

La médiation du livre à l'ère du numérique : l'enjeu des algorithmes de recommandation

Mediating the Book in a Numeric Era: The Challenge of Recommender Algorithms

La mediación del libro en la era digital: el desafío de los algoritmos de recomendación

Stéphane Labbé

Volume 61, numéro 1, janvier–mars 2015

URI : <https://id.erudit.org/iderudit/1029000ar>

DOI : <https://doi.org/10.7202/1029000ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Labbé, S. (2015). La médiation du livre à l'ère du numérique : l'enjeu des algorithmes de recommandation. *Documentation et bibliothèques*, 61(1), 15–21. <https://doi.org/10.7202/1029000ar>

Résumé de l'article

Par la réalisation d'une réflexion théorique, nous avons analysé la médiation en ligne du livre à l'aide des algorithmes de recommandation. Nous avons adopté une approche communicationnelle, celle-ci portant un regard sur le processus même de la médiation. Nous avons tenté de circonscrire le concept d'algorithme de recommandation et décrit son fonctionnement. Cela nous a permis d'identifier les enjeux centraux que soulève le recours aux algorithmes de recommandation dans le processus de médiation en ligne du livre, c'est-à-dire l'enjeu de la diversité culturelle et celui du cloisonnement des individus dans des identités et des communautés numériques.

La médiation du livre à l'ère du numérique : l'enjeu des algorithmes de recommandation

STÉPHANE LABBÉ

Doctorant en communication sociale, Département de lettres et communication sociale
Université du Québec à Trois-Rivières
stephane.labbe@uqtr.ca

RÉSUMÉ | ABSTRACT | RESUMEN

Par la réalisation d'une réflexion théorique, nous avons analysé la médiation en ligne du livre à l'aide des algorithmes de recommandation. Nous avons adopté une approche communicationnelle, celle-ci portant un regard sur le processus même de la médiation. Nous avons tenté de circonscrire le concept d'algorithme de recommandation et décrit son fonctionnement. Cela nous a permis d'identifier les enjeux centraux que soulève le recours aux algorithmes de recommandation dans le processus de médiation en ligne du livre, c'est-à-dire l'enjeu de la diversité culturelle et celui du cloisonnement des individus dans des identités et des communautés numériques.

Mediating the Book in a Numeric Era: The Challenge of Recommender Algorithms

Using a theoretical thought process, we analysed online mediation of the book using recommender algorithms. We developed a communicational approach that examines the mediation process. We attempted to circumscribe the recommender algorithm and to describe its function. This allowed us to identify the main challenges when using the recommender algorithm during online mediation of the book, namely cultural diversity and the grouping of individuals according to identities and numeric communities.

La mediación del libro en la era digital: el desafío de los algoritmos de recomendación

A través de una reflexión teórica, hemos analizado la mediación en línea del libro, utilizando algoritmos de recomendación. Para ello, hemos adoptado un enfoque comunicacional, que ofrece una mirada sobre el proceso de mediación. Asimismo, hemos intentado circunscribir el concepto de algoritmo de recomendación y describir su funcionamiento. Esto nos ha permitido identificar los principales desafíos que representa el uso de algoritmos de recomendación en el proceso de mediación en línea del libro, es decir, el desafío de la diversidad cultural y el de la agrupación de las personas en identidades y comunidades digitales.

Au sein de l'industrie culturelle du livre, la médiation joue un rôle central dans la mise en contact d'une œuvre avec son public. Ce processus émane en partie de la nature expérientielle du livre, celle-ci faisant référence au fait que le livre, en tant qu'œuvre unique, est un bien qui nécessite qu'on amorce sa consommation, et donc qu'on se le procure, avant d'en apprécier ses qualités (Colbert 2012).

La médiation culturelle est un objet interdisciplinaire. La sociologie, la psychologie, la bibliothéconomie et la communication sont autant d'approches permettant son étude. Notre angle d'approche est communicationnel et, de ce point de vue, la médiation est conçue comme un processus qui met en relation des contenus avec des individus dans l'objectif d'une éventuelle appropriation de ces contenus par les individus. La prémisse de ce processus est celle d'une intention de communiquer, et celle-ci sous-tend la sélection d'éléments qui seront effectivement transmis, ainsi que le recours à une stratégie de communication. Dit autrement, la médiation culturelle évoque « *un ensemble de pratiques plus ou moins reconnues, entre certaines offres culturelles et une partie du public à qui elles sont destinées* » (Chaumier & Mairesse 2013, 5-6). Ces précisions autour du concept de médiation impliquent que l'émetteur sait à qui il s'adresse, ou qu'il en a à tout le moins une représentation, une idée, que celle-ci soit juste ou erronée. À cet égard, l'approche communicationnelle est cohérente en ce qu'elle considère comme central « *l'usage que font les publics d'objets culturels* » (Lucke-hoff & Jacobi 2014, 59-60).

