

LES ENJEUX D'UNE BIBLIOTHÈQUE SUR LE WEB

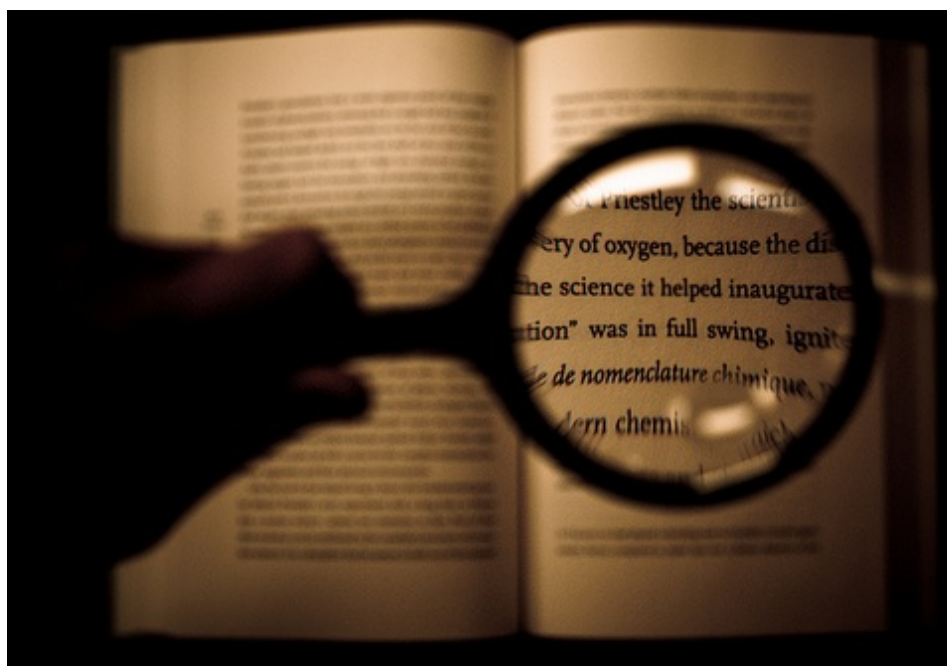
LE 10 JUIN 2010 CHRISTIAN FAURÉ

Comment utiliser les métadonnées pour permettre aux bibliothèques de mieux mettre en valeur les ressources ? Par la mise en place d'"orages sémantiques" captant l'ensemble des discussions, polémiques, argumentations sur un sujet.

La plupart des débats sur le rôle et la présence des bibliothèques sur web s'est jusqu'à présent concentré fortement sur les enjeux de numérisation des fonds et sur son corolaire de diffusion des œuvres numériques. Autre thème, mais dans une moindre mesure (surtout à l'échelle du grand public) : la mise en ligne des catalogues et la disponibilité des métadonnées des institutions qui les gèrent. Ces questions sont importantes, mais j'aimerais en proposer une troisième, qui s'appuie sur l'exposition des métadonnées, et constitue à mes yeux la clé de voute pour une stratégie des bibliothèques sur le web.

1. Avoir des ressources facilement identifiables et consultables sur le web

Chacun aura remarqué qu'il n'est pas facile de faire un lien vers l'URL d'un livre qui pointe vers les catalogues en ligne que proposent les bibliothèques. Moi-même, quand je parle d'œuvres ou d'auteurs, je privilégie un lien vers la page associée chez Amazon ou Wikipedia, c'est tellement plus rapide. Mais, à part amener du trafic à ces sites, ce geste de faire un lien n'est pas valorisé en lui-même. Une bibliothèque doit donc commencer par augmenter la visibilité de ses ressources pour ensuite pouvoir donner de la valeur à mon lien.



L'exposition des métadonnées passe par la mise à disposition d'URLs pérennes pour l'ensemble des ressources signifiantes. Ces URLs doivent donc être pérennes (je vous renvoie chez **figoblog** pour la littérature et les réflexions associées), mais aussi facilement identifiable (voire inférençable : par exemple, je peux déduire l'URL d'un auteur à partir de celle d'un autre), ainsi que déréféençables (cf. **Qu'est-ce qu'une URI déréféençable ?**). Point d'autre salut dans cette démarche que d'embrasser les standards du web sémantique et d'emboîter le pas à l'initiative **Linked Data**.

2. Exploiter la vie de ces ressources sur le web

On n'envoie plus un satellite dans l'espace pour pouvoir dire « je sais le faire », mais pour exécuter une mission bien précise qui repose sur la collecte et la récupération de données.

