

## **Apports et limite du « social bookmarking » dans la veille sur internet.**

Grégoire CLICQUET, Henri SAMIER, Pascal FRION

\* Enseignant-Chercheur , Ecole de Design de Nantes  
rue Christian Pauc

BP 30607 - 44306 Nantes Cedex 3 - France

Tel.: +33 (0)2 51 13 50 70, g.cliquet@lecolededesign.com

\*\* Enseignant-Chercheur , directeur ISTIA Innovation & Master Information Stratégique  
ISTIA Innovation, Université d'Angers

62 avenue notre dame du lac, 49000 Angers, tél. : 33 2 41 22 65 49

samier@istia.univ-angers.fr

\*\*\* Chercheur IcomTech et Animateur Acrie Réseau

Aéroport Nantes atlantique, rue Nungesser et Coli, 44860 Saint Aignan de Grand Lieu - France

Tél. : 33 (0) 2 40 04 25 25

Mots-clefs : Web2.0, réseaux sociaux, signets sociaux, folksonomie, veille, recherche d'information

### **Résumé :**

Les « signets sociaux » peuvent-ils contribuer à enrichir la veille sur internet ? Qu'apportent ces outils comparativement aux moteurs de recherche ? Nous proposons un nouvel outil d'exploration des « signets sociaux » nommé « social explorer ». Dans une analyse comparative nous verrons dans un premier temps quelles sont les différences fondamentales entre une approche de recherche sur la plateforme « delicious » et celle plus traditionnelle via Google. Dans un second temps, nous verrons que delicious nous permet d'identifier les individus les plus en amont de la chaîne de l'information, des prescripteurs d'un domaine donné à intégrer dans un processus de veille documentaire.

## **1 : Le Web 2.0 : Espace social de services**

Le web 2.0. (O'Reilly, 2005) est avant tout un concept, qui vise à qualifier une étape significative dans l'évolution du Web et plus largement d'Internet. Le terme a pour but de poser un repère temporel dans l'histoire de l'application « Web », conçue au départ comme une extrapolation d'un vaste système documentaire qui évoque le « Memex » de Vannevar Bush (Bush, 1945), mais dont la topologie s'apparente désormais à un espace ouvert de services. L'avènement des technologies Web dites « dynamiques » ont contribué à sa mutation rapide en système d'information et de communication ; La démocratisation des accès à Internet a favorisé sa rencontre avec des usagers dont les pratiques encouragent à leur tour la création de nouveaux services. Aujourd'hui, l'internaute ajoute très facilement des contenus en utilisant des applications en ligne spécifiques animé par une volonté de création, de partage et d'enrichissement de contenus (Davis, 2005). Nous pouvons tous, très simplement, mettre en place notre propre système d'information sur le Web tout en suivant en quasi temps réel son évolution (commentaires) via des flux RSS. L'« usager spectateur » dans la consultation des documents du Web 1.0, devient « acteur » d'un système qui lui confère des capacités d'ajout, d'interprétation et d'appropriation des ressources d'un Web devenu « 2.0 », d'un système réparti décentralisé qui favorise l'enrichissement et l'exploitation de la connaissance (Ertzscheid, 2008) et apporte des nouveau comportement de recherche d'information.

Folksonomie : activateur des réseaux sociaux

Les agrégats d'individus qui constituent des communautés de fait, se fondent sur le principe de la « folksonomy », un « système de classification collaborative décentralisée et spontanée » (wikipedia). Le terme proposé par Thomas Vander Wal (Vander Wal, 2004), qui regroupe les mots « folk » (peuple) et « taxonomy » (taxinomie), incarne une méthode de classification des données décentralisée, ouverte à l'appréciation de chaque internaute. L'ajout d'un ou plusieurs mots-clés également appelés "tags" permet en effet de qualifier subjectivement des ressources, de rompre avec la traditionnelle rigidité des classements réalisés par des « médiateurs-indexeurs » (Le Deuff, 2007) et enfin, de rassembler des internautes par « étiquettes interposées ». Vander Wal (Vander Wal, 2005) distingue deux types de folksonomies :

- Les folksonomies « étroites » (narrow folksonomies) ou « personomies » (Le Deuff, 2006) qui visent prioritairement à satisfaire des besoins individuels en terme de classification des informations.
- Les folksonomies « larges » (broad folksonomies) qui visent à étendre la portée d'un mot-clef à l'ensemble de la communauté, agissant alors tels des « aimants à individus ».

