

**Jean-Philippe Gervais (INA) : "Le Web 3.0 sera sémantique"**

[Retour au sommaire de la lettre](#)

*Jean-Philippe Gervais est responsable de la filière multimedia à la direction de la formation de l'Ina (Institut National de l'Audiovisuel) et auteur du livre "Web 2.0, les internautes au pouvoir". Il revient dans cette interview sur de nombreuses thématiques et de nombreux usages touchant à la recherche d'information : les blogs, le "social search", la folksonomie, etc. pour arriver à une définition de l'"entreprise 2.0" qui s'orientera vers une prise en compte de nos souhaits par les moteurs de recherche qui mèleront à l'avenir sémantique et langage naturel...*

**Jean-Philippe Gervais**, auteur de l'ouvrage intitulé « Web 2.0, les internautes au pouvoir » (<http://www.dunod.com/pages/ouvrages/ficheouvrage.asp?id=50701>) paru aux éditions Dunod en 2007, répond à nos questions.



Responsable de la filière multimédia à la direction de la formation de l'Ina qu'il rejoint en 1997 après avoir été chef de projet multimédia sur de nombreux cédroms et sites web, Jean-Philippe Gervais est également co-auteur de *Gestion des médias numériques* (Dunod 2006), et a publié des articles dans le Film Français. Titulaire d'un DEA multimédia, il termine une thèse à la Sorbonne (Paris 1) sur les formes audiovisuelles émergentes sur les réseaux et est intervenant dans la Licence Documentation Audiovisuelle (INTD), le Mastère Création et Production Multimédia (Telcoms Paris/Ina) et Ina'Sup.

***Pouvez-vous nous dire quel est le sujet de votre livre ?***

L'idée à la base de cet ouvrage était de dresser à un moment donné, un panorama des usages du web 2.0. Ces usages concernent la géolocalisation, les réseaux sociaux, les blogs, la folksonomie et les moyens d'accès à l'information, le partage de vidéos et de photos. Le livre, rédigé en 2006, a été publié en février 2007. Il propose également une ouverture sur le thème de l'entreprise 2.0, c'est à dire la façon dont ces outils peuvent s'intégrer au sein de l'entreprise.

***Quelle est votre définition du web 2.0 ?***

Tim O'Reilly, le concepteur du web 2.0, a toujours souhaité ne pas en donner de définition. On peut juxtaposer une série d'adjectifs comme : collaboratif, participatif, personnel... etc. Ce sont en réalité les usages que font les internautes des services proposés, qui modèlent la réalité du web 2.0. Il se constate donc plus qu'il ne se définit. Il me semble ainsi plus important de parler des usages du web 2.0 que de sa définition conceptuelle.

De plus, en matière de recherche d'information, le web 2.0 offre de nouvelles modalités d'accès à l'information, comme la folksonomie, les nuages de mots ou la représentation graphique des informations (*voir les liens qui figurent à la fin de cet article*). Ces nouvelles façons de présenter l'information facilitent l'accès au sens et participent à la clusterisation de l'information.

***Pouvez-vous parler du partage de l'information sous forme de blogs et syndication de contenus et des effets que cela a sur la recherche d'information ?***

Il existe mondialement plus de 100 millions de blogs et on en compte 175 000 nouveaux par jour. C'est une information diluée que les moteurs traditionnels au début de la blogosphère indexaient plus ou moins bien. La raison est technique : le blog est généré par un système de *content management*. Le moteur Technorati s'est alors spécialisé dans l'indexation des blogs et l'accès à leur contenu. Souvent les internautes ne vont pas directement sur les blogs voir si un nouveau billet a été posté, mais s'abonnent à des flux RSS qu'ils reçoivent alors par mail. Il n'était donc pas simple pour un moteur de recherche d'accéder à ce type d'information. Désormais Google propose une interface en version beta d'accès aux blogs (<http://www.google.fr/blogsearch>) qui propose d'effectuer une recherche en temps réel sur les billets postés une heure, deux heures ou douze heures plus tôt. Face à cette concurrence, Technorati a alors cessé d'indexer uniquement des blogs et s'est progressivement orienté vers un type plus classique d'accès à l'information.

