

SESSION 2014

---

**CAPES  
CONCOURS INTERNE  
ET CAER**

**Section : DOCUMENTATION**

**ÉPREUVE À PARTIR D'UN DOSSIER THÉMATIQUE**

Durée : 5 heures

---

*L'usage de tout ouvrage de référence, de tout dictionnaire et de tout matériel électronique (y compris la calculatrice) est rigoureusement interdit.*

*Dans le cas où un(e) candidat(e) repère ce qui lui semble être une erreur d'énoncé, il (elle) le signale très lisiblement sur sa copie, propose la correction et poursuit l'épreuve en conséquence.*

*De même, si cela vous conduit à formuler une ou plusieurs hypothèses, il vous est demandé de la (ou les) mentionner explicitement.*

**NB :** *La copie que vous rendrez ne devra, conformément au principe d'anonymat, comporter aucun signe distinctif, tel que nom, signature, origine, etc. Si le travail qui vous est demandé comporte notamment la rédaction d'un projet ou d'une note, vous devrez impérativement vous abstenir de signer ou de l'identifier.*

**Tournez la page S.V.P.**

## **Titre du dossier : Les folksonomies**

A partir de ce dossier thématique comportant 5 documents, vous devez :

1. rédiger une note de synthèse dégageant une problématique d'ensemble (3 pages maximum) ;
2. développer une réflexion personnelle sur l'utilisation pédagogique, par le professeur-documentaliste, de la folksonomie ;
3. élaborer pour le document 3 la référence bibliographique suivie des éléments d'analyse.

Les références bibliographiques de ces documents seront établies, en application des normes en vigueur, en renseignant pour chacun d'eux, les zones pertinentes du bordereau de saisie.

Mots clés : indiquer exclusivement dans le champ du bordereau prévu à cet effet les mots clés qui permettront une recherche efficiente en langage naturel.

Le document 3 fera l'objet d'un résumé indicatif de 50 mots.

*Règles pour le comptage des mots :*

*Les chiffres : une date = 1 mot (ex. 2007 = 1 mot) ; un pourcentage : 50 % = 2 mots*

*Les sigles : CNDP = 1 mot. Il est déconseillé d'utiliser des sigles peu connus dans l'éducation nationale.*

*Les articles, même élidés : 1 mot.*

*Les mots composés avec un trait d'union (ex. sino-soviétique) = 1 mot, mais c'est-à-dire = 4 mots.*

Document 1

BiblioDoc. Francophonie. Qu'est-ce que la folksonomie ?

Document 2

Sylvie Dalbin. Séminaire folksonomies et tagging au Cnam. Des balbutiements du Web 2.0 à nos jours, premiers bilans interdisciplinaires.

Document 3

Olivier Ertzscheid. Culture documentaire et folksonomie : l'indexation à l'ère industrielle et collaborative.

Document 4

Web 2.0 : le web sémantique.

Document 5

Anne Delannoy. Le « Social bookmarking » : une pratique incontournable.



Actualités

Portail

- > [Bibliothèques](#)
- > [Associations professionnelles](#)
- > [Formation](#)
- > [Revue et ressources électroniques](#)
- > [Dossiers](#)

AIFBD

FIAB / IFLA

Rechercher sur ce site

ok

Design du site : Allioo

> [Liens utiles](#) > [Liste de discussion](#) > [AIFBD](#) > [Traductions](#) > [Accueil](#)

Portail

## Folksonomie

[Retour au menu principal](#)

[Page précédente](#) / [Page suivante](#)

## Qu'est-ce que la folksonomie ?

### Définition

Jusqu'à tout récemment, la création de métadonnées était réalisée

- soit par les professionnels de l'information travaillant avec des normes incompréhensibles pour l'utilisateur et impossibles à appliquer sur l'immensité des documents disponibles sur le Web. Ces métadonnées sont reconnues de qualité mais ont un coût prohibitif en termes de temps et d'effort ;
- soit par les auteurs-mêmes des documents électroniques. Il existe là un problème d'expertise. L'auteur n'est pas un expert en catégorisation ; de plus certains peuvent volontairement mettre des étiquettes incongrues, inadéquates, etc.

