louis Rosenfeld [5], se montrent très critiques à l'égard des folksonomies : « *Il est facile de dire que les usagers des réseaux sociaux sont parvenus à réaliser ce que les bibliothécaires n'ont pas réussi : un moyen de faire fonctionner des métadonnées dans des collections de contenu largement distribuées et non connectées les unes aux autres. Facile mais faux : les folksonomies sont clairement convaincantes en favorisant une forme de navigation basée sur la sérendipité² qui peut être tout à fait utile. Mais elles ne favorisent pas la recherche et les autres types de navigation de manière aussi précise que des tags émanant de vocabulaire contrôlé par des professionnels.* »

Clay Shirky lui répond sur un autre blog³ : « *Le problème de l'analyse de Rosenfeld provient de son manque total de sens économique. [...] L'avantage des folksonomies n'est pas qu'elles sont meilleures que les vocabulaires contrôlés, mais qu'elles sont meilleures que rien, car les vocabulaires contrôlés ne sont pas adaptables dans la majorité des cas où un tag est requis. Construire, maintenir et améliorer un vocabulaire contrôlé revient très cher par rapport aux folksonomies, à la fois en temps de développement et pour l'utilisateur, notamment l'utilisateur amateur dans l'utilisation d'un système normé.* »

2. De l'anglais *serendipity* (1754) : « *Pure luck in discovering things you were not looking for* » (dict. Webster), en français l'art ou la faculté de trouver ce qu'on ne cherche pas.

3. La totalité de la réponse de Shirky se trouve sur ce blog : http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php

Des moyens de veille et de mesure

Cependant les folksonomies présentent un intérêt notable en ce qui concerne la traque de termes précis (*tracking*), notamment pour faire de la veille. C'est un élément intéressant pour des veilles spécialisées. Par

Le nouveau danger des folksonomies est que certaines entreprises tentent de faire croire, via les systèmes de recommandation par les internautes, à la fiabilité de leurs produits

exemple, nous utilisons le tracking de Technorati concernant le mot-clé *information literacy*. Nous pouvons ainsi consulter de nombreux blogs où ce terme est employé.

Les folksonomies sont aussi à la base des mesures de popularité. En effet, des sites comme Digg ou l'équivalent francophone Fuzz permettent de mesurer la popularité d'un message sur un blog et affichent sur leur page d'accueil les billets de blogs les plus vus du moment. C'est aussi un moyen de voir comment et à quelle vitesse se propage une information. Il existe pour cela plusieurs sites tels des agrégateurs de tags comme Guten Tag, qui affichent la popularité des tags utilisés.

Mais il ne faut pas confondre popularité et pertinence, qui sont parfois des notions opposées. De même que pour les métadonnées, ce type d'usages ne garantit pas la fiabilité de l'information et est source de ma-

nipulations évidentes pour figurer « en haut de l'affiche ». En effet, ces usages reposent aussi sur l'ego des blogueurs, qui cherchent à être lus et référencés. D'ailleurs, le site Technorati oblige les blogueurs qui veulent référencer leurs derniers billets à effectuer de régulières mises à jour de leur site par un *ping* (Packet Internet Groper)⁴ sur le site.

Les nuages de tags permettent d'afficher l'essentiel des thèmes abordés de manière visuelle sur un blog, ce qui est pratique pour voir en un seul coup d'œil l'intérêt du site. Autre avantage de ce système, l'aspect linguistique. En effet, cela permet de voir quels sont les champs lexicaux prédominants ainsi que les langues les plus employées. Au vu du succès des tags en anglais, nous pouvons nous demander si les usagers ne vont pas être d'ailleurs tentés de n'employer que des termes anglais, voire de n'écrire qu'en anglais afin d'être lus par le plus grand nombre, ce qui n'est pas anodin en termes de rentabilité publicitaire du blog.

