

Pourrais-je avoir un nuage de tags sur mon site web 2.0 ? ou faire du neuf avec du vieux

Publié le 17/12/2006

Petite remise en perspective historique des innovations du Web 2.0 pour. Sous couvert d'innovation, se présentent bien souvent des fonctionnalités ou possibilités déjà existantes, mais un peu rajeunies...

La brumeuse prétendue innovation des nuages de tags À en croire les commentateurs du Web 2.0, une des grandes innovations mise au service de l'internaute, ce sont les *tags*. Et de fait, on voit fleurir sur les meilleurs sites ou blogs ces "*nuages de tags*" permettant de visualiser la pertinence et le poids des sujets les plus traités sur le site.

Le snobisme consistant à jargonner techniquement en anglais plus que de raison, on ne comprend pas au premier abord que ce fameux système de tags n'est autre que la bonne vieille méthode d'indexation à l'aide de mots-clés, enseignée en bibliothéconomie et en documentation sous le nom d'indexation matière : il s'agit d'apposer des mots-clés décrivant un objet mis à disposition du public : texte, image, son, vidéo...

Une fois de plus, nous découvrons qu'après avoir prophétisé - voire préconisé - la fin des métiers de l'information-documentation dont l'existence serait rendue caduque par l'émergence du Web, celui-ci ne tient que par les techniques documentaires à tous les étages.

Nous passons ici en revue les diverses techniques documentaires utilisées pour sauver, dès l'origine le Web du chaos.

Les balbutiements de l'Internet grand public - c'est-à-dire le Web - furent placés sous le sceau des techniques documentaires classiques, même si d'aucuns ont pu croire que l'ère des techniques documentaires étaient révolues. Illusion d'optique, relayée par quelques penseurs pas forcément au fait des réalités professionnelles (1). C'est ainsi que les tout débuts du Web furent assistés des techniques documentaires réinventées sans le savoir par les pionniers du net.

Les répertoires : classifications inavouées La première initiative - en termes méthodologiques, car

historiquement, les deux naissent en même temps - pour suivre l'extraordinaire démarrage des sites web, fut celle des répertoires. Deux étudiants américains décidèrent de suivre la création de tous les sites Web et de les répertorier. La structure de description des sites ainsi référencés reprenait peu ou prou les éléments essentiels de la description bibliographique : titre du site, auteurs, mots-clés, description...

Mieux : ils allaient structurer la navigation dans le répertoire à partir de "*categories*", comme on dit en anglais, c'est-à-dire tout simplement d'une classification quasi-universelle, à ceci près qu'elle n'avait pas la rigueur des classifications connues en bibliothéconomie. Ainsi s'est développé le répertoire de *Yahoo !*, toujours vivant, même si les dirigeants de la société le masquent pudiquement aujourd'hui, pour des raisons hors de propos ici. Tous les grands répertoires - encore nommés annuaires ou guides web - obéissent à cette logique classificatoire directement héritée des classifications bibliothéconomiques et documentaires. Comme quoi les métiers de l'information-documentation apportaient quelque chose de bon...

Les moteurs de recherche : logiciels documentaires masqués L'initiative concurrente a consisté à indexer automatiquement tout le Web mondial, à l'aide d'outils puissants nommés moteurs de recherche (*search engines*). Peu se sont avisés que ces moteurs ne sont en fait que les formes les plus avancées des logiciels documentaires (2), regroupés sous le concept œcuménique d'*informatique de contenu*... De sorte que lorsqu'on cherche une information à partir d'un moteur comme Google, on fait de la documentation sans le savoir...

Bien sûr, le traitement a pris des proportions mondiales et industrielles telles qu'on est loin de l'informatique artisanale des bases de données internes aux centres de documentation. Mais les techniques de base sont exactement les mêmes.

De la sorte, le Web mondial n'est devenu ce chaos très organisé qu'avec l'aide - bien involontaire parfois - des techniques professionnelles issues des métiers de l'information-documentation. L'erreur de bien des professionnels a été de ne pas s'en apercevoir et de croire que l'Internet allait "leur prendre leur travail", plutôt que de tenter de se placer aux positions stratégiques de concepteur de systèmes d'information sur le Web. Et pourtant, il en aurait bien fallu. Rares sont les sites web qui soient correctement décrits et par conséquent, correctement référencés ; parce que non correctement conçus sous l'angle de leur repérage par les moteurs.