Dans les deux cas, les utilisateurs des documents sont exclus du processus de création de métadonnées. Désormais, les défenseurs du web 2.0 proposent une troisième voie : la création des métadonnées par l'utilisateur, en fonction de ses besoins particuliers, mais qu'il partage avec la communauté. Le terme de folksonomie vient de « folk » et « taxonomy » ; la taxonomie du peuple. Le terme a été proposé par Thomas Vander Wal en 2004 pour désigner la catégorisation réalisée par les internautes lorsqu'ils étiquettent (tag) des ressources Web sur des sites de référence comme Del.icio.us (<http://del.icio.us/>), Furl (<http://www.furl.net>), Frassle (<http://www.frassle.org>), Yoono (<http://www.yoono.com>), etc. C'est donc une classification décentralisée et spontanée qui repose sur l'appréciation de chacun contrairement aux classifications pré-établies comme celle des répertoires. L'utilisateur décrit lui-même une URL à l'aide de mots-clés. Ses étiquettes sont proposées à la communauté via les sites de références. Si ses étiquettes sont utilisées par d'autres, elles deviennent populaires. Ainsi, lorsque des étiquettes ont obtenu un certain consensus dans la communauté, le système les suggère aux utilisateurs. Quand une personne étiquette une ressource, elle voit combien de personnes ont étiqueté cette ressource, qui a étiqueté cette ressource et peut finalement regarder toutes les ressources qu'une personne en particulier a étiquetées. C'est ainsi un moyen de trouver de nouvelles ressources intéressantes et de voir quelles personnes partagent les mêmes centres d'intérêt. Les folksonomies créent en ce sens des communautés. Les étiquettes ne sont pas reliées entre elles. Mais certains logiciels permettent de créer des hiérarchies des étiquettes. Certains systèmes permettent également de créer un groupe d'étiqueteurs. Cela peut être utile pour un projet réalisé par un groupe particulier. Il existe notamment des sites d'étiquetage spécifiques aux chercheurs et professeurs universitaires <http://www.connotea.org> et <http://www.citeulike.org>.

### Comment procéder ?

Il suffit de créer un compte sur les systèmes tels que Del.icio.us. L'utilisation est gratuite. Seul un nom et un mot de passe sont requis. Une fois enregistré, l'utilisateur télécharge une barre d'outils permettant de créer les étiquettes et de visualiser ses ressources étiquetées. Lorsque l'utilisateur veut ajouter à sa liste une ressource particulière, il l'ajoute au système via sa barre d'outils, puis il lui assigne les étiquettes qu'il souhaite pour décrire la ressource. Les étiquettes ne sont pas obligatoires ; certaines sont proposées, issues à la fois des précédentes personnes ayant indexé la même ressource et des étiquettes déjà utilisées par l'utilisateur. Il suffit de cliquer sur ces étiquettes proposées pour les assigner à la ressource à étiqueter. Les étiquettes nouvellement enregistrées sont donc celles de l'utilisateur en particulier, mais sont désormais disponibles pour la communauté. Les étiquettes reflètent les besoins de leurs utilisateurs. L'étiquetage peut ainsi être utilisé au lieu des favoris habituellement utilisés, mais qui deviennent ingérables dès lors que les sites référencés deviennent nombreux. En anglais, on parle de Bookmarking. L'utilisateur peut aussi chercher les étiquettes par sujet ou bien par le nom d'une personne en particulier. Il est possible d'indexer pareillement tous les types de documents (audio, vidéo, photo, texte)

### Forces et faiblesses de la folksonomie

Les forces de la folksonomie sont indéniables. En indexant ainsi les ressources, les personnes leur donnent du sens, les rendent plus faciles à retrouver et, par extension, rendent plus facile la découverte de ces ressources. La folksonomie est basée sur le langage quotidien, le langage de chacun ; c'est là sa principale force. Elle reflète les choix

des utilisateurs en termes de diction, terminologie et précision. De plus, La folksonomie a une force incroyable concernant la découverte. Balayer les documents référés par les différentes étiquettes qui ont précédemment été utilisées pour étiqueter un document, est un moyen efficace pour trouver des documents que l'utilisateur n'aurait pas trouvé autrement. Les sites de folksonomie permettent une exploration plus ouverte et hasardeuse du contenu que les moteurs de recherche. De plus, le vocabulaire évolue parfois rapidement et la folksonomie permet de refléter en temps réel cette évolution. La folksonomie ne requiert aucune formation préalable, elle ne coûte rien, ni financièrement ni cognitivement. Le succès des folksonomies est attribuable au besoin d'indexer les nombreux documents disponibles via la Web, et à la démocratisation de cette indexation. Ainsi, l'étiquetage des sites par les utilisateurs est la méthode la plus populaire pour enregistrer l'information électronique désirée pour un accès et une utilisation ultérieurs. Contrairement aux favoris, les étiquettes sont enregistrées sur le système Web, et non sur un poste de travail en particulier. L'utilisateur peut donc y avoir accès là où il se connecte. Mais la folksonomie a aussi ses faiblesses. Elle est en prise avec la synonymie, la polysémie, le niveau de spécificité et de généralité des étiquettes et ne propose pas de relations entre les étiquettes. Les acronymes sont aussi une source d'ambiguïté certaine (ONF est l'office national du film au Québec, mais l'office national des forêts en France). Les systèmes sont insensibles à la casse des lettres ; ce qui peut poser problème lorsqu'un acronyme est un mot existant (exemple : AMI = aide internationale médicale ou aide aux musiques innovatrices ou ami). Souvent on retrouve plusieurs formes d'un même mot, pluriel et singulier notamment, ce qui rend les choses difficiles. De plus, les systèmes comme Del.icio.us ne permettent pas toujours de créer des étiquettes contenant plusieurs mots. Ainsi, étiqueter un document « construction sociale » serait compris comme « construction » et « sociale », ce qui n'a aucun sens. Les personnes utilisent alors des trucs pour contourner ce problème, comme d'accoler les mots d'une expression « constructionsociale », ce qui ne veut rien dire non plus. Il n'existe aucune règle de syntaxe pour lier plusieurs mots. Ainsi histoire philosophie et philosophie histoire ne signifie pas la même chose. Les étiquettes en différentes langues sont utilisées, ce qui peut là encore poser problème de sens (« pain » en anglais signifie « douleur »), outre la multiplication des étiquettes. Enfin, les folksonomies acceptent les usages incorrects de la langue. Par exemple, le terme archéologie est parfois utilisé pour décrire les dinosaures et les microbes primitifs.