Au Québec, c'est sous l'impulsion des politiques de démocratisation du gouvernement québécois que la *Loi sur le développement des entreprises québécoises dans le domaine du livre* a été adoptée au début des années 1980 afin d'assurer, d'une part, un accès aux livres sur tout le territoire et, d'autre part, le développement d'une production littéraire nationale (Québec (Province) 2014). Ainsi, la médiation du livre a d'abord été ancrée dans des lieux physiques sur un territoire donné, notamment dans les bibliothèques, et elle a été envisagée surtout dans la mise en contact des livres avec les éventuels lecteurs, par une stratégie de développement de l'offre et de l'accès à celle-ci.

Dans ces lieux que sont les bibliothèques publiques, les stratégies de médiation se sont traduites sous différentes formes : de la classification au catalogage, de la signalétique au bureau d'information ou de service

(Bertrand 1995). Qui plus est, le rôle de médiateur est vu comme central dans les fonctions du bibliothécaire, en ce qu'il « traite toutes sortes de documents et a pour vocation de mettre en relation des usagers et un contenu, quel que soit son vecteur » (Leroux et al. 2009).

En bibliothèque, comme en librairie d'ailleurs, la médiation du livre constitue une forme de discussion qui se traduit souvent en une recommandation de lecture : que celle-ci soit verbalisée par le bibliothécaire, mise en valeur par une signalétique spécifique ou, plus généralement, qu'elle se traduise par une sélection de livres mise en disponibilité dans les collections. La recommandation de lecture s'insère dans le processus de médiation du livre comme une stratégie communicationnelle qui, nous l'avons évoqué plus haut, prend en compte les besoins de l'utilisateur-lecteur.

À l'ère du numérique, le processus de médiation du livre, incluant les recommandations de lecture, est systématisé. Cette systématisation n'évacue en rien toute intervention humaine : le processus que sous-tend la technique nécessite qu'on le programme en amont, en fonction d'un objectif et d'une vision, lesquels demeurent le fruit d'une réflexion humaine. Aussi, ce type de procédé n'est pas exclusif à la médiation en ligne. Le recours à la signalétique par les bibliothèques, comme par les musées par ailleurs, constitue un dispositif de médiation (Davallon 2000), et celui-ci est systématisé et a été conçu en vue de l'atteinte d'un objectif.

Le dispositif de médiation actuellement privilégié pour la médiation en ligne du livre demeure les algorithmes de recommandation. La nature et le fonctionnement des algorithmes de recommandation, notamment en matière de culture, sont des objets de recherche récents et la littérature scientifique en la matière est émergente. Nous proposons ici une réflexion théorique sur les enjeux que soulève le recours actuel aux algorithmes de recommandation dans le processus de médiation du livre. Cette réflexion nous amènera, dans un premier temps, à définir le concept d'algorithme de recommandation et, dans un deuxième temps, à faire la description de son fonctionnement. Cela nous permettra d'entamer ensuite notre analyse des enjeux de la médiation en ligne du livre.

Approche méthodologique

Cet article constitue une réflexion théorique de type analyse de pertinence, c'est-à-dire un type de recherche « très utile pour permettre aux chercheurs d'identifier des questions de recherche d'intérêt pour la communauté des chercheurs » (Raïche & Noël-Gaudreault 2008, 486). Ainsi, nous visons l'identification de problématiques et d'enjeux relatifs à la médiation culturelle à l'ère du numérique, et plus spécifiquement à la médiation en ligne du livre. L'analyse est réalisée sous l'angle de l'approche communicationnelle des faits culturels, et « c'est [donc] sous l'angle de la communica-

tion, de la socialisation et de la médiation que les institutions, les productions, les pratiques [et] les publics sont étudiés » (Luckerhoff & Jacobi 2014, 59-60).

Nous avons procédé à une recherche documentaire portant sur les algorithmes de recommandation dans le milieu culturel. Notre objectif consistait à identifier la nature des algorithmes, et donc à les définir, ainsi qu'à mieux comprendre leur fonctionnement. Ainsi, la sélection des textes tout comme la grille de lecture ont été dirigées par cette visée.