## **2. La recherche d'information et le Web (Google)**

Sur Internet on accède à l'information par 2 techniques aujourd'hui complémentaires :

- La première de type « Pull » vise littéralement à « tirer » vers soi l'information, c'est-à-dire à la rechercher activement. Pour cela l'utilisateur explore des banques de données spécialisées, procède par requêtes auprès des principaux moteurs et méta-moteurs de recherches ou plus rarement y accède au hasard de sa navigation .
- L'autre méthode « Push » qui découle notamment de ce travail préalable de recherche vise à obtenir l'information de façon passive, par abonnement à des canaux spécialisés, via e-mails, mailing-lists , newsgroups ou encore par flux RSS . Ces derniers pouvant être consultés pour une meilleure efficacité via des agrégateurs tels que Google Reader .

L'utilisateur détermine ensuite la pertinence de l'information aux regards de critères déterminés de ses besoins présents ou à venir (prospective) et de son rapport avec le sujet étudié (relations connexes). Si l'information comporte un intérêt quelconque, l'utilisateur l'analyse puis la classe pour l'archiver.

Google ne propose en définitive qu'une liste de liens ordonnés selon leur autorité respective (sans la mentionner) agrémentés des informations contenues dans le code source de la page (meta description). L'ouverture proposée par google est réduite à l'ajout de liens commerciaux ou à l'affichage en bas de page des recherches apparentées. Delicious, comme beaucoup de plateformes sociales, propose des listes apparemment non ordonnées mais enrichies par des informations

supplémentaires : en quantifiant la popularité d'une ressource, en l'associant à un usager (primo bookmarker) et en proposant d'étendre la recherche à des tags connexes. Ces caractéristiques méritent d'être étudiées en vue de proposer une nouvelle expérience utilisateur dans le domaine de la recherche d'information.

Dans notre étude nous considérons les signets sociaux comme une réponse aux étapes d'analyse et d'archivage du processus de veille décrit précédemment. L'analyse d'une ressource est formalisée par l'ajout d'un commentaire et son archivage est organisé par le biais de mots-clefs, l'ensemble est externalisé sur le Web sur la plateforme delicious. D'autre part, l'identification de bookmarkers prescripteurs devient ici un élément d'« input » de la technique « pull » évoquée précédemment, dans la mise en œuvre d'une stratégie de veille documentaire.

### **3 : Recherche d'informations sur delicious**

delicious.com permet aujourd'hui à plus de 5,5 millions d'internautes de sauvegarder et d'organiser quelques 200 millions d'adresses Web. L'utilisateur détermine également s'il souhaite partager une ressource (défaut) ou non (do not share). Le site propose la liste des signets récemment mémorisés (recent) ainsi que celle des liens les plus populaires (popular). Dans le premier cas, la liste est logiquement organisée du lien le plus récent au plus ancien et le système propose de filtrer les résultats par le nombre d'occurrences d'une URI (saved by at least). Par contre, la hiérarchisation des liens populaires ne dépend pas seulement de leur popularité au sens strict du terme (saved by), mais d'une combinaison entre leur popularité et la fréquence à laquelle ils sont mémorisés. Ce mode de classification des résultats est appliqué à l'ensemble des recherches effectuées sur delicious (Benhuman-Fick, 2010).