### Usages actuels

L'étude de Brooks et Montanez montrent que les étiquettes aident les utilisateurs à regrouper les documents ou sites en larges catégories (2006). Les plus largement utilisées sont des mots-clés, variables selon le niveau de spécificité décrivant le contenu. Certains documents sont étiquetés selon le genre ou la forme. Mais les étiquettes sont particulièrement efficaces dans la recherche d'information non spécifique, donc pour le balayage plutôt que pour la recherche. Ce qui aide les utilisateurs est de ne pas avoir à se creuser la tête pour trouver le mot-clé adéquat et de pouvoir utiliser les suggestions proposées, notamment pour les personnes pour lesquelles l'anglais est une seconde langue (Sinclair et Cardew-Hall, 2007).

### Améliorations attendues

Les systèmes de folksonomie ont donc du pain sur la planche pour proposer des synonymes, proposer d'autres étiquettes, proposer un groupe d'étiquettes, élargir le droit des personnes à étiqueter dans le cas des folksonomies étroites, etc. de manière à éviter les écueils de l'indexation non-contrôlée. Par exemple, Fac.etio.us propose aux utilisateurs d'organiser leurs étiquettes en les regroupant par facettes de lieu, technologie, ou attributs. La formation des utilisateurs est une autre voie à explorer. Certains sont optimistes : la folksonomie aboutit à un consensus ou tout du moins une stabilisation du vocabulaire dans la mesure où des étiquettes sont proposées à l'utilisateur pour indexer un document particulier ; l'étude de Golder et Huberman le montre (2006). À mesure qu'un site est étiqueté, l'ensemble des étiquettes utilisées pour le décrire devient stable et est utilisé par les utilisateurs. Cette stabilisation résulte de l'imitation des utilisateurs entre eux (utiliser les étiquettes déjà utilisées par les utilisateurs précédents) et de la volonté de partage des connaissances chez les utilisateurs. Si l'utilisateur n'a pas l'expertise d'un professionnel de l'information, la collaboration entre les utilisateurs permet d'atteindre une catégorisation de qualité. Russell propose lui un système qui reconnaisse l'autorité (l'expertise) de la personne qui étiquette. C'est-à-dire que les personnes qui étiquettent sont étiquetées à leur tour pour être reconnue dans leur expertise (2005).

## Les applications de la folksonomie en bibliothèque

<http://www.librarything.com> est un service qui permet aux utilisateurs d'étiqueter leurs propres documents, incluant les données bibliographiques d'Amazon.com et/ou des catalogues de bibliothèques. Les personnes étiquettent leurs documents de manière à les retrouver plus facilement et pour collecter des items similaires sous la même étiquette. Les utilisateurs peuvent ainsi découvrir quels sont les autres personnes qui ont le même document. LibraryThing démontre que les folksonomies et les vocabulaires contrôlés peuvent coexister et se compléter pour aider l'utilisateur à trouver ce dont il a besoin. À tel point qu'une folksonomie peut être un point de départ pour développer un vocabulaire contrôlé propre à une organisation, pour un Intranet par exemple. La folksonomie est un domaine en pleine expansion dans les recherches en sciences de l'information. Ses défenseurs proposent aux bibliothèques plusieurs applications :

1. Les étiquettes assignées par les usagers aux documents de la bibliothèque pourraient être un plus pour les usagers qui ne comprennent pas toujours l'indice de classification et l'indexation matières. Cela permettrait aussi de combler les manques des vocabulaires contrôlés (comme « drame » ou « action » pour les genres de films qui sont impossibles avec LCSH). Par contre, si l'on permet aux

usagers d'étiqueter le catalogue, il faut un modérateur qui vérifie la pertinence des étiquettes. Mais ce peut être aussi les bibliothécaires qui proposent des étiquettes sur le catalogue. Voir en anglais le Librarians' Internet Index (<http://li.org>), ou C&RL News (<http://www.ala.org/ala/acrl/acrlpubs/crlnews/internetresources.htm>).