Par analogie, **c'est cet objectif qui manque souvent aux stratégies d'ouverture et d'exposition des données**. Je suis le premier à dire qu'il faut mettre à disposition les données sur le web, tout comme je suis le premier à dire qu'il faut du très haut débit. Mais si ces initiatives sont nécessaires, elles n'en sont pas pour autant suffisantes, car elles ne sont que les prémices d'une politique, il manque encore quelque chose. C'est notamment la raison pour laquelle je terminais mon texte dans « **Pour en finir avec la mécroissance** » en écrivant :

“

« Pire peut-être, ces politiques se limitent à favoriser l'accès à Internet et à offrir des débits de connexion toujours plus importants, mais pour quoi faire ? » p. 278.

”

Je crois qu'il faut partir de là : accéder à des contenus ou des données sur le web, d'accord, mais « pour quoi faire ? » Je balaye ici d'un revers de main toute la rhétorique qui consiste à dire : « ouvrez vos données et vous verrez ce sera formidable ». Oui, je crois effectivement que ce sera formidable, mais ce discours ne favorise pas la prise de décision pour celui qui doit faire l'effort : il faut donc trouver d'**autres motivations**.

La logique et la stratégie de l'accès orientent trop souvent les débats autour d'une **stratégie de diffusion**. C'est très bien de diffuser, surtout quand il s'agit de ressources culturelles comme celles que gèrent les bibliothèques, mais je pense que ce n'est pas là que les choses se jouent. La plupart des bibliothécaires ne peuvent pas se satisfaire de concevoir **le web comme un simple outil de communication et de diffusion**. Le métier de bibliothécaire croule tellement sous la charge de travail que représente la politique d'acquisition, de conservation, de gestion des catalogues, d'archivages, plus tous les aspects fonctionnels d'accueil du public, de relation avec les chercheurs etc. que la perspective du web comme nouvel outil de diffusion ne représente à leur yeux qu'une contrainte de plus qui va surtout donner lieu à une énième refonte du système informatique.

Faire tous ces efforts **simplement parce que cela semble inéluctable** n'est pas la meilleure des motivations qui soit. On revient donc à notre question « pour quoi faire ? », car si la réponse est « parce que c'est nécessaire », on peut être sûr que c'est l'inertie qui va s'installer (pourquoi faire des efforts si la situation est tellement inéluctable ?).

Il faut donner une motivation aux bibliothèques : pénétrer dans l'économie générale du web, cela ne doit pas représenter uniquement un surplus de travail et d'effort, même si la diffusion et la valorisation sont des missions essentielles d'une bibliothèque.



Concernant les bibliothèques, ma

proposition sera donc la suivante : il faut développer les « **orages sémantiques** ». Par cette expression on entend l'ensemble des discussions, polémiques, argumentations autour d'une ressource (auteur, oeuvre, thème, etc.). Dans cette perspective, il faut considérer que **chaque ressource disponible en ligne est un paratonnerre** dont le but est de capter les polémiques et les discussions dont elle fait l'objet.

Avec cette approche, l'activité de catalogage s'étend **au-delà du catalogage des œuvres** puisqu'il couvre le **catalogage des débats sur autour des ressources sur le web**. Grâce à ce catalogage des « orages sémantiques », une bibliothèque peut commencer à fournir de nouveaux services, comme par exemple une sorte de « Zeitgeist », un esprit du temps.

Aujourd'hui la Library of Congress archive, plus qu'elle ne catalogue, les messages publiés sur la plateforme de Twitter. La question qui est posée aux bibliothèques est la suivante : souhaitez-vous être condamnées à négocier avec des acteurs privés **le catalogage des orages sémantiques** via leur plateforme commerciale, ou souhaitez-vous développer vous-même ces dispositifs ? C'est-à-dire être pro-actif dans la conception de ces dispositifs pour réinventer le catalogage et les services d'une bibliothèque sur le Web. Voulez-vous n'être que des archivistes sous-traitants des plateformes commerciale ou des promoteurs d'une politique économique-culturelle de la contribution ?

Parlant de **contribution**, je me dois ici de préciser que je ne parle pas de crowdsourcing, de site participatifs ou autres espaces personnalisés de contribution dont pourrait se doter un site web de bibliothèque. Il ne s'agit pas de ici de rajouter des fonctions de tags ou des folksonomies car le squelette de la démarche repose sur **l'autorité des métadonnées de la bibliothèque. Le dispositif de captation des orages sémantiques doit reposer sur les acquis des catalogues et des notices d'autorités pour faire la révolution copernicienne du catalogage** : elle ne se fait pas en marge de lui ou contre lui mais avec lui, avec ce **trésor des métadonnées**.