Il est également possible d'explorer les ressources mémorisées par les usagers du site à partir de mots-clefs (<http://delicious.com/tag/monmotclef>) et de parcourir ainsi les espaces publics d'autres usagers sur une thématique donnée. L'assignation de plusieurs mots-clefs à une même ressource permet d'affiner le tri des résultats et d'étendre la recherche à des domaines connexes (Cloud tags / Related tags). Enfin, il est possible de se créer un sous-réseau en faisant figurer le compte d'un autre usager dans la rubrique « your network » et profiter ainsi de l'ensemble des informations qu'il mémorise en temps réel ou de s'abonner, via la syndication de contenus, plus spécifiquement à l'un de ses mots-clefs (subscription). L'un des membres de votre réseau peut également vous soumettre des liens en vous signifiant comme destinataire lors de leur mémorisation « link for you ».

### **4 : Présentation du nouvel outil « Social Explorer »**

Le « Social Explorer » est une application Web que nous avons développée en XHTML et PHP/MySQL qui procède dans un premier temps à l'analyse syntaxique du code source de chacune des pages d'un site de signets sociaux afin d'en isoler les données pertinentes : Ressources, Commentaires, Tags et l'identifiant des usagers. Les données ainsi recueillies permettent d'effectuer des recherches et de proposer de nouveaux modes de classement dans le but de les analyser.

L'application comprend un champ de saisie des tags qui après validation, lance le processus de récupération de l'ensemble des URI s bookmarkées, du nombre total de fois où une URI a été enregistrée, des mots-clefs associés et du nom du premier usager l'ayant mémorisé. Ces informations sont stockées dans une table et permettent, dans un second temps, l'exploration de chacun des comptes des usagers afin de les classer en fonction des profils types définis précédemment. Pour chaque tag recherché, 2 tables MySQL sont créées : tag\_uri et tag\_users, elles vont permettre l'accumulation des données recueillies au cours des recherches effectuées successivement sur un même tag.

Nous proposons une étude comparative de la nature des ressources Web mémorisées sur delicious et celles qui résultent d'une recherche sur Google. Pour cela nous choisissons de restreindre notre approche à des recherches informationnelles (Gwizdka 2008) (Krause 2007) et (Broder, 2002), nous excluons donc les recherches « navigationnelles » dont l'objet vise à trouver une page en particulier et les recherches « transactionnelles » qui visent à accomplir un acte d'achat. Le contexte de notre étude concerne le domaine de la petite enfance à travers l'exploration des résultats produits par une recherche sur les termes : Breastfeeding (allaitement), Chilcare (protection de l'enfance) et Parenting (élever ses enfants).

Terme	Google	Delicious
Breastfeeding	13,200,000	11,991
Childcare	16,300,000	12,158
Parenting	61,500,000	95,383

Figure : nombre d'occurrences par outil de recherche

Afin de procéder à une comparaison réaliste des résultats, nous garderons dans Google les 30 premiers résultats obtenus par la saisie de l'un de ces mots-clefs, nous procéderons à l'identique sur delicious en récupérant l'ensemble des URIs associées au tag recherché pour ne conserver ensuite que les 30 premiers résultats triés par popularité en éliminant ainsi l'influence du facteur temporel sur le classement des résultats. Il est à noter que le facteur temporel pourrait à lui seul faire l'objet d'une étude spécifique qui permettrait d'analyser la dynamique d'agrégation des individus autour d'une ressource et les logiques de propagation de l'information.

Le Social Explorer récupère les données issues de l'interface publique de consultation de delicious et par conséquent est sujette aux limitations fixées dans l'utilisation du service. Ainsi, nous ne pouvons récupérer qu'un maximum de 2000 URIs (20 pages x 100 URIs) récemment mémorisées, d'où la nécessité de renouveler régulièrement l'opération afin d'accroître le nombre d'URIs collectées. D'autre part, nous ne pouvons solliciter la plateforme delicious de façon intensive qui pour garantir l'équité dans l'accès à son service rejette les sollicitations intensives. La conception de notre application prend en compte ces paramètres et permet donc de poursuivre les opérations de collecte malgré des déconnexions aléatoires.