2. Une formation peut être offerte aux usagers sur la folksonomie, notamment dans les bibliothèques universitaires et bibliothèques spécialisées.
3. La folksonomie peut être utilisée par le personnel de la bibliothèque pour organiser leurs ressources de travail, les ressources pour la référence, créer des groupes aux intérêts communs. La folksonomie est en effet un moyen efficace pour aider le travail collaboratif et partager les connaissances dans une organisation.
4. La folksonomie est aussi un moyen de chercher la littérature grise.

Date de mise à jour : **mardi 12 mai 2009**



[Nous joindre](#)

■ Sylvie Dalbin

## Séminaire folksonomies et tagging au Cnam

Des balbutiements du Web 2.0 à nos jours, premiers bilans interdisciplinaires

Les folksonomies <sup>1</sup> et l'activité d'indexation de ressources par les usagers deviennent aujourd'hui objets d'étude pour différentes communautés de chercheurs. Dans ce contexte, la journée du 26 mars 2010, organisée par Alexandre Monnin et Évelyne Bourdoux, de l'équipe de recherche Dicen <sup>2</sup> du Centre national des arts et métiers, visait à « amorcer un questionnaire et un dialogue entre [ces] différentes communautés scientifiques pour créer des passerelles entre les approches ». Sujets – et querelles ! – de laboratoires se sont mêlés au cours de cette journée passionnante.

L'ambiance était sérieuse dans cet amphithéâtre du Cnam bien rempli par des chercheurs et de nombreux praticiens, mais qui aurait pu être bondé avec des étudiants. Il est en effet toujours étonnant que l'on ne profite pas de ces journées pour amener – au moins pour quelques interventions – les étudiants en Infodoc afin de leur faire partager ce moment privilégié avec ceux qui deviendront leurs collègues, mais aussi pour leur permettre d'expérimenter ce type d'événement.

La journée était organisée en deux temps : une matinée réservée à des exposés théoriques de chercheurs, et une après-midi dédiée à des présentations de réalisations par d'autres chercheurs.

Fabien Gandon, qui fait toujours l'unanimité lors de ses présentations, démarrait avec une étude sur les ontologies informatiques ou, pour être plus juste, une étude mettant en exergue le rapprochement des ontologies et des folksonomies. Dès cette première intervention, on comprenait que la réconciliation va passer par un double processus : « folksonomiser » les ontologies et « ontologiser » les tags !

## Les techniques informatiques au service d'un dispositif social

C'est ainsi qu'une grande part des travaux ou systèmes présentés exploitent des ontologies informatiques ou, en tous les cas, des structures qui s'en approchent fonctionnellement pour « formaliser » les tags. Le système de tags de programmation (ou *machine tags*) de Flickr <sup>3</sup>, présenté par Patrick Peccatte, propose d'injecter, au milieu des tags, d'autres tags structurés par des micro-schémas tels que des éléments de Dublin Core associés à un vocabulaire contrôlé. L'ontologie informatique NiceTag <sup>4</sup>, présentée par Alexandre Monnin, capture les tags et les documente. Le tag ne se promène plus tout seul : il est accompagné de son identifiant et de son système de référence (ici du couple auteur-date), le système pouvant fournir lui-même automatiquement ces informations. Dans le cas des *machine tags*, le choix du tag est contrôlé puisque issu d'un vocabulaire choisi et structuré par des éléments d'un schéma, mais il se libère de son système source pour vivre sa vie avec la ressource à laquelle il reste associé, et surtout avec d'autres tags non contrôlés. Dans NiceTag, le choix du tag reste libre, mais on documente sa production en lui fournissant un peu de contexte de production. Une constante de ces interventions – et qui nous change du « tout automatique » ou du « tout humain » – est le déploiement de techniques informatiques au service d'un dispositif social.

