



## Madame Machine, pouvez-vous me conseiller un bon livre ? : les nouveaux outils Web de recommandation de lectures...

Date de publication : 27/06/2011

Jusqu'à ce qu'Amazon<sup>1</sup> lance ses « Ceux qui ont acheté ce livre ont également acheté... » et « Alexandre, bienvenue sur Votre Amazon.fr », les recommandations personnalisées constituaient la chasse gardée des libraires, des bibliothécaires... et des copains. Désormais, votre meilleur ami pour les conseils de lecture se nomme « Al Gorithme de recommandation »... et il vous veut du bien !

En réalité cela fait déjà un certain temps qu'en utilisant un mot-clef de la description bibliographique d'une monographie, vous pouviez faire afficher à l'OPAC une liste d'autres documents sur le même sujet (ou de manière plus générale de la même catégorie). D'abord il a fallu resaisir le mot ou l'expression dans la zone de recherche – ce qui fut ô combien moins fastidieux lorsque la technique du copier-coller est apparue ; ensuite, il a suffi de cliquer sur le mot-clef, devenu avec l'évolution technologique « un hyperlien », pour effectuer un « rebond » et le tour était joué. Oui mais...

Oui mais si les mots-clefs liés au sujet sont fréquents (en RAMEAU, le plus souvent) au sein de nos catalogues, ceux liés aux genres littéraires et aux nationalités de littérature le sont beaucoup moins, sans doute parce qu'il n'y a pas de standard faisant l'unanimité en ces domaines. Pour les œuvres de fiction, c'est tout à fait regrettable, d'autant plus que lorsqu'il y en a, les catégories sont souvent trop grossières : les romans noirs, les thrillers et les enquêtes policières sont fourrés dans le « policier » alors que la dystopie, le space opera, l'uchronie et – pire ! – même la fantasy aboutissent dans le grand sac « science-fiction ». Au demeurant, que ce soit pour les genres littéraires ou les nationalités de littérature, certains recourent à RAMEAU, d'autres à la classification Dewey (voire CDU) et d'autres encore à un fichier d'autorité « maison ».

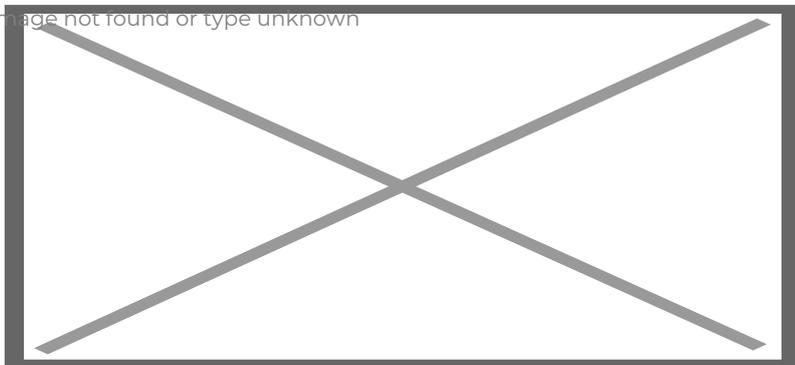
Par ailleurs, on ne peut pas dire que les rebonds liés aux sujets proposés au sein des notices de nos catalogues soient toujours satisfaisants : les limites de RAMEAU (et d'autres thesaurus matière « savants ») ne sont que trop connues. Heureusement l'indexation en langage naturel a connu un considérable essor avec l'arrivée de la « folksonomie » et certains ont fait le choix raisonnable de proposer une liste de mots-sujets mixtes à leurs usagers. Voilà qui est déjà mieux ; cela reste toutefois insuffisant. Sans parler des genres littéraires et des nationalités de littérature...

### Une nouvelle offre à nos (anciens et futurs) lecteurs : les œuvres « proches »

**Amazon** donc, outre qu'il propose sur le marché anglo-saxon l'une des liseuses parmi les plus convaincantes, le Kindle, a fait preuve d'une grande innovation il y a quelque temps avec ses recommandations de lecture, qu'elles soient liées à un titre ou adaptées à un lecteur (et client, en l'occurrence !).

