

"Tagging", classer de manière décentralisée

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 1 Avril 2006. Dernière mise à jour le 2 Avril 2006

<http://www.bortzmeyer.org/tagging.html>

Le "*tagging*", le marquage de ressources par les utilisateurs, sans schéma de nommage et sans thesaurus standard est un des concepts qui a récemment enrichi l'Internet en permettant de créer des services d'accès à l'information, alimentés et classés par leurs utilisateurs.

Traditionnellement, pour s'y retrouver dans une vaste quantité d'informations (par exemple un ensemble de signets ou bien une immense galerie de photos), on confiait à un petit groupe de professionnels de la documentation la tâche d'établir un thesaurus, une liste de mots-clés, et un groupe plus vaste (mais quand même fermé, n'incluant pas la totalité des utilisateurs) était chargé d'appliquer ces mots-clés à chaque ressource, travail très long.

Ce travail peut typiquement être parallélisé et c'est là qu'Internet est une innovation fantastique. On peut désormais enrôler des milliers de bénévoles pour faire le marquage des ressources.

Mais le "*tagging*" moderne va plus loin : il n'y a plus de thesaurus, la liste des mots-clés n'est plus figée, elle est établie par les utilisateurs, de manière décentralisée. Je peux "*taguer*" une ressource avec le mot-clé "*foobar*" ? Je n'ai pas à demander d'autorisation, je le fais. Cela donne ce qu'on nomme la "*folksonomie*", la classification par le peuple, pour le peuple.

Cette idée de "*tagging*" est tellement à la mode (tout le monde s'en réclame désormais) qu'elle a même donné lieu à un pastiche, <<http://www.tagtagger.com/>>.

Les deux sites à folksonomie les plus connus sont del.icio.us <<http://del.icio.us/>> et Flickr <<http://www.flickr.com/>>.

del.icio.us est un site qui ne contient pas d'informations, uniquement des signets. Chacun enregistre les signets de son choix (vous pouvez voir les miens en <<http://del.icio.us/bortzmeyer>>). Et chacun peut marquer ses signets avec des "*tags*". On peut ensuite consulter del.icio.us par utilisateur (comme le lien précédent) ou par "*tag*" (par exemple, pour voir les liens sur Xen, on peut regarder

<<http://del.icio.us/tag/Xen>>. Et il existe de nombreuses autres possibilités (pas toujours bien documentées).

Flickr <<http://www.flickr.com/>>, quant à lui, contient de l'information en quantité, c'est devenu le principal outil de publication de photos. Toute manifestation contre le CPE est immédiatement suivie d'une publication sur Flickr <<http://www.flickr.com/photos/tags/CPE>> de toutes les photos prises par les nombreux appareils numériques présents. Ces photos peuvent ensuite être "*taguées*" et cherchées par "*tag*", par exemple <<http://www.flickr.com/photos/tags/Sarlat>> si je cherche des photos de Sarlat dans le Périgord, ou par utilisateur ou par une combinaison des deux.

Comme avec tous les systèmes où les utilisateurs peuvent non seulement accéder à l'information mais peuvent participer à sa création, les sceptiques se demandent si cela va vraiment marcher. Comment géra-t-on les synonymes (le tiers des photos d'un numbat "*taguées*" avec "*numbat*", un autre tiers avec "*fourmilier*" et le dernier tiers avec "*myrmécobie*"), par exemple? Ce qui est curieux, et montre la nécessité d'essayer avant de prédire "*ça ne marchera pas*" est que, en pratique, la plupart des folksonomies tendent à s'uniformiser. Les logiciels aident à cette uniformisation en proposant des "*tags*", par exemple si j'enregistre dans del.icio.us le site <<http://www.bravepatrie.com/>>, del.icio.us me propose le "*tag*" "*humour*" (car c'est le "*tag*" le plus souvent utilisé par les autres qui ont "*tagué*" ce site) qui est tout à fait adapté.

Ce mécanisme d'interaction entre utilisateurs (parfois appelé un peu pompeusement "*La sagesse des foules*") a été bien décrit dans l'article *Folksonomies; Tidying up Tags?* <<http://www.dlib.org/dlib/january06/guy/01guy.html>> du magazine D-lib.

Un avertissement au passage : vous ne verrez pas de photos de moi sur Flickr car je préfère les mettre sur Wikimedia Commons <<http://commons.wikimedia.org/>> où ne se trouvent que des photos libres (celles de Flickr ont des licences très variées et je ne vois pas de moyen de chercher par licence, pour pouvoir les réutiliser).

Autre avertissement : tout lien que vous mettez sur del.icio.us enrichit une base de données que Yahoo a racheté très cher <<http://blog.del.icio.us/blog/2005/12/yahoo.html>>. Comme le souligne Paul Graham <<http://www.paulgraham.com/>>, autrefois, l'informatique était dominée par le "*hardware lock-in*" : les fabricants de matériel contrôlaient le jeu. Après, on est passé au "*software lock-in*". Les éditeurs de logiciels commandaient. Désormais, on est au "*data lock-in*" : toutes tes données appartiennent à Google ou Yahoo.

Comme le "*tag*" peut être apposé par un autre que l'auteur, il fournit une base à des systèmes d'accréditation non-hiérarchiques, en réseau. Par exemple, si j'apprécie les "*tags*" de l'utilisateur "*nws-mith*" sur del.icio.us, je peux regarder tout ce qu'il a "*tagué*" en <<http://del.icio.us/nwsmith>> et le prendre ainsi pour guide. (Au fait, mes liens à moi sont en <<http://del.icio.us/bortzmeyer>>.) Chacun peut ainsi être accréditeur et recommander des sites. Un tel mécanisme avait déjà été proposé pour le projet Interpedia <<http://groups.google.de/group/comp.infosystems.interpedia/msg/9ba401aacf26250b?output=plain>> en 1993 (les "*tags*" se nommaient "*seals of approval*") et pourrait fournir une base de solution très intéressante au problème récurrent de la validation des données dans une encyclopédie ouverte comme Wikipédia <<http://www.bortzmeyer.org/wikipedia.html>>. Au lieu de vouloir interdire les articles supposés "*de mauvaise qualité*" (ce qui est très subjectif et entraînerait des disputes sans fin), on "*taguerait*" les bons articles et on permettrait une navigation par "*tag*".