Freddy Limpens, doctorant, centre son travail sur l'établissement de synergies entre des calculs automatiques et des contributions faites par des spécialistes de divers domaines à l'Ademe (Agence de l'environnement et de la maîtrise de l'énergie), en cherchant à résoudre les problèmes liés aux variations linguistiques ou sémantiques des tags et ceux liés au coût des ontologies. Le prototype traite les divergences entre les tags proposés, en leur donnant un statut de « candidat-concept ». Le système sociotechnique s'appuie tout à la fois sur des outils techniques structurants (SKOS ou Simple Knowledge Organisation System : ontologie légère dérivée du format de données des thésaurus) <sup>5</sup>, sur des automates suggérant aux utilisateurs des rapprochements entre concepts-tags, mais aussi sur des processus (une gestion de cycle de vie de la folksonomie devenue structurée) et sur une organisation entre différents acteurs avec, par exemple, un « ReferentUser » qui offre un point de vue consensuel de référence sans exclure les points de vue divergents. Tout à la fin de l'intervention, on apprend que les « documentalistes offrent leurs ressources » (un thésaurus), gèrent et récoltent ces candidats pour enrichir leur thésaurus. Le dispositif est en définitive tripartite. Reste à traiter dans la suite du projet la question essentielle des outils de gestion de ces vocabulaires. Cette intervention, la seule qui se déroule dans un contexte d'entreprise, marie des systèmes de concepts non contrôlés (tags) et contrôlés (thésaurus), une grande diversité d'acteurs (utilisateur, utilisateur référent, professionnel de l'infodoc et automates) et un processus s'articulant autour des usages.

Fabula <sup>6</sup>, riche site internet construit et alimenté par « des chercheurs [...] s'intéressant à l'articulation entre théorie et histoire littéraires » n'utilise pas, selon Alexandre Gefen, de vocabulaires contrôlés, mais des mots-clés ou tags. Il est intéressant de noter ici l'usage des mots « tags » ou « mots-clés » pour désigner, au sein d'une communauté de spécialistes, des mots-clés qui ne sont pas imposés par des personnes extérieures à la communauté... mais qui le sont par la communauté elle-même, car peut-on vraiment dire que le vocabulaire utilisé par une communauté aussi restreinte que celle des chercheurs francophones en théorie et histoire littéraires n'est pas « contrôlé » ? Dans un contexte où des experts parlent aux experts, il est souvent peu utile de déployer un vocabulaire plus contrôlé qu'il ne l'est déjà par nature. Mais peut-être peut-on penser que l'indexation n'est pas contrôlée ? Une fois encore, le dispositif semble dans la réalité très contrôlé : le vocabulaire déjà validé est suggéré à l'utilisateur qui alimente le site avec un outillage simple, mais, en back-office, quelqu'un « réorganise » ces indexations, c'est-à-dire corrige, supprime ou fusionne les mots-clés des utilisateurs. Un fonctionnement qui reporte après coup le nettoyage d'une indexation non contrôlée. Ces quelques remarques ne retirent rien à l'intérêt du travail fourni et à l'intelligence de la conception d'un système qui allie automatisme et interventions humaines.

## Un désaccord entre communautés scientifiques

Assez rapidement dans la journée, nous avons pu saisir le désaccord entre deux communautés scientifiques, désaccord exposé par Manuel Zacklad. Il repose sur le postulat suivant : deux termes de deux langues culturelles ou de spécialités différentes correspondent-ils ou non au même concept ? Ainsi, « Car » et « Voiture », selon Manuel Zacklad, renvoient à des notions différentes. Dans cette communauté, le programme scientifique du web sémantique, qui énonce que « pour partager des informations, il faut remonter à des degrés de formalisation qui nous permettraient d'atteindre des concepts et ensuite de faire des inférences logiques », est réfuté. Le sens d'un terme est analysé comme dépendant à la fois du contexte et de la situation. Le web sociosémantique s'appuie ici sur deux fondamentaux : la cohabitation de plusieurs systèmes d'organisation des connaissances (SOC) suivant les différents points de vue, plutôt qu'une approche par une vision globalisante, et la notion de ROI (recherche ouverte d'information), correspondant à une assistance à la recherche, à l'intersection de la requête (la recherche vue du web sémantique), de la fouille et de la navigation.

La recherche scientifique, c'est aussi l'évaluation et l'analyse des objets de recherche, ici des folksonomies. L'étude exploratoire comparative de folksonomies scientifiques, menée par Dominique Besagni, Cécilia Fabri, Claire François et Évelyne Broudoux, fournit d'intéressantes données sur ces pratiques, malgré les grandes difficultés liées au terrain d'étude (mobilité forte des dispositifs de folksonomies).

Cette journée s'est achevée sur une intervention d'Olivier Le Deuff, qui centra son approche sur les réseaux, les aspects sociaux des tags, plutôt que sur le tagging lui-même (le « folklore de la folksonomie »), et l'usage « mémoriel » (hypomnemata numériques), individuel ou collectif du tag.

L'articulation entre travaux scientifiques et réalisations concrètes, toujours délicate à réaliser, nous semble avoir été un élément-clé de la réussite de cette journée, dont on espère la reconduite.