Commençons par analyser les recommandations établies au départ d'un titre d'ouvrage précis.

Image not found or type unknown



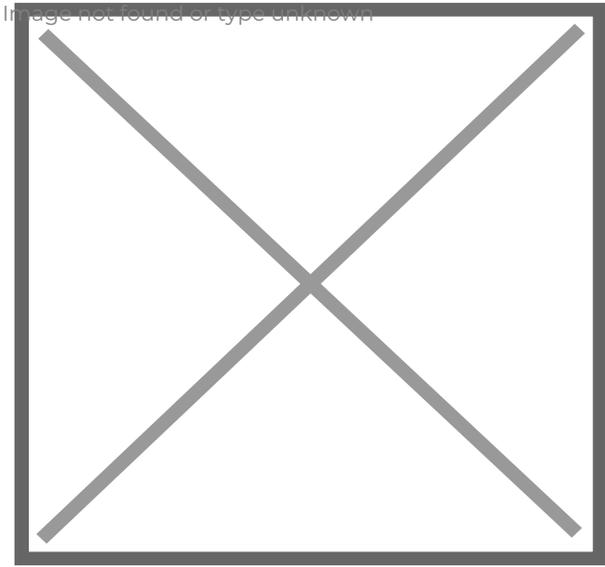
Les recommandations de titres « proches » de Fahrenheit 451 sur Amazon.fr

Le résultat est plutôt concluant sur Amazon : pour *Fahrenheit 451* de Ray Bradbury, l'algorithme de recommandation propose notamment deux autres célèbres dystopies (sous-genre de la SF parfois nommé aussi « contre-utopies ») : *Le meilleur des mondes* et *1984*.

Les récents réseaux sociaux de lecture, comme Librarything dans le monde anglo-saxon (une version française existe mais son catalogue de titres en français est encore peu étoffé à ce jour) et Babelio<sup>2</sup> dans le monde de la littérature française se sont lancés dans ce type de recommandations eux aussi.

Il me paraît, avant d'aborder les recommandations automatisées de ces réseaux sociaux, utile de rappeler une source d'inspiration traditionnelle qui trouve son pendant sur le Net : l'espionnage consenti de la bibliothèque d'un copain (ou d'une copine) dont on sait qu'il / elle a des goûts proches des nôtres. En fait, les réseaux sociaux permettent même d'aller bien au-delà : on peut désormais connaître les membres qui partagent le plus de lectures avec nous...et ensuite faire les voyeurs chez ces inconnus qui nous ressemblent !

**Babelio** nous propose un outil très intéressant dans cette perspective (et plus, si affinités !) : le « comparateur de lecteurs ». Lorsque vous avez constitué au sein du réseau votre bibliothèque virtuelle (les livres de votre bibliothèque physique ou la liste des livres que vous avez lus ou vos livres fantastiques préférés, ou...), vous pouvez demander au système d'afficher les lecteurs les plus proches de vous. La comparaison se fait sur base du recouvrement entre vos bibliothèques virtuelles respectives et aussi sur base des notes que vous avez attribuées aux livres de cette intersection. Ensuite, vous pourrez aller « mater », si votre alter ego littéraire joue le jeu de l'exhibitionnisme encouragé par ce type de réseau, en tout bien tout honneur bien entendu !