Les bases de données Scopus, SocINDEX, Web of Science, Érudit, Google Scholar, et plus généralement le moteur de recherche Google, ont été utilisés.

Les mots-clés « médiation », « numérique », « culture » et « algorithme », ainsi que leurs traductions anglaises, ont été systématiquement saisis dans chacune des bases de données.

Pour une définition de l'algorithme

La terminologie utilisée par les chercheurs en matière d'algorithme de recommandation semble relativement commune en ce que notre revue de la littérature sur le sujet converge vers les concepts d'algorithme *per se* et de système de recommandation (*recommender system*). Ce vocabulaire commun est d'ailleurs probablement à l'origine de la convergence des définitions en la matière.

Parmi ces définitions, nous retenons celle de Tarleton Gillespie, suggérant que les algorithmes sont des « procédures encodées qui utilisent des formules mathématiques dans le but de transformer des données entrantes (input data) en des résultats escomptés (desired output) » (Gillespie 2012, 1, notre traduction). Cette riche définition nécessite d'être déconstruite afin d'en mieux comprendre les tenants et aboutissants. Quatre concepts sont ici amenés par l'auteur : les procédures encodées, les formules mathématiques, les données entrantes et enfin les résultats escomptés.

Les procédures encodées font appel non seulement au fait que la nature des algorithmes demeure procédurale, et donc dynamique, mais surtout au fait que ces procédures sont programmées (encodées) par des individus ou des organisations, et qu'elles sont donc le reflet de leurs interprétations, de leurs visions ou encore de leurs objectifs. À cet égard, on pointe vers le fait que les « systèmes de recommandation produisent des suggestions orientées vers l'utilisateur et sont reconnus utiles tant par ces derniers, particulièrement par la simplification de la recherche, que par les organisations les produisant, notamment par le fait qu'ils contribuent à augmenter leur chiffre d'affaires » (Campochiaro, Casatta, Cremonesi & Turrin 2009, 648, notre traduction). Cet objectif de satisfaction de l'utilisateur se traduit par le fait d'augmenter la visibilité de certains produits parmi les très larges catalogues, notamment en proposant un échantillon réduit d'articles dont la probabilité qu'ils satisfassent aux besoins de l'utilisateur soit élevée. Cette

procédure est connue sous le nom de « *top-N recommendation task* » (Campochiaro et al. 2009, 648).

Le deuxième concept, celui des formules mathématiques, vient préciser que l'algorithme est caractérisé par une forme de systématisme, sans égard aux éventuels impacts de cette systématisation. Cela implique, par exemple, que les données qui peuvent être identifiées par un algorithme sont standardisées, c'est-à-dire formatées et prêtes à être utilisées par l'algorithme, et ce, sans intervention humaine.

Le troisième concept, les données entrantes, appelle trois prémisses : d'abord que l'algorithme ne peut opérer sans l'existence de ces données, ensuite que les résultats escomptés sont déterminés par les données entrantes, et enfin que ces données entrantes peuvent être de plusieurs formes. En effet, il peut s'agir de données qu'un usager aura introduites lui-même dans la base de données ou l'outil de recherche, ou de données que la plateforme, à l'aide d'un algorithme, aura collectées auprès de ce même usager. Nous n'avons pas recensé de typologie catégorisant les différents types de données entrantes. Le contexte semble jouer un rôle déterminant dans les types de données qui seront développés et agrégés par un algorithme. Par exemple, dans une étude portant sur les algorithmes de recommandation dans le domaine musical, les chercheurs ont identifié trois types de données (métadonnées dans le texte), à savoir les métadonnées éditoriales (*editorial metadata*), culturelles (*cultural metadata*) et acoustiques (*acoustic metadata*). Les premières réfèrent aux données générales telles que produites par le producteur, notamment celles décrivant l'album et le genre musical, et celles-ci peuvent varier selon la vision et la culture du producteur. Les deuxièmes constituent des données implicites qui réfèrent à la culture d'appartenance. Et, enfin, les données acoustiques proviennent d'analyses des contenus audio, le rythme et la mélodie en constituant des exemples (Baumann & Hummel 2005).

Enfin, le dernier concept amené par Gillespie dans sa définition de l'algorithme, celui des résultats escomptés, positionne ces derniers comme étant la finalité de l'algorithme et met au jour la volonté de satisfaire les besoins des usagers, laquelle s'arrime à la manière d'une boucle au premier concept présenté. Ainsi, la définition proposée par Gillespie laisse poindre que l'algorithme est un processus itératif, c'est-à-dire un processus qui évolue au fil du temps et, surtout, au fil des données qu'il collecte et de la pertinence des résultats qu'il produit.