1. « 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. » Olivier Le Deuff, « Folksonomies », *BBF*, 2006, n° 4, p. 66-70. (retour)

2. Dicen : Dispositifs d'information et de communication à l'ère numérique. <http://dicen.cnam.fr/> (retour)

3. <http://www.flickr.com> (retour)

4. <http://www.slidshare.net/aamonnz/> (retour)

5. [http://fr.wikipedia.org/wiki/Simple\\_Knowledge\\_Organisation\\_System](http://fr.wikipedia.org/wiki/Simple_Knowledge_Organisation_System) (retour)

6. <http://www.fabula.org> (retour)

**Notice bibliographique :**

Dalbin, Sylvie, « Séminaire folksonomies et tagging au Cnam », *BBF*, 2010, n° 5, p. 89-91  
[en ligne] <<http://bbf.enssib.fr/>> Consulté le 01 septembre 2013

Le Bulletin des bibliothèques de France est publié par l'École nationale supérieure des sciences de l'information et des bibliothèques  
Enssib - 17-21, boulevard du 11 novembre 1918 - 69623 Villeurbanne Cedex  
© enssib - [www.enssib.fr](http://www.enssib.fr)



olivier.ertzscheid@univ-nantes.fr

Auteur d'une thèse sur l'hypertexte et les nouveaux modes d'accès aux connaissances sur les réseaux, **Olivier Ertzscheid** est maître de conférences en sciences de l'information et de la communication, actuellement en poste à l'université de Nantes. Ses travaux de recherche portent notamment sur le processus de veille et de recherche d'information, sur les nouvelles modalités d'organisation et d'accès aux connaissances sur le Net ainsi que sur l'émergence de pratiques documentaires dédiées. Depuis cinq ans, il tient son blog d'enseignant-chercheur, [affordance.info](http://affordance.typepad.com), à l'adresse <http://affordance.typepad.com>.

**[décryptage]** Des folksonomies aux *hashtags*, quelles cultures informationnelles ? Ou comment la présence numérique revient de plus en plus à exercer des fonctions documentaires – des fonctions de tri, d'évaluation, d'indexation, etc. Comment cette évolution interagit-elle avec les savoirs professionnels ?

## Culture documentaire et folksonomie : l'indexation à l'ère industrielle et collaborative

### À l'index

Il a déposé des photos de ses vacances sur FlickrR ; recherché la vidéo d'un extrait de colloque sur YouTube ; partagé des signets sur Delicious ; publié un article sur son blog ; bavardé sur Twitter à propos d'un événement récent ; consulté des photos ou des profils de ses amis sur Facebook... Dans chacun de ces cas de figure et dans bien d'autres encore, l'utilisateur, en sus de son activité de dépôt, de recherche, de publication, de consultation ou de simple conversation, a été invité à pratiquer une indexation libre. Une indexation sur ses propres traces informationnelles ou sur celles produites par d'autres. Une indexation qui traverse nos espaces numériques publics, privés et intimes, désormais réunis en une même sphère d'indexabilité. Une indexation à l'unisson de la cinétique des traces auxquelles elle s'attache : synchrone, instantanée, fragmentaire, plurielle. Enfin, une indexation parfois collaborative et le plus souvent, transparente aux autres, à tous les autres.

### C'est en forgeant qu'on devient forgeron et c'est en indexant... qu'on finit sur Twitter

Les folksonomies désignent « un processus de classification collaborative par des mots-clés librement

choisis, ou le résultat de cette classification<sup>1</sup> ». Les raisons de leur succès sont connues<sup>2</sup> : leur faible coût cognitif, leur aspect orienté-tâche et le bénéfice qu'elles apportent en terme de filtrage collaboratif. Soit l'avènement de la communauté comme indexeur.

Les folksonomies font le choix d'un affichage de nuages de mots, plutôt que celui d'une zone de saisie laissée vierge. Le mot plutôt que l'absence du mot. Or un grand nombre de scénarios de recherche d'information se construisent sur l'absence d'un besoin documentaire clairement circonscrit, que les folksonomies permettent donc de combler ou d'orienter.

Leur succès repose en outre sur une triple économie : économie de compétence documentaire et informatique<sup>3</sup>, et économie d'échelle (chacun travaille d'abord pour soi, ce qui évite la tentation du spam ou de l'usage détourné, le résultat de ce travail restant toujours possiblement mais non nécessairement bénéfique à tous).