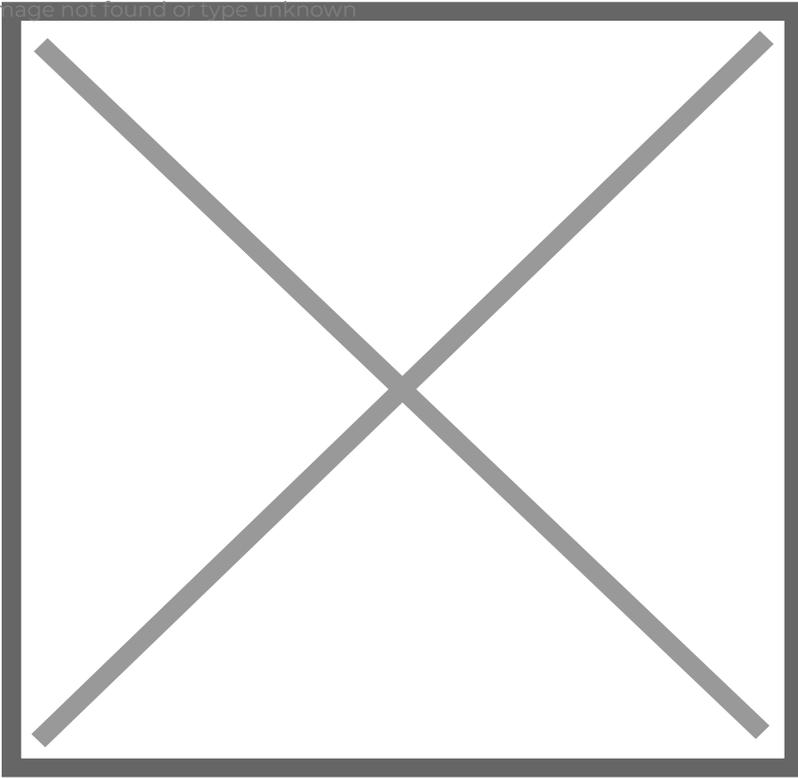


Les recommandations de titres « proches » de *Fahrenheit 451* sur Babelio

Les recommandations par rapport à un titre précis du réseau Babelio montrent elles aussi une certaine pertinence (relevons notamment au passage, pour les amateurs, l'excellente dystopie de Robert Silverberg, *Les monades urbaines*). On ne retrouve toutefois pas ici, les deux grandes dystopies d'Orwell et Huxley car la recherche a été faite en mode connecté et que le membre a signalé dans sa bibliothèque les avoir déjà lues. On notera par ailleurs l'intéressante option proposée en bas de liste : « Signaler une suggestion hors sujet ». Le réseau fait donc ici appel, selon l'expression consacrée du Web social, à « l'intelligence collective », dans un contexte où cela apparaît comme tout à fait raisonnable.

**Librarything** recourt lui aussi d'ailleurs à l'assistance de ses membres, en proposant une seconde liste de recommandations, en sus de celle qui est constituée automatiquement : les recommandations directes des membres par rapport à ce titre. Elles sont classées par ordre décroissant du nombre de validations effectuées par d'autres membres pour chaque titre proche proposé.

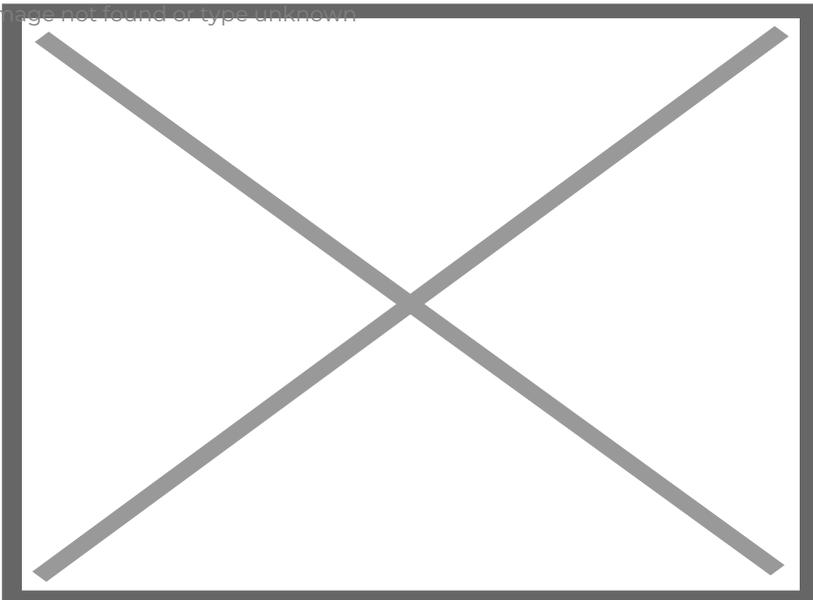
Image not found or type unknown



Les recommandations de titres « proches » de *Fahrenheit 451* sur Librarything (le site en anglais, ici)