Aujourd'hui Google propose également aux utilisateurs qui émettent une requête de s'abonner aux flux qui concernent cette requête. Pour disposer de cette information en temps réel, le moteur en fait, s'abonne automatiquement aux flux RSS des blogs et les indexe quasiment en temps réel. La seule chose, en effet, qui différencie une page classique d'une page de blog est le flux RSS et le fait qu'une page de blog soit liée au temps. J'insiste particulièrement sur ce point dans mon livre car on n'a sans doute pas suffisamment pris en compte le fait que le blog inclut la dimension du temps. Chaque billet est daté, participe à la constitution d'archives dans le système de *content management*, ce qui n'est pas le cas sur une page classique. La seule différence entre une page classique et un blog est cette dimension temporelle qui ouvre de nouveaux horizons.

### ***Pouvez-vous nous parler des outils de folksonomie ?***

Folksonomie est la contraction de « folks », les gens, et « nomie » qui est la racine de taxonomie. La taxonomie constitue une hiérarchie de classements d'éléments les uns par rapport aux autres. La taxonomie la plus connue est celle de la classification des plantes par Buffon, chaque plante classifiée appartenant à une catégorie.

Et l'idée de la folksonomie est de parvenir à une classification de l'information réalisée par les internautes eux-mêmes. Les internautes qui toutefois ne sont ni des scientifiques ni des professionnels de la taxonomie, vont modestement mettre des étiquettes (ou tags) sur des éléments. C'est cela qu'il faut retenir de la folksonomie. Ces éléments peuvent être des articles de blogs identifiés soit par leur titre, soit par des mot-clés. La folksonomie devient plus spécifique au web 2.0 lorsqu'elle intéresse des médias de type photo ou vidéo. Les internautes taggent leurs images, leurs sons et leurs vidéos. Une description plus ou moins fine permettra alors d'identifier et d'accéder à ces contenus grâce à ces tags ou mot-clés. Les limites de la folksonomie sont dans le fait que tous les internautes ne jouent pas forcément le jeu.



Une nouvelle folksonomie qui tourne autour de la géolocalisation est en train de voir le jour. Il est aujourd'hui possible sur certains sites de partage de photos de localiser directement les lieux photographiés. La procédure est simple. Si j'ai une photo de la Tour Eiffel, il suffira de la déposer sur une carte géographique au bon endroit pour que le tag s'insère automatiquement. Au-delà de la géolocalisation des photos on voit maintenant aussi se développer aussi des usages de géolocalisation d'objets ou de personnes soit par GPS, soit à l'aide de puces RFID. Il est évident qu'avec la diffusion massive des puces GPS dans les téléphones portables ce type de géolocalisation est amené à se développer. Il est ainsi possible de retrouver des personnes partout en France grâce, par exemple à des sites comme Ootay (<http://www.ootay.fr/Accueil.asp>). Les puces GPS sont aujourd'hui d'une taille inférieure à celle d'une pièce de un centime d'euro. On peut les mettre dans un cartable ou même dans une chaussure. Ces puces sont même utilisées aujourd'hui par exemple pour vérifier que les coureurs du marathon de New-York n'ont pas triché par rapport au tracé du parcours. Il est vrai qu'avec la généralisation de ce type de technologies, on est en droit de penser que Big Brother n'est plus très loin.

### ***Les outils de folksonomie proposent-ils un accès fiable à l'information ?***

La folksonomie qui consiste techniquement en l'accès à des données par le biais de mot-clés pose un problème pour les moteurs, car l'inexactitude de certains tags met en danger la viabilité des informations.