Et puis vinrent les *hashtags*. Un *hashtag* est un mot-clé, ou plus précisément une fonctionnalité d'indexation liée au service de *micro-blogging* Twitter. Il s'agit, au sein d'un message (un *tweet*), d'un mot ou d'une concaténation de mots<sup>4</sup> précédée du symbole dièse (#), permettant de l'indexer, pour suivre l'ensemble des messages ainsi balisés ou pour leur ajouter un niveau de sens différent : #exemple

Tout comme les « tags » des folksonomies, les *hashtags*<sup>5</sup> participent pleinement au processus de redocumentarisation aujourd'hui à l'œuvre sur le Web. Eux-mêmes, d'ailleurs, se prêtent à de nouvelles scénographies documentaires. Ainsi le site WhatTheHashtag<sup>6</sup> permet de retrouver la signification d'un //



///// hashtag ainsi que d'autres informations contextuelles à son sujet : contributeurs l'utilisant le plus, fréquence d'utilisation pendant les sept derniers jours, autres *hashtags* et sites web associés.

## Du marquage au rebond, de la rareté à l'abondance

Hier, l'indexation était autant une technique documentaire qu'une trace mémorielle inscrite – et parfois sanctuarisée. Aujourd'hui, avec l'indexation sociale, avec la synergie presque immédiate entre les temps jadis distincts de la recherche, de l'accès et de la consultation, elle a effectuée une mue cardinale. Sans se départir de ses vertus premières, elle a désormais pour vocation principale de faire naître ou d'amplifier la capacité de rebond inscrite à chaque carrefour de nos navigations sur le Web, dans la plus petite unité d'information disponible. C'est une véritable ingénierie de la sérendipité qui se met en place.

En d'autres termes, l'indexation est passée d'une économie de la rareté (rareté du sens permise par les différents techniques de désambiguïsation, le travail sur les renvois), à une économie de l'abondance, où l'on ne se préoccupe guère des « risques » d'ambiguïté, d'homonymie. On compte sur l'aspect massivement collaboratif et ouvert du processus pour les atténuer ou, à l'inverse, on accepte ces risques comme autant de chances nouvelles de s'orienter différemment dans de gigantesques entrepôts de ressources hétérogènes.

Ce basculement est fondamental : *il ne s'agit plus seulement de gérer des hiérarchies documentaires mais des hétérarchies<sup>7</sup> de ressources et de parcours.*

1 Définition issue de la page <http://en.wikipedia.org/wiki/Folksonomy>.

2 Wikipédia. Pour une analyse approfondie de ces raisons, se reporter à : O. Ertzscheid, G. Gallezot, Étude exploratoire des pratiques d'indexation sociale comme une renégociation des espaces documentaires. Vers un nouveau big bang documentaire ? [http://archivesic.ccsd.cnrs.fr/sic\\_00091679/fr](http://archivesic.ccsd.cnrs.fr/sic_00091679/fr), ainsi qu'à : O. Ertzscheid, Indexation sociale et folksonomies : le monde comme catalogue, [www.slideshare.net/olivier/oe-abes-mai2008](http://www.slideshare.net/olivier/oe-abes-mai2008)

3 À l'inverse des balises <Meta> qui requéraient de la part des usagers de maîtriser le code HTML ainsi que des habits d'indexation professionnels.

4 Pour fonctionner et être reconnus comme tels, les *hashtags* ne doivent contenir ni accent, ni espace, ni signe de ponctuation.

5 Comme les tags folksonomies, les *hashtags* peuvent aller de l'explicite : #iran-elections à l'allusif : #MJ (pour Michael Jackson), au contextuel : #RIP (pour signaler la mort d'une personne) jusqu'au plus farfelu : #memepasmal.

6 <http://withashtag.com>.

7 On parle d'hétérarchie à partir du moment où, dans une organisation, il n'y a pas de « niveau supérieur ». C'est Warren Mc Culloch, l'un des premiers cybernéticiens, qui avait créé ce terme pour décrire certains programmes informatiques.

8 Citons comme exemples les « machine-tags » de Flickr (<http://code.flickr.com/blog/2009/07/06/extraextraextra>), le « collaborative structured tagging » d'Amapedia, service d'Amazon (<http://amapedia.amazon.com/view/Meta>About/id=114244>), ou encore « SemanticScuttle » (<http://sourceforge.net/projects/semanticscuttle>), une plate-forme de gestionnaire de signets qui autorise les tags structurés (avec des relations d'héritage sémantique).

## Le processus de redocumentarisation

« L'objectif de la documentarisation est d'optimiser l'usage du document en permettant un meilleur accès à son contenu et une meilleure mise en contexte. Le numérique, par nature, implique une re-documentarisation. Dans un premier temps, il s'agit de traiter à nouveau des documents traditionnels qui ont été transposés sur un support numérique en utilisant les fonctionnalités de ce dernier. Mais le processus ne se réduit pas à cette simple transposition. En effet, bien des unités documentaires du Web ne ressemblent plus que de très loin aux documents traditionnels. Dans le Web 2.0 [...] la stabilité du document classique s'estompe et la redocumentarisation prend une toute autre dimension. Il s'agit alors d'apporter toutes les métadonnées indispensables à la reconstruction à la volée de documents et toute la traçabilité de son cycle. Les documents traditionnels eux-mêmes, dans leur transposition numérique, acquièrent la plasticité des documents nativement numériques et peuvent profiter des possibilités de cette nouvelle dimension. »