Bon, très bien, et donc, on surfe avec deux fenêtres ouvertes (ou deux onglets, pour éviter les courants d'air !) en passant allègrement d'un site tel que ceux évoqués supra vers l'OPAC Web de sa bibliothèque ou du catalogue collectif régional. Comme on peut le lire sur le site d'une société travaillant sur les enrichissements de catalogues par du contenu Web 2.0 : « vous connaissez la réponse à cette question ! ». Un peu d'aide au cas où vous ne seriez pas bien réveillé : non, l'internaute souhaite voir tout au sein du même site. Ah d'accord, ben alors on l'intègre dans le catalogue, pas de souci, on avait déjà les SIGB et les portails, on a l'habitude d'intégrer. Allez zou, c'est parti, voici le catalogue de la Médiathèque de Toulouse<sup>3</sup> :

Image not found or type unknown



A droite, les recommandations de titres « proches » de Fahrenheit 451 dans le catalogue de la Médiathèque de Toulouse

La Médiathèque de Toulouse a en effet adhéré à « Babelthèque », une offre qui permet, via un Web service<sup>4</sup>, d'aller chercher à la volée dans Babelio les recommandations liées à un titre affiché dans le catalogue de la bibliothèque au départ de l'envoi de l'ISBN de cette édition de l'œuvre. Il est même possible si la bibliothèque fournit la liste des ISBN des éditions présentes dans son catalogue de mettre en œuvre un filtre qui ne retiendra pour l'affichage que les titres d'œuvres proches présents dans le catalogue de la bibliothèque. Ainsi si « Les monades urbaines » par exemple n'est pas présent dans le catalogue de la bibliothèque cliente, ce titre ne s'affichera pas parmi les titres conseillés dans le catalogue de la bibliothèque (alors qu'il s'affiche, on l'a vu, dans Babelio même, qui est la source de cet affichage).

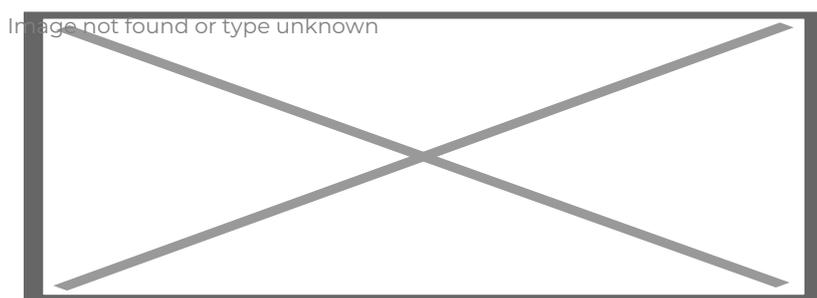
Par contre, 1984 d'Orwell n'apparaît pas dans cette liste alors qu'une requête parallèle effectuée a permis de relever sa présence dans le catalogue de la Bibliothèque de Toulouse. Aurait-il été victime du terrible système de censure de l'état totalitaire du roman ?!

On notera enfin que l'OPAC affiche également les commentaires de membres du réseau social de lecture auquel la médiathèque est affiliée (Babelio, ici) et la moyenne des notes attribuées par ces membres ainsi que les étiquettes apposées par leurs soins, deux informations fort intéressantes mais sur le titre affiché lui-même cette fois.

#### « Bonjour chez vous, numéro 6 ! »

C'est à nouveau à l'inventivité d'Amazon que nous ramène la petite phrase de politesse rabâchée dans cette vieille mais excellente série anglaise de et avec Patrick Mac Gohan, « Le prisonnier ».