Si sur Daily Motion vous recherchez une vidéo sur une personnalité, cela peut vous poser un problème. Certains internautes se contentent, en effet, d'identifier la vidéo par le prénom, d'autres par le surnom, voire par un nom qui n'a aucun rapport avec le contenu. Les fautes d'orthographe

posent aussi un véritable problème et pour l'instant les moteurs ne disposent pas de listes d'autorité portant sur des entités nommées. Quand l'INA ou une autre entreprise audiovisuelle indexe ses images, elle utilise des procédures spécifiques au traitement des noms propres. Aucune erreur ne peut être commise sur les noms propres grâce à ces listes d'autorité. Mais sur Daily motion ou You tube, une recherche sur une personnalité du spectacle ou politique donne lieu à des variations importantes sur la façon de nommer cette personne et les sites de partage de vidéos ou de photos n'ont pas mis en place des procédures d'indexation suffisamment strictes. Un mode d'accès aux requêtes les plus formulées chaque mois pourrait déjà simplement permettre d'accéder à des contenus plus pertinents au regard d'une actualité. Les sites de partage de photos sont toutefois en train de pallier à ce déficit en information grâce à la géolocalisation.

### ***Quel est l'avenir de la folksonomie ?***

Le défi de la folksonomie est d'arriver à "hiérarchiser" et "définir" les tags comme dans une taxonomie classique, ce qui paraît antagoniste avec la définition même de la folksonomie où chacun est libre d'étiqueter comme bon lui semble. C'est le fameux web sémantique ou web 3.0. La première étape pour les moteurs, est d'identifier les tags en tant que tel, c'est le cas par exemple du moteur Icerocket (<http://blogs.icerocket.com/tag/>).

Ce moteur indexe les tags sur les blogs, grâce à l'utilisation de "micro formats", technologies qui permettent l'interopérabilité, ou échange des données. Bien sûr tout n'est pas encore standardisé, mais à court ou moyen terme, les moteurs devraient être capables d'isoler les tags. Si l'on se souvient que Flickr appartient à Yahoo! et YouTube à Google, l'optimisme est de mise !

### ***Et le social search ?***

Le social search s'appuie sur un des paradigmes du web 2.0 : l'intelligence collective, qui veut que la somme des savoirs soit supérieure à celle de chaque individu pris séparément.

Néanmoins, dans notre pratique quotidienne, le confort de la recherche d'information sur les moteurs est assuré par la réactivité quasi instantanée de ces derniers, donc on imagine mal revenir à un système où il faudrait patienter pour obtenir des informations.

Par contre, dans le cas de recherche plus générique sur des thématiques données, disposer de réponses formulées par des internautes et non d'une simple url est véritablement un plus. A ce titre Yahoo Answers (<http://answers.yahoo.com/>) est intéressant. Ce qu'il faudrait réussir c'est l'équivalent de l'Open Directory ([www.dmoz.org](http://www.dmoz.org)) version web 2.0. Le social bookmarking est déjà une première étape dans le social search, l'avenir est peut être aussi du côté des réseaux sociaux comme Facebook où les internautes "jouent" le jeu de partager des infos.

### ***Vous évoquez l'identité numérique, de quoi s'agit-il exactement ?***

C'est un concept que j'évoque dans mon livre. Lorsqu'un utilisateur poste des photos et des vidéos et se met en scène sur le web il construit une identité numérique qui lui est propre et la construction de cette identité numérique passe par le fait d'assumer les effets produits par les documents personnels mis en ligne sur internet. Par exemple, les universités américaines mettent les étudiants en garde à propos des informations qu'ils peuvent poster et qui mettent en scène leurs éventuels excès en tant qu'adolescents et qui demeureront en ligne le jour où ils rechercheront un emploi. Cela met en valeur la nécessité d'une éducation des acteurs du web à la construction de cette identité numérique, car toute intervention laisse des traces. Aujourd'hui ce qui est dit et posté sur le web, que ce soit sous sa propre identité ou sous un pseudo, demeure et les technologies, quel que soit le cas, sont capables de remonter à la source et seront de plus en plus à même de localiser les personnes qui se sont exprimées. Cela peut dérapier si, il n'y a ni contrôle, ni autocensure du système, ce qui n'est heureusement pas toujours le cas.