L'affichage des résultats peut être consulté :

<http://infoviz.lecolededesign.com/humanhub/delicious/admin/archives.php>  
(login : visitor / mdp : adminutde)

Pour qualifier les ressources obtenues, nous proposons de les décrire en fonction de 3 critères principaux qui déterminent :

Critère n°1 : Le type de système d'information qui héberge la ressource : Page, portail, blog, wiki, forums, listes, outils (application), galeries puis du positionnement de cette source au regard de la recherche effectuée : généraliste, connexe, ou spécialiste.

Critère n°2 : La nature du message : Page d'accueil, article, image, vidéo, son et son caractère : Informatif, officiel ou commercial

Critère n°3 : La proximité de ce message avec l'objet de la recherche : aucun, éloigné, polémique, pratique, ou avancé.

## 5. Analyse des résultats

D'un point de vue général, on ne constate pas de différences significatives en ce qui concerne les types de systèmes d'information. Dans leur grande majorité, les ressources sont hébergées au sein de portails horizontaux donnant accès à un ensemble d'articles et de services associés : Création de comptes utilisateurs, inscription à une liste de diffusion, abonnement aux différents flux RSS, etc... Les liens indépendamment de la plateforme étudiée pointent vers la page d'accueil de ces portails ou directement sur les articles dans des proportions similaires.

Les différences notables se situent :

a. Sur les caractéristiques des sources :

Dans le cas d'une recherche opérée via Google, la majorité des URIs pointent sur des portails qui ont un lien très fort avec la recherche effectuée, à la différence de delicious qui conduit à des portails majoritairement « connexes » c'est-à-dire ayant une relation périphérique au domaine étudié.

b. Sur le caractère des messages :

Hormis la proportion importante de messages à caractère informatif, on constate que les usagers de delicious ont tendance à mémoriser des ressources publicitaires ou commerciales, tandis que Google privilégie des sources officielles (institutions spécialisées, organismes de recherches). Ce phénomène est en corrélation directe avec la logique d'indexation de Google basée sur la notion d'autorité d'une ressource mesurée en fonction de la quantité et la qualité des liens sortants et entrants (« page rank »).

c. Sur la pertinence ou plus exactement la proximité des résultats avec la recherche effectuée :

Google et delicious conduisent vers des liens essentiellement pratiques (conseils sur la nutrition, l'hygiène, l'éducation...), cependant les sources officielles repérées par Google produisent davantage d'articles spécialisés qualifiés ici d'« avancés », tandis que delicious propose des liens vers des ressources « polémiques » quand celles-ci ont un rapport avec le mot recherché (aucun rapport : 6%). Cette dernière catégorie nous permet cependant d'affirmer que les résultats fournis par delicious viennent compléter ceux fournis par Google. De plus, nous constatons que le taux de recouvrement entre ces deux méthodes de recherche est très faible puisqu'il ne concerne seulement que 3 URIs concernant les recherches sur le terme « breastfeeding ».

La mise en place d'une veille suppose la détection de sources d'informations pertinentes. Dans le contexte du Social Bookmarking ces sources sont des usagers qui mémorisent activement des liens relatifs au thème de recherche. Nous devons par conséquent identifier ces prescripteurs d'informations, ceux que nous considérons comme faisant autorité dans un domaine en particulier. Nous distinguons ainsi 3 types de « profils usagers » :

Celui qui possède un grand nombre de liens étiquetés avec le tag recherché (« Top bookmarker ») qui illustre l'intérêt de l'utilisateur pour le thème étudié.

L'utilisateur qui privilégie ce tag en particulier par rapport à ses autres tags (« Active bookmarker »), il consacre ainsi une part non négligeable de son travail de veille au domaine concerné. Ne sont considérés que les usagers qui possèdent au moins 10 bookmarks sur le tag concerné.