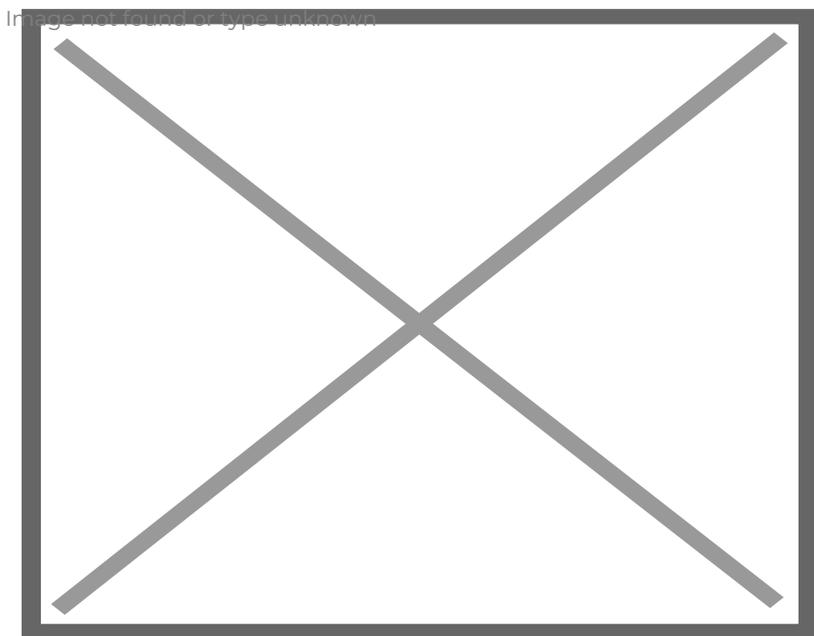
En effet, si vous vous inscrivez comme membre d'Amazon et que vous vous identifiez comme tel par la suite, vous serez salué tout de suite d'un tonitruant (on imagine, c'est muet) « Bonjour Victor Hugo » (à supposer, pour l'exemple, que vous vous prénommez Victor et que votre patronyme soit Hugo). Un peu plus bas on vous propose un hyperlien « Chez vous » et par contraction de ces deux expressions, vous vous sentez tout de suite « au village »<sup>5</sup> !



Les recommandations personnalisées d'Amazon apparaissent sous la forme d'une liste dans laquelle on peut naviguer horizontalement ; à noter : la possibilité de corriger une recommandation

Toute plaisanterie au tiroir, la personnalisation proposée en aval de l'identification n'a rien d'une prison, en fait. On vous propose des livres spécialement pour vous, des DVD qui pourraient avoir l'heur de vous plaire et toutes sortes d'autres propositions sur mesure dont des nouveautés et même des livres à paraître qu'on croirait édités uniquement pour nous...

Mais comment font-ils pour connaître si bien nos goûts ? Les fameuses caméras du village du prisonnier seraient-elles parvenues jusque dans nos chaumières ? C'est un peu ça, en fait, si ce n'est qu'ici la caméra vous filme uniquement dans la librairie, qu'elle ne filme que la bibliothèque des livres que vous avez acheté chez ce libraire en ligne, que vous pouvez montrer à une caméra la liste de vos « envies d'achat » ainsi que la liste des livres que vous aimez et vous désinscrivez du site quand vous voulez. Certains sites de librairies en ligne présentent donc désormais une dimension de réseau social simplifié. Les libraires aussi ont compris que les lecteurs préfèrent voir tout au même endroit (et eux préfèrent que cet endroit soit leur librairie !)..



Les recommandations personnalisées d'Amazon pour les nouveautés et les documents à paraître

Ainsi, ce type de site ne fait pas que retenir ce que vous lui avez acheté, il vous propose de lui fournir des informations lui permettant de mieux cerner vos goûts de lecture. Et, sauf si vous estimez que cela vous pose un problème de vie privée – ce qui est votre droit évidemment bien qu'ici cela porte sur des données relativement peu sensibles – vous avez tout intérêt à le faire afin qu'on vous propose des documents susceptibles de vous plaire.