Jean-Michel Salaün, « La redocumentarisation, un défi pour les sciences de l'information », *Études de communication*, 2007, n° 30, <http://hdl.handle.net/1866/1724>

Dès lors, l'indexation sociale s'affranchit d'un certain nombre de savoirs et d'héritages professionnels. N'importe qui peut en effet se trouver en situation d'indexeur : simples usagers de services, producteurs ou consommateurs de contenus, novices ou experts. Ensuite, la granularité des ressources indexables s'élargit considérablement : au-delà des ressources scientifiques et des objets culturels, c'est l'ensemble des photos, vidéos, documents de travail, billets de blogs et autres « profils » sur les réseaux sociaux qui entrent dans le champ des objets indexables, qu'ils « m'appartiennent » ou non. Enfin ce sont les finalités et la nature de la tâche d'indexation qui se diversifient : on peut indexer à la volée (en même temps que l'on parcourt ou découvre des ressources), pour son usage personnel ou à destination explicite d'une communauté d'intérêt, en parfaite connaissance de cause ou dans une totale ignorance du sujet ou de l'objet du processus.

## L'indexation : industrielle ou sociale ?

On oppose souvent l'indexation sociale, participative, contributive, à l'indexation « sérieuse » (c'est-à-dire normée) et à l'indexation machinique et aujourd'hui industrielle, pratiquée par les moteurs de recherche. Opposition factice, à bien y regarder.

Ainsi Google, qui revendique comme un *motto* de la firme de ne pratiquer « aucune intervention humaine dans [ses] résultats de recherche », pratique pourtant sans cesse l'indexation humaine et collaborative : chacune de nos requêtes couplée aux clics sur tel résultat issu de cette requête équivaut littéralement à une qualification humaine du résultat de la requête, qualification qui sera ensuite intégrée dans l'ensemble des paramètres algorithmiques permettant, pour l'utilisateur identifié et/ou pour l'ensemble des requêtes semblables, de faire varier l'ordonnement des résultats. Dit autrement, le couple

« requête / activation d'un lien de la page de résultat » équivaut exactement à la procédure qui, dans les folksonomies, fait correspondre un ou plusieurs mots-clés librement choisis à une page web. L'indexation à l'ère industrielle est donc *nécessairement* sociale.

## À suivre...

La culture documentaire du Web est celle d'une bibliothéconomie de masse, d'un passage à l'ère industrielle des processus et compétences liées à l'indexation, dans un environnement où tout est document, tout devient « fait » documentaire, jusqu'à nos plus infimes traces identitaires. Dans ce monde, l'horizon de l'indexation se déplace ; la question n'est plus celle de l'autorité (qui a compétence pour indexer) ni même de l'expertise (qui a compétence pour indexer), mais celle de l'usage (qui a besoin d'indexer).

Dès lors, trois sphères de pratiques sont amenées à cohabiter et à interagir : l'indexation professionnelle de fonds documentaires parfaitement circonscrits, l'indexation industrielle de l'ensemble des contenus du Web telle qu'elle est opérée par les moteurs de recherche, et l'indexation sociale. La ten-

dance de fond étant celle d'une perméabilité chaque jour plus grande et plus réciproque entre ces sphères.

L'interaction avec les savoirs professionnels peut être aujourd'hui prolongée selon trois axes. Tout d'abord, « ceci ne tuera pas cela ». Il faut laisser s'opérer toutes les substitutions nécessaires entre d'anciennes compétences et de nouvelles habiletés, pour autant que ces dernières répondent à un besoin réel que les premières ne peuvent suffire à combler. Ensuite, ne pas jeter les anciens formalismes avec l'eau du tout participatif. La compétence d'indexation sera de plus en plus déterminante, tout au moins pour une partie de la morphogénèse complexe et permanente des contenus et des interactions se donnant à lire sur la toile. Enfin, ces deux cultures documentaires ne sont pas nécessairement antagonistes. Il existe une voie médiane. Des modèles hybrides sont d'ores et déjà disponibles, capables par exemple d'ajouter de la structuration ou des notions d'héritage à l'intérieur de folksonomies, tout en bénéficiant de l'effet d'échelle et de la simplicité d'amorçage de ces dernières comme des communautés qu'elles fédèrent.

En tout état de cause, ces perpétuels déplacements de compétences, ces réinventions de la chaîne de traitement du document préfigurent les nouvelles granularités sociales que l'on peut déjà observer dans les usages du Web et dans la manière dont elles impactent notre rapport au monde, et pas simplement de manière virtuelle. •