L'utilisateur qui est le plus prompt à mémoriser des ressources (« Earliest bookmarker »), il occupe ainsi une position de leader par sa capacité à trouver des nouvelles ressources et/ou dispose de sources d'informations pertinentes.

Après analyse des corpus d'URIs de ces individus ainsi discriminés, en privilégiant dans un premier temps ceux qui appartiennent à nos 3 catégories, notre stratégie de veille consistera à nous abonner, à leur flux RSS sur le tag concerné . Dès lors, nous bénéficierons automatiquement et en temps réel de l'enrichissement de nos ressources par celle de ces usagers.

Le Social Explorer nous permet ainsi d'identifier :

- Breastfeeding : 775 bookmarkers
- Chilcare : 890 bookmarkers
- Parenting : 1263 bookmarkers

## 6. Conclusion et perspectives

Nous montrons à travers cette étude la pertinence d'approches alternatives, dans la recherche d'informations sur le Web et plus particulièrement sur un Web devenu « 2.0 » par l'exploitation d'outils sociaux en ligne. La complémentarité des approches techniques (moteurs) et sociales (Web 2.0) nous invite à encourager dans le cadre d'une nouvelle approche dans le domaine de la recherche sur le Web, des solutions hybrides qui s'appuieraient sur une dynamique communautaire afin de constituer dans un premier temps un corpus de liens pertinents, pour utiliser ensuite des techniques plus classiques, propres aux moteurs de recherche de « crawling » vertical pour procéder à une indexation des ressources préalablement discriminées.

En reformulant le classement des URIs proposé par delicious nous obtenons des embryons d'annuaires thématiques co-constitués et composés de ressources avant tout pratiques et polémiques qui élargissent par conséquent le champ d'exploration documentaire. Nous pouvons par conséquent émettre l'hypothèse selon laquelle, une veille exploitant « l'effet de bookmarks » ferait émerger des pratiques ou des besoins d'utilisateurs formulés à travers la mémorisation d'un lien Web. Dans le domaine industriel, cette approche peut donc contribuer à enrichir la phase amont de conception de produits et de services innovants d'un aspect « social » ou encore permettre une remontée de « l'innovation horizontale » définie par Eric Von Hippel (Hippel, 2005) vers les entreprises et envisagée par Dominique Cardon (Cardon, 2005). D'autre part, le social explorer nous permet d'identifier des « lead users » qui peuvent être invités à participer à des tests d'usages, ou, à posteriori, être relais de communication dans la promotion d'un nouveau produit ou service.

Le très faible taux de recouvrement avec Google, nous invite une nouvelle fois à nous interroger sur la place des moteurs de recherches dans l'environnement « Web » considérés encore à tort comme les seuls outils de recherche documentaire.

Cette première approche, nous a conduit à mettre en place un outil graphique de type « treemap » permettant la visualisation journalière des sites les plus mémorisés (popular) sur delicious. Cette veille très généraliste nous permet de connaître les sites les plus mémorisés au jour le jour, de percevoir le « zeitgeist » quotidien de delicious. D'autre part, nous avons développé à l'Ecole de design Nantes Atlantique notre propre plateforme de signets sociaux basée sur « Semantic Scuttle » . Le site Web « [library@lecolededesign.com](mailto:library@lecolededesign.com) » a pour objet, la constitution par les étudiants et enseignants, d'un corpus de signets dans les domaines du design produit, d'interactivité et d'espace. L'accès direct à la base de données nous permettra la mise en place d'outils statistiques et sémantiques plus complets nous permettant d'approfondir l'analyse de la dynamique collaborative de production de ressources documentaires au sein d'un secteur d'activité, en vue de sa transposition en entreprise.

D'autre part, nous travaillons actuellement à la mise en place d'un nouvel outil qui, sur la base d'agrégation de divers flux RSS nous permettra cette fois de recueillir, croiser et analyser différentes sources d'information et ne pas nous limiter à la seule plateforme delicious. Ce nouvel outil nous permettra de mettre en œuvre une veille spécifique au Web 2.0 destiné à stimuler l'innovation d'une entreprise.