Néanmoins ces usages constituent avec les flux RSS une alternative intéressante à la seule recherche via les moteurs de recherche. D'ailleurs, les firmes comme Google et Yahoo surveillent de près les folksonomies et n'hésitent pas à acheter les plus populaires, comme ce fut le cas pour *del.icio.us* racheté par Yahoo.

Le web sémantique du pauvre ?

Le succès des folksonomies vient sans doute du fait qu'il s'agit, en quelque sorte, d'un web sémantique du pauvre. Il s'agit aussi d'un système qui change les méthodes de navigation et de recherche, puisqu'il permet d'accéder à des ressources pertinentes qu'un moteur de recherche aurait pu ignorer. Malgré des limites manifestes, les folksonomies mettent

en évidence la possibilité de développer un système plus élaboré, auquel les internautes pourront être tentés de participer. Elles ont aussi le mérite d'exister face aux projets de web sémantique dont les applications concrètes sont encore limitées.

Il est donc probable que nous ne sommes qu'à une étape intermédiaire du système. Les micro-formats, avec *X-folk*⁵, qui permet de définir des tags dans n'importe quelle page HTML, ou bien encore l'extension pour Mozilla Firefox *Tagsy*⁶, sont parmi les projets les plus attendus.

Comment éduquer à l'indexation ou la *Tag literacy*

Une des solutions pour améliorer le système serait de former les utilisateurs à l'indexation par tag. Pour cela, il faudrait aussi que les sites permettent l'usage de plusieurs mots-clés pour définir un document. Mais il faut d'abord comprendre quels sont les mauvais tags.

Nous pouvons établir une liste de mots-clés mal indexés. Nous nous sommes inspirés des travaux du consultant en systèmes d'apprentissage *Ulises Mejias* [3] pour établir une liste de tags inefficaces : les tags mal orthographiés, les tags mal conçus, notamment les groupes de mots collés ensemble ou séparés sans utiliser l'*underscore*, ceux qui ne respectent pas les règles d'usage ou les conventions, les tags personnels n'ayant aucun intérêt collectif ou encore un tag unique qui n'apparaît qu'une seule fois dans une base de données.

Des règles de bonne indexation par tags ont été édictées :

- l'utilisateur doit penser collectivement : les tags sont certes personnels mais peuvent également être utilisés par d'autres ;
- employer le pluriel pour définir

des catégories. Le pluriel est plus approprié car la catégorie peut contenir différentes variations ;

- ne pas employer de majuscules, à moins que le mot ne puisse être compris sans ;
- utiliser l'*underscore* pour définir un groupe de mots ;
- inclure des synonymes afin d'éviter les confusions ;
- observer et utiliser les conventions d'indexation des sites et des réseaux sociaux utilisés ;
- contribuer à ce que les efforts d'indexation soient efficaces en collaborant et en ajoutant des tags à d'autres ressources.

Outre ces règles, des sites comme *Tagyu* tentent de mettre de l'ordre, en conseillant l'utilisateur sur les tags à utiliser, par analyse des mots-clés utilisés par d'autres.

Il s'agit aussi
d'un système
qui change les méthodes
de navigation
et de recherche,
puisque'il permet
d'accéder à des ressources
pertinentes qu'un moteur
de recherche
aurait pu ignorer

Cet ensemble de règles peut être utile. Toutefois, contribuer à une indexation efficace implique non seulement des compétences mais aussi du temps. C'est ce que souligne *Ian Davis*, spécialiste en bibliothéconomie⁷. Selon lui, l'indexation par les usagers semble moins coûteuse en

4. Outil permettant de vérifier si un serveur répond, et donc s'il est accessible par le réseau.

5. Vous pouvez retrouver plus d'informations sur le projet : http://thecommunityengine.com/home/archives/2005/05/xfolk_entry_04.html

6. <http://tagsy.com/>

7. *Ian Davis*, « Why tagging is expensive », in *Silkworm blog*, http://silkworm.talis.com/blog/archives/2005/09/why_tagging_is.html

temps, mais c'est le temps passé à retrouver l'information qui s'accroît, contrairement aux systèmes d'informations hiérarchisées, où le coût d'indexation par les professionnels est plus élevé, mais la recherche d'informations facilitée pour l'utilisateur. Partant de l'adage « *le temps, c'est de l'argent* », il en conclut que le système de tag devient dès lors onéreux.