### ***Cette identité numérique est en quelque sorte une seconde vie des acteurs du web ?***

Une doctorante américaine de l'université de Berkeley, Danah Boyd (<http://www.danah.org/>) travaille sur la construction psychologique du monde qu'est susceptible de réaliser un adolescent, avec les outils que le numérique met à sa disposition. L'adolescence est une mutation de la perception du corps et de l'environnement. Avant cela passait parfois par des conduites à risque. On observe depuis quelques années une substitution des conduites à risque par une mise en scène de la construction intérieure en ligne. Il faut éviter de prendre cela trop au premier degré en

renvoyant aux jeunes quelques années plus tard, l'image produite par les mises en scène qui correspondaient à cette phase de construction de leur identité de futur adulte. Il s'agit aussi d'une construction du lien social. L'identité n'est pas uniquement représentée sur un mode « avatar » comme c'est le cas dans « second life », mais par la mise en scène du vrai « moi » avec ou sans pseudo. Il s'agit d'une communication à la fois synchrone et asynchrone. Celui qui possède un blog et un espace sur « myspace » est tout le temps présent, tout le temps en communication même si l'internaute n'est pas en temps réel sur son ordinateur. Il y a de plus une dimension locale, lorsqu'on convoque ses amis et connaissances, et un prolongement parfois dans la vraie vie des gens rencontrés en ligne. Il existe certains groupes de musiciens dont le public en concert est presque exclusivement constitués de fans qui les ont connus sur « myspace ».

***Comment les moteurs de recherche gèrent-ils la fantaisie du langage de type « texto » qu'on trouve parfois sur les blogs ?***

Je crois que ces variantes orthographiques sont très peu prises en compte par les moteurs de recherche. Cela dit, certains moteurs, comme Google ou Exalead suggèrent une autre orthographe. C'est pourquoi il serait important de proposer des listes d'autorité ou une meilleure gestion des entités nommées. Une fonctionnalité utile serait que des sites de mise en ligne de contenus comme Flickr ou Daily Motion proposent une variante orthographiques aux utilisateurs qui taggent les contenus. Corriger une orthographe défectueuse à la source serait une façon astucieuse de contourner le problème de la recherche d'information en orthographe erronée. Cela permettrait un meilleur accès au contenu, car ce problème est aujourd'hui générateur de bruit. Aujourd'hui sur Daily Motion, si un utilisateur ne met pas de tag pour identifier une vidéo, le système reprend automatiquement le nom du fichier. Ainsi une vidéo dont le titre serait « toto.mov » serait taguée « toto » sur le système, ce qui n'est en rien annonciateur de son contenu. De même sur Flickr certaines photos sont identifiés par le nom par défaut de type « DSC 00234 » donné par l'appareil photo.

***Pouvez-vous nous parler de l'entreprise 2.0 ?***

C'est un concept très nouveau. Le particulier en effet, dispose souvent d'un accès plus riche à internet chez lui que dans sa propre entreprise. Le concept d'entreprise 2.0 transpose ainsi les usages des particuliers en les adaptant à l'entreprise. Les usages d'internet en entreprise sont souvent centrés essentiellement sur le mail. L'un des usages les plus courants en entreprise est également la recherche d'information sur internet. De plus, les informations disponibles sur un intranet sont souvent des informations descendantes avec très peu de remontées de la part des utilisateurs. Les usages du web en entreprise sont le plus souvent des usages de type web 1.0. L'entreprise 2.0 est un concept qui permet aux usagers des entreprises de bénéficier de tous les usages collaboratifs et participatifs au sein de l'entreprise dans laquelle ils travaillent. L'usage participatif le plus courant en entreprise est le wiki. Il s'agit de la transposition du mode encyclopédie collaborative aux usages de l'entreprise. On observe donc une remontée des informations de la part du terrain. Ainsi des commerciaux peuvent enrichir une base wiki sur les retours clients ; les ingénieurs est la chaîne de production enrichissent également ces données et l'ensemble constitue alors une masse d'information beaucoup plus importante sur un produit donné que celles disponibles sur la fiche ou notice du produit.