## Bibliographie

- Al-Khalifa H-S., Davis H-C, (2007), Delicious Learning Resources, in Innovations in E-learning, Instruction Technology, Assessment, and Engineering Education Book, Springer Ed., pp. 139-143.
- Barnes, J. A. (1954). Class and Committees in a Norwegian Island Parish. *Human Relations*, (7).
- Benbunan-Fich R., (2010). Koufaris M., An empirical examination of the sustainability of social bookmarking websites, *Information Systems and e-Business Management*, Vol. n°8, Issue 2, January 2010, Pages 131-148.
- Braun S., Zacharias V., Happel H-J., (2008), Social Semantic Bookmarking, in Practical Aspects of Knowledge Management book, Springer Ed., pp.62-73.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2), 3-10. doi: 10.1145/792550.792552.
- Davis, I. (2005, Juillet 4). Talis, Web 2.0 and All That. *Internet Alchemy*. Retrouvé Janvier 24, 2009, de <http://iandavis.com/blog/2005/07/talis-web-20-and-all-that>.
- Ertzscheid, O. (2008). Moteurs de recherche : des enjeux d'aujourd'hui aux moteurs de demain. [http://hal.archives-ouvertes.fr/index.php?view\\_this\\_doc=sic\\_00325690&extended\\_view=1&version=1&halsid=e119lccs9j2o0jhrig058tukk2](http://hal.archives-ouvertes.fr/index.php?view_this_doc=sic_00325690&extended_view=1&version=1&halsid=e119lccs9j2o0jhrig058tukk2).
- Guy, M., & Tonkin, E. (2006). Folksonomies Tidying up Tags? *D-Lib Magazine*, 12(1).
- Gwizdka J., Cole M., (2007), Finding It on Google, Finding It on del.icio.us, in the book « Research and Advanced Technology for Digital Libraries », Éditeur Springer Ed., pp. 559-562.
- Hammond, T. H. (2005). Social Bookmarking Tools (I):A General Review. *D-Lib Magazine*, 11(4). Retrouvé de <http://www.dlib.org/dlib/april05/hammond/04hammond.html>.
- Krause B., Hotho A., Stumme G., (2008) A Comparison of Social Bookmarking with Traditional Search, in « Advances in Information Retrieval », SpringerEd., 2008, pp. 101-113.
- Le Deuff, O. (2006). Folksonomies : Les usagers indexent le web. *BBF*, (4), 66-70.
- Le Deuff, O. (2007). Folksonomies et communautés de partage de signets Vers de nouvelles stratégies de recherche d'informations. Dans colloque h2ptm07 : Collaborer, Echanger, Inventer : Expériences de réseaux. Hammamet, Tunisie. Retrouvé de <http://archivesic.ccsd.cnrs.fr/docs/00/31/59/47/PDF/ledeuffh2ptm07.pdf>.
- O'Reilly, T. (2005, Septembre 30). What Is Web 2.0 ? *Design Patterns and Business Models for the Next Generation of Software*. O'Reilly. Retrouvé de <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- Technorati. (2008, Juin 17). Technorati Media. Technorati. Retrouvé Janvier 31, 2009, de <http://technoratimedia.com/>.
- Vallet D., Candator I., Jose J.M., (2009), Exploiting Social Tagging Profiles to Personalize Web Search, in *Flexible Query Answering Systems*, Vol. n° 5822, Springer De., pp. 629-640.
- Wetzker R., Zimmermann, C., Bauckhage, C., Detecting trends in social bookmarking systems: A del.icio.us endeavor, *International Journal of Data Warehousing and Mining*, Volume 6, Issue 1, January 2010, Pages 38-57.
- Yee R., (2008), Exploring Social Bookmarking and Bibliographic Systems, in *Pro Web 2.0 Mashups* book, Apress Ed., pp. 395-416.