Une réalité presque dépassée : les metatags Il est amusant de se rappeler que les robots des premiers moteurs de recherche exploraient notamment les métadonnées (*metadata*, logées dans des

metatags) situées dans l'en-tête (partie cachée) d'une page web, dans laquelle le créateur de la page avait la possibilité de fournir ce qu'on nomme une notice bibliographique en bibliothéconomie :

Titre, Auteur, Description (petit résumé) et Mots-clés.

Rappelons qu'un groupe de travail s'est penché sur ces questions de métadonnées : le Dublin Core, du nom de la ville des USA (et non la capitale irlandaise) où ils se sont réunis.

Aujourd'hui, si les bons concepteurs de sites pratiquent toujours les métadonnées, celles-ci ne sont plus exploitées par de nombreux moteurs. Pourquoi ? Tout simplement parce que l'indexation qu'elle contient, réalisée par l'auteur du site, n'est pas fiable.

Il y a d'abord les cas de *spoofing* (exagération) : abus de mots-clés intempestifs destinés à piéger les moteurs de recherche et les internautes. Certains ont cru bon d'aligner vingt fois le même mot-clé pour obtenir la meilleure place dans le classement des résultats. Les moteurs ont donc d'abord choisi de ne retenir qu'un maximum de deux occurrences pour chaque mot. Certains ont aussi choisi d'introduire des mots-clés hors sujet, dans le seul but d'attirer de nombreuses connexions (critère pour vendre de la publicité sur son site). Ainsi Pamela Anderson aurait-elle été le sujet de très nombreux sites, si du moins on en croit le champ "mot-clé" des *metatags*...

Il y a ensuite le fait que tout concepteur de site - fût-il bon développeur web - n'a pas forcément les compétences pour bien rendre compte du contenu d'un site par quelques mots-clés soigneusement choisis. On peut oublier certains aspects du site, en exagérer involontairement d'autres. Bref, nous n'allons pas plaider pour le professionnalisme de l'indexation ; s'il existe deux professions pour s'y employer (bibliothécaires et documentalistes), c'est qu'il y a une raison !

De la limite des tags du Web 2.0 Aujourd'hui on nous présente les tags comme une innovation sans précédent, là où on réinvente une technique documentaire séculaire (la classification Dewey est née en 1876). L'innovation réside en effet dans la possibilité de voir s'afficher de manière originale (sous forme de "nuage") les mots-clés les plus usités. Ce système de mots-clés associés aux documents mis en ligne sur les sites personnels ou collaboratifs permettra en effet de mieux les retrouver dans une certaine mesure.

Il n'en demeure pas moins qu'on reste dans un certain amateurisme, puisque les publicateurs de ces informations n'ont pas le savoir-faire pour correctement indexer. Cela prend des proportions tangibles lorsqu'il s'agit d'indexer des photos ou des vidéos. L'indexation de l'image animée est une des choses les plus délicates. Hormis les données objectives (lieu, date, météo, circonstances objectives telles qu'une fête) - dont les auteurs peuvent déjà omettre certains aspects, l'image véhicule un non-dit que l'analyste d'image sait repérer et décrire, avec une formation appropriée.

Souhaitons en outre fortement que la pratique des tags ne dérive pas en *spoofing* comme les *metatags*, au point qu'on soit obligé de les ignorer. Le bel apport du Web 2.0 perdrait de sa crédibilité, et ce serait dommage.

Une autre difficulté se pose.

La pratique des tags est laissée à l'appréciation de chaque producteur d'information, ce qui est l'esprit même du net. Mais cette liberté même impose des limites à l'efficacité du système.