Amazon va même encore plus loin dans l'affinage par vos soins de votre propre profil de « consommateur » de produits culturels : il vous offre la possibilité de spécifier, pour chaque livre que vous avez acheté, indiqué comme livre que vous aimiez ou comme livre qui vous fait envie, que vous ne souhaitez pas qu'il soit pris en compte dans l'élaboration de votre profil personnel de lecteur. C'est un pas supplémentaire important car c'est le lecteur dans ce cas qui précise quels sont parmi ces comportements passés, goûts et intentions signalées ceux qu'il y a lieu de prendre en considération. Chacun contribue donc, s'il le souhaite à ce que la librairie connaisse mieux son profil de lecteur. Un terme récent recouvre bien cette nouvelle notion : nous sommes devenus (si on le veut bien) des « consommacteurs »<sup>6</sup>...

### **Petite typologies des recommandations**

Avant de conclure cet article par l'impact que ce nouveau phénomène pourrait avoir en bibliothèques, il faut – que les technophobes me pardonnent ! – que je touche un mot des techniques qui sous-tendent ces algorithmes de recommandation. En fait, on relève à l'heure actuelle deux grandes approches pour la constitution des listes de recommandation : les méthodes basées sur le contenu (« content-based », en

anglais) et celles basées sur le filtrage collaboratif (ou « collaborative filtering based »).

Pour donner une idée du principe, sans entrer ici dans les détails, les premières méthodes considèrent comme proches d'un titre, les titres qui ont en commun avec lui un nombre important de mots au sein de certaines zones de leurs notices bibliographiques. Par exemple, « Le langage Java pour les nuls » sera considéré comme proche de « Java, premier langage » parce qu'ils ont en commun au sein des zones de titre, de l'indexation RAMEAU, du résumé et des étiquettes attribuées par les internautes, les mots « Java », « langage », « programmation », programmer », « développement », algorithmes » et « Web ».

Le principe des secondes est fondé sur le croisement, pour un titre donné, de listes de livres contenant ce titre. Ces listes sont constituées sur base des historiques des livres achetés / empruntés par des lecteurs, sur les bibliothèques virtuelles (ou listes) constituées par les membres sur un réseau social du livre, etc. Interviennent notamment dans le calcul du degré de proximité de deux titres, le nombre de fois que les deux livres se retrouvent dans une même liste mais aussi, dans certains cas, les notes attribuées à ces deux titres par les lecteurs qui les ont indiqués dans leur liste. Ainsi, avec un tel système, il est possible qu'un livre comme « Le grand Meaulnes » figure en tête de liste de recommandations par rapport à un titre de départ qui serait « Le diable au corps ». Par contre les livres de la collection Harlequin n'y figureront vraisemblablement pas, bien que l'on puisse les cataloguer comme romans d'amour également et qu'à ce titre un simple rebond sur le genre littéraire affiché dans la notice aurait pu entraîner leur affichage, ainsi d'ailleurs qu'un système basé sur le contenu...

Il est au demeurant important de signaler que des méthodes mixtes semblent actuellement donner de meilleurs résultats encore...

### **Vers des catalogues 2.0 avec recommandations en bibliothèques ?**

Bon, et dans nos bibliothèques, que fait-on de cette nouvelle tendance ? Plusieurs voies se présentent : soit on laisse cela aux librairies en ligne et aux réseaux sociaux du livre, soit on se l'approprie au départ des ressources propres du catalogue et des historiques de prêts (en respectant bien évidemment l'anonymat des données), soit on l'achète à un fournisseur (sous la forme d'un Web service, comme à Toulouse, par exemple).