Ne cherchez pas l'outil magique pour faire cela, je crois qu'il n'existe pas, et il reste à faire. Il y a en fait deux dispositifs qui peuvent répondre à cet enjeu des orages sémantiques :

Le premier est un **dispositif indirect** basé sur l'analyse traces, celui dont je parle ici. Le deuxième est un **dispositif direct** basé sur des outils critiques offrant des fonctionnalités d'annotation et de traçabilité des polémiques (le modèle que j'ai en tête étant les outils de gestion des sources dans les projets de développement informatique). De celui ci je ne parle pas dans cette note.

La démarche indirecte doit se construire à partir de **plusieurs briques fonctionnelles** :

du « **web analytic** » au travers de l'ensemble des services qui permettent d'analyser des trafics de sites web ;

du **text-mining** pour dégager des métadonnées des sources qui pointent vers les ressources exposées de la bibliothèque ;

du **data-mining**

des **technologies d'indexation**

bien sûr, si l'effort a été fait d'avoir les **données structurées en RDF**, la granularité des informations de consultations et de requêtes n'en sera que plus fine et plus facilement exploitable.

Il s'agit donc bien d'un **panaché de technologies** pour ne pas avoir en sortie un simple *hit parade* ou un moteur de recommandation à la Amazon. De plus, les expérimentations de publication des statistiques brutes de consultations ont montré que c'est *Mein Kampf* ou *Le Kamasutra* qui vont trôner en haut des classements. Je précise également qu'il ne s'agit pas uniquement d'utiliser des informations de consultations, mais surtout **des informations provenant de la source des liens** qui pointent vers ce lieu de référence qu'est une bibliothèque (pensez ici à BackRub, l'ancêtre de Google).

Un petit résumé de ce qui change avec cette approche des orages sémantiques :

Les bibliothécaires et conservateurs en « back office » ont un *feed-back* de ce qui se passe sur le web autour des ressources dont ils ont la gestion. Retour appréciable, me semble-t-il, quand on doit gérer l'évolution de son catalogue : les orages sémantiques placent le bibliothécaire au coeur de son temps, des polémiques et de ce qui fait débat ;

La bibliothèque peut ajouter des services innovants sur son site web. Par exemple, je n'irai pas spontanément sur la page d'un auteur sur le site web d'une bibliothèque mais plutôt sur Wikipedia à cause de la richesse des informations. Mais si la page en question me donne une « météo culturelle » de cet auteur, alors cela peut changer mes pratiques ;

enfin, au delà du *Zeitgeist*, c'est une extension de la pratique de catalogage qui révolutionne l'activité d'une bibliothèque en la plaçant **au coeur de l'économie générale du web**.

Je termine par un dernier point car, les bibliothèques, en ces temps budgétaires difficiles, cherchent à augmenter leurs fonds propres, or je serai le premier à acheter une oeuvre (papier ou numérique) qui comprenne un appareil critique issu des informations collectées

via le dispositif des orages sémantiques géré par une institution publique telle qu'une bibliothèque. Cela changerait à coup sûr les rapports entre les bibliothèques et les éditeurs, ces derniers ayant beaucoup trop tendance à les mépriser. C'est d'ailleurs quelque chose qui m'a frappé en m'intéressant à **la chaîne du livre** : l'indifférence et le mépris règne entre les acteurs ce milieu tandis que de nouvelles industries arrivent et raflent la mise, le sourire en coin.

J'espère que les bibliothécaires me pardonneront mon ingérence dans leur domaine de compétence.

—

Billet initialement publié sur le blog de **Christian Fauré** ; images CC Flickr **beelaineo**, **Reini68**, **jakebouma**

2 pings

Les tweets qui mentionnent Les enjeux d'une bibliothèque sur le web » Article » owni.fr, digital journalism -- Topsy.com le 10 juin 2010 - 11:08

[...] Ce billet était mentionné sur Twitter par Silvae, Frédéric Guerrier, Xavier BESSON, Lecteur du Val, Michèle Battisti et des autres. Michèle Battisti a dit: RT @owni: [#owni] Les enjeux d'une bibliothèque sur le web <http://goo.gl/fb/Ant2G> [...]

links for 2010-06-12 « Ikan66 le 12 juin 2010 - 17:02

[...] Les enjeux d'une bibliothèque sur le web » Article » owni.fr, digital journalism (tags: bibliothèque web) [...]