Tel internaute va produire par exemple un article sur l'informatique documentaire. Il choisira comme tag associé : "informatique documentaire", un autre auteur, à sujet identique, pourrait choisir avec autant de raison "informatisation documentaire", surtout si son article évoque plus la démarche que les outils, cependant qu'un troisième pourrait bien choisir "logiciel documentaire". On touche ainsi du doigt les limites de l'indexation dite libre, en dehors de tout langage documentaire contrôlé (classification ou thésaurus uniformisé) permettant de rattacher une même réalité conceptuelle au même mot, "descripteur" du concept.

Remarquons aussi qu'un même internaute pourra, au fil de ses publications, utiliser un jour "informatique documentaire" et un autre jour "documentation informatisée", tout simplement parce qu'il ne s'est pas souvenu de son ancienne indexation à quelques mois d'écart ; tous les documentalistes ayant pratiqué l'indexation libre connaissent ce phénomène. Une analyse des nuages de tags sur certains sites laisse perplexe à cet égard : des mots-clés très proches coexistent sans qu'on sache ce qui a motivé le distinguo. Leurs auteurs non plus, sans doute... Parfois même, c'est une simple question d'écriture "Web 2.0" et "Web2.0"... ou une simple question de langue : "e-mail" et "courriel".

Cette disparité est en partie rattrapée grâce à l'affichage des mots-clés, sur les fameux nuages, ou lorsqu'il est possible de voir la liste complète des mots-clés utilisés. À condition de se trouver sur le site concerné. Mais pour une recherche via un moteur, ces mêmes mots-clés resteront presque aussi imprécis que le langage naturel contenu dans les textes.

En d'autres termes, la pratique des tags est une innovation qui rend grâce aux techniques documentaires. Même si elle n'est pas aussi affinée, elle permet une certaine amélioration de l'accès à l'information.

Taxonomie et ontologie Avec l'arsenal des tags, émergent les mots ronflants de taxonomie (ou taxinomie) et d'ontologie. Ces mots savants cachent des réalités là aussi séculaires.

La taxonomie (de *taxis* = placement, mise en ordre et *nomos* = règle, en grec) est l'art du

classement d'objets selon une hiérarchie, c'est-à-dire qu'elle est notamment la science des classifications.

Quant à l'ontologie (du grec *ontos* = existant et *logos* = discours), elle permet à l'origine de décrire des objets, intellectuels ou matériels et leurs relations entre eux. C'est donc - pour simplifier, sous notre angle d'étude - la science qui préside à l'élaboration des thésaurus.

Ces termes sont utilisés aujourd'hui par dérivation pour désigner des essais de classification (taxonomie) d'objets, notamment pour s'y retrouver dans les tags. Les ontologies sont utilisées en informatique pour cerner les relations logiques qui existent entre des objets qui doivent être traités par les systèmes automatisés.

Autrement dit, même si nous pouvons paraître réducteur, ces essais mettent au jour les limites de l'indexation libre, dénoncées plus haut. Ils tentent donc *a posteriori* de structurer les notions pour mettre de l'ordre dans les nuages de tags, ou pour améliorer la navigation en introduisant des relations entre les notions.

Somme toute, on réinvente une nouvelle fois les notions de classification et de thésaurus. N'est-ce pas à la fois le plus bel honneur qu'on puisse faire aux pratiques les plus avancées des professionnels de l'information-documentation, mais aussi un constat de malentendu puisque ces métiers devraient - une fois de plus - être présents sur ces terrains sur lesquels on réinvente et remet au jour ce qu'ils pratiquent, même imparfaitement, depuis des lustres ?

|cc| Didier Frochot - décembre 2006

Notes : [1](#). Nous pensons notamment à ce grand penseur du management de l'information qui affirmait devant des parterres de managers que les documentalistes ne serviraient plus à rien puisqu'aujourd'hui Internet permettait de tout trouver au bout des doigts... C'était méconnaître à la fois les techniques documentaires et les sites Web en profondeur, comme nous le montrons ici. Repensons aussi à cet article des Échos qui prophétisait, en 1994, la fin des secrétaires, des documentalistes, et même des agents de voyages...

[2](#). Lire : Catherine Leloup, *Moteurs d'indexation et de recherche*. - Eyrolles, 1997.