Une seconde dimension est celle des blogs d'entreprise ou intra blogs. Les salariés s'expriment sur une thématique donnée. Ces blogs peuvent aussi être utilisés en mode extranet. Il s'agit alors de réaliser un blog pour un client donné. Il s'agit alors de faciliter la communication avec le client sur une thématique donnée. Il existe aussi des blogs d'entreprise ouverts vers l'extérieur. Il existe ainsi un blog de Mc Donald sur la gouvernance d'entreprise. Par exemple, Mc Donald distribue des petits jouets dans menus pour enfants. A un moment donné des minis 4X4 ont été distribués aux petits garçons. Ces voitures sont mal vues par les lobbies écologistes en raison de la pollution qu'elles génèrent. Les utilisateurs ont réagi sur ce thème. L'entreprise a ainsi retiré les 4X4 incriminés de ses menus enfants et communiqué sur ce thème auprès de ses clients. Il s'agit d'un retour client immédiat. Aujourd'hui tout le monde a les moyens de connaître les conditions de production d'un produit et le blog est un moyen de réaction rapide et efficace sur les sujets sensibles.

***Pour quand l'avènement du web 3.0 ?***

Le web 3.0 sera un web sémantique. Tim Burners Lee explique que la notion de web sémantique a été interprétée dans le mauvais sens dans la mesure où les usagers ont imaginé qu'il s'agissait de

créer du sens. Or il s'agit plutôt d'un web de données qu'il faut organiser et structurer pour mieux les croiser, qu'il s'agisse de texte ou de données multimédia. Il faudrait pouvoir préparer par exemple un voyage en exprimant simplement ses besoins dans un moteur de recherche, et ceci de façon très simple : « j'ai deux enfants, l'un aime le tennis, nous voulons être à côté de la plage pour le petit qui a cinq ans et fait des châteaux de sable » etc. Ainsi le véritable web sémantique me permettrait de disposer d'une offre de voyage correspondant réellement à mon besoin à partir d'une requête, guidée ou non, exprimée simplement en langue naturelle.

Cela commence à exister dans les moteurs de comparateurs de prix qui ont travaillé très finement sur le classement des objets, de leurs fonctionnalités et utilisations. On retrouve ces classifications fines aussi dans le monde du voyage ou de l'industrie pharmaceutique. Cela ne se fera certainement pas en un jour car il est indispensable pour que tout cela fonctionne, de créer des définitions mondiales des objets de recherche et d'inciter les acteurs à prendre conscience de l'importance de la mise à disposition de ces informations. Le web 3.0 consistera donc à organiser toutes ces données disséminées de façon éparses au travers des outils mis à disposition des usagers, pour mieux les croiser et les mettre à disposition de ceux-ci.

Nous sommes aujourd'hui dans une économie de la méta donnée. Le contenu est important en tant que tel, mais ce qui est plus important encore, ce sont les modalités mises en place pour l'accès à ce contenu et la façon dont il peut être contextualisé par rapport à un environnement économique ou juridique ou social. C'est cette contextualisation dans le cadre du web 3.0 qui permettra de passer d'une économie de la méta donnée à une économie de la méta langue.

**Merci, Jean-Philippe Gervais**

### **Liens complémentaires**

Interface innovante autour des "tags" :

Les travaux de Jonathan Harris

<http://www.fredcavazza.net/2007/09/19/connaissez-vous-lergonomie-artistique/>

<http://labs.digg.com/> et principalement l'essai <http://labs.digg.com/swarm/> avec une interface à la "kartoo"

Interface recherche image dans flickr par dessin

<http://labs.systemone.at/retrievr/>

Toujours flickr, mettre un mot par exemple "sand" et les mots associés apparaissent en images

<http://www.quasimondo.com/tagnautica.php>

Mashups :

<http://www.programmableweb.com/>

### **Interview réalisée par Marianne Dabbadie**

*Directrice Innovation i-KM*

*Laboratoire GERIICO – EA 1060*

**Réagissez à cet article sur le blog des abonnés d'Abondance :**

<http://abonnes.abondance.com/blogpro/2007/11/jean-philippe-gervais-inale-web-30.html>