Il émet même l'hypothèse que, dans cinq ans, les folksonomies seront petit à petit abandonnées pour un retour à la médiation des bibliothécaires et de Google (*sic*). On peut trouver cette vision tout aussi exagérée que celle de Shirky.

Conclusion

Les folksonomies s'inscrivent donc dans une « redocumentarisation » [4] qui voit l'utilisateur conférer un sens au document, à partir de l'usage qu'il souhaite en faire. Cela entraîne des mutations qui peuvent laisser la place aussi bien à une nouvelle souplesse documentaire qu'à des manipulations.

Comme l'écrit Roger T. Pédaque [4], « *les bouleversements induits par les nouveaux usages du web affectent autant la valeur attribuée aux contenus (crédit, autorité, représentativité) que les modes de médiation eux-mêmes (conditions spatio-temporelles de l'interaction, brouillage des rôles et des sphères « public/privé », camouflage des identités, rupture dans les genres, les discours et les usages, etc.) On se trouve désormais confronté à un jeu de rôles dont les règles changeraient en permanence, à l'insu*

même des acteurs qui tirent profit de cette permissivité ou subissent les désagréments des transgressions. »

On peut sans doute parier sur la coexistence des différents systèmes d'indexation, par les usagers et par des professionnels: les deux opérations ne répondent pas toujours aux mêmes objectifs. Les folksonomies présentent de forts avantages pour

Les folksonomies
présentent de forts
avantages pour l'accès
à de l'information rapide,
voire dans le cas
de la veille, tandis que
l'intérêt de l'indexation
professionnelle demeure,
pour sélectionner
de l'information
à caractère plus durable

l'accès à de l'information rapide, voire dans le cas de la veille, tandis que l'intérêt de l'indexation professionnelle demeure, pour sélectionner de l'information à caractère plus durable.

Dans ce contexte, les folksonomies peuvent nous permettre d'envisager sur la Toile des systèmes de catalogues hybrides, permettant les deux types d'indexation.

Mai 2006

RÉFÉRENCES BIBLIOGRAPHIQUES

1. **MATHES**, Adam. — *Folksonomies. Cooperative Classification and Communication Through Shared Metadata*.
<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
2. **MARIEKE**, Guy, **TONKIN**, Emma. — « Folksonomies. Tidying up Tags? ». — *Dlib Magazine* Volume 12, Number 1, 2004
<http://www.dlib.org/dlib/january06/guy/01guy.html>
3. **MEJIAS**, Ulises Ali. — *Tag literacy*. — 2005
http://ideant.typepad.com/ulises_mejias.html
4. **PÉDAUQUE**, Roger T. — *Document et modernités*. — 2006
<http://rtp-doc.enssib.fr/IMG/pdf/Pedauque3-V4.pdf>
5. **ROSENFELD**, Louis. — *Folksonomies. How about metadata ecologies*. — 2005
http://louisrosenfeld.com/home/bloug_archive/000330.html
6. **SHIRKY**, Clay. — *Ontology is Overrated; Categories, Links and Tags*. — 2005
http://www.shirky.com/writings/ontology_overrated.html
7. **SUTTER**, Éric. — « Pour une écologie de l'information ». — *Documentaliste - Sciences de l'information*, vol. 35, n° 2, 1998, p. 83-86.
8. **VANDER WAL**, Thomas. — *Explaining and Showing Broad and Narrow Folksonomies*. — 2005
<http://www.vanderwal.net/random/entrysel.php?blog=1635>