Personnellement je suis très favorable à ce que les bibliothèques publiques proposent ce nouveau service en ligne à leurs usagers car il rejoint entièrement deux rôles importants du bibliothécaire qui consistent d'une part à aider l'utilisateur dans ses recherches et d'autre part, de manière plus proactive, à celui de faire découvrir d'autres œuvres aux lecteurs. Pour ce second rôle, une question pertinente fut énoncée récemment à un débat sur la nouvelle « démocritique » littéraire (les critiques de livres déposées sur le Web par les internautes), à Charleroi lors d'un atelier des « rencontres wallonnes de l'Internet citoyen » : y a-t-il un risque d'enfermement au sein de domaines de connaissance ou de genres littéraires spécifiques ? Un des postulats de départ des recommandations personnalisées est en effet que ce dont le lecteur a envie est du même tonneau que ce dont il a eu envie... Mais après tout, ce ne sont là que des suggestions, quand un garçon de café vous dit, dans l'établissement dont vous êtes un habitué, « un petit noir bien serré, comme d'habitude », rien ne vous empêche de dire « je vais plutôt prendre un thé vert, aujourd'hui » !

Quant à l'alternative entre le développement d'un algorithme qui part des données collectées par la bibliothèque et l'abonnement à un Web service (ou une autre forme de fourniture de ce type de service), chaque bibliothèque intéressée par des recommandations de lecture devra faire son choix. Parmi les questions fondamentales à se poser dans cette réflexion relevons en quelques-unes :

- Disposons-nous en interne des types de données requis et de la masse critique afin qu'un algorithme interne tourne de manière efficace ?
- Le fournisseur de notre OPAC propose(ra)-t-il un tel service ?
- Pouvons-nous (en interne ou en externalisant) développer et assurer une maintenance de ce type d'algorithme et l'intégrer à notre OPAC ? Si oui, à quel coût de développement, avec quel tarif annuel de maintenance ? avec quelle autonomie par rapport au fournisseur et/ou au Service informatique de notre institution ?
- Quels sont les coûts d'utilisation de tels services fournis par un réseau social ? Un tel choix d'outsourcing n'implique-t-il pas une perte de la maîtrise de l'évolution et de la souplesse du paramétrage de ce service ?

Mais avant de se poser toutes ces questions, une réflexion me paraît essentielle : à quel niveau (local, départemental, régional, ...) convient-il d'organiser ce service ? N'est-ce pas en mutualisant leurs efforts et leurs données que les bibliothèques pourront assurer leur position sur le nouveau Web ?

[1] A tout le moins c'est sur Amazon que j'ai pu observer les premiers mouvements en ce sens mais il est possible que d'autres aient été plus précoces ; qu'ils veuillent bien m'en excuser le cas échéant. Je serais heureux par ailleurs que tout lecteur de cet article qui aurait connaissance de tels précurseurs ait l'amabilité de m'en faire part...

[2] Un autre grand réseau social de lecture français, Libfly, prépare lui aussi une offre, aux dernières nouvelles. D'autres réseaux avancent peut-être également dans cette direction mais cet article recherche davantage à donner un aperçu des nouvelles tendances significatives qu'à être exhaustif.

[3] A nouveau, ce catalogue est pris à titre d'exemple illustratif du propos, sans aucune autre considération.

[4] Technique qui permet à un site, lorsqu'on affiche la notice détaillée d'un livre (ou lorsqu'on clique sur un bouton ou un hyperlien sur l'écran), d'envoyer automatiquement l'ISBN de ce livre par Internet à une base de données distante qui renvoie alors, en temps réel, des informations qu'elle contient sur ce livre et que la base de données de votre catalogue ne contient pas (par exemple, l'image de couverture). Parmi les avantages de cette technique on citera le fait que vous ne devez pas stocker ces infos en local et surtout que la « fraîcheur de l'info » peut être meilleure si la base de données distante est mise à jour en continu.

[5] Les moins de trente ans voudront bien m'excuser pour cette référence un peu « antédiluvienne » !

[6] On notera pour être précis que ce terme est polysémique et revendiqué par plusieurs mouvances, notamment la commerciale et celle de la défense de l'éthique dans le commerce. Voir à ce sujet l'article en ligne de François Laurent, « Marketing is dead ».