Nous terminons cette section descriptive par la présentation des types d'algorithmes recensés dans la littérature. De la même manière que pour la typologie des données utilisées par les algorithmes, ces derniers n'ont pas fait l'objet d'une catégorisation arrêtée et cela émane probablement du caractère récent de l'objet. Cela étant, Campochiaro et al. (2009), inspirés de Adomavicius et Tuzhilin (2005), ont proposé une typologie en deux familles d'algorithmes : les algorithmes de contenus (*content filtering*) et les algorithmes de relations (*col-*

laborative filtering). Nous proposons ces traductions en cohérence avec le fonctionnement de ces algorithmes : les premiers établissent des liens de similitude entre les différents contenus, les seconds établissent des liens entre les contenus et les informations relatives aux usagers. Pour apprécier la transposition de ces deux familles d'algorithmes en des applications concrètes, c'est-à-dire en des algorithmes spécifiques, on consultera les travaux de Doooms, De Pessemier et Martens (2011a, 2011b) portant sur les algorithmes de recommandation dans le domaine des événements culturels.

Enfin, à une échelle globale, lire mondiale, on a également proposé le concept d'algorithmes culturels (*cultural algorithms*), lesquels réfèrent autant à une application dans un espace culturellement ou géographiquement défini qu'à un espace virtuellement circonscrit par le partage de croyances et de valeurs communes (Zhang 2012).

Fonctionnement des algorithmes de recommandation

L'algorithme est une procédure dynamique, et nous tâcherons ici d'en comprendre le fonctionnement, non pas dans ses moindres détails puisque ce n'est pas l'objectif de cet article, mais bien son fonctionnement en tant que système, et donc en analysant comment sont mises en relation chacune des composantes du système. Rappelons que notre approche est communicationnelle, et qu'à cet égard nous nous intéresserons surtout au processus mettant en relation les contenus avec les usagers.

L'axiome du système algorithmique est à l'effet que l'algorithme en soi n'a qu'une faible valeur alors que les données et métadonnées qu'il utilise ont une haute valeur. Dit autrement, « *les algorithmes sont inertes, des machines sans signification tant qu'ils ne sont pas arrimés à des bases de données pour fonctionner* » (Gillespie 2012, 3, notre traduction).

Le processus d'arrimage de contenus à des usagers implique trois entités, à savoir les contenus, l'outil de recherche utilisé et les utilisateurs de cet outil (Pan et al. 2007). Dans le cas d'une recherche à l'aide de Google, ce processus implique les contenus des pages Web référencées par Google, l'outil de recherche de Google *per se* et l'utilisateur produisant la requête de recherche. De la même manière, mais dans le cas d'une recherche de titres sur une plateforme de vente de livres, le processus implique les livres référencés par la plateforme, l'outil de recherche de la plateforme et l'utilisateur de celle-ci.

Le fonctionnement de l'algorithme a pour point de départ la requête d'un usager. Cette requête se traduit généralement par la saisie de mots-clés dans un outil de recherche, mais il peut également s'agir de la sélection ou de la désélection d'une catégorie dans un index de furetage. La plupart des usagers saisissent généralement deux ou trois mots-clés dans l'outil de recherche,

mais ces comportements présentent des variances selon la culture et l'âge de l'individu (Pan et al. 2007). Toutefois, « *ces mots-clés sont souvent insuffisants et imprécis* » (Zhang 2012, 3641-3642) et cela se traduit par l'échec de la mise en relation entre les contenus et l'utilisateur. Pour pallier cet éventuel insuccès, on peut avoir recours à l'expansion sémantique (*query expansion*), soit un procédé arrimant les mots-clés saisis par l'utilisateur à des termes implicitement ou explicitement pertinents à la recherche (Zhang 2012).

Lorsque les données de la requête de l'utilisateur sont saisies dans le système, l'algorithme débute les opérations de forage dans les données. Selon la procédure préétablie par les programmeurs, notamment l'objectif visé, l'algorithme réalisera une recherche des données les plus pertinentes à la requête de l'utilisateur. Ce qui nous importe ici, dans le processus de recherche effectué par l'algorithme, c'est l'établissement du niveau de pertinence des données, et non le procédé technique de forage. En effet, l'existence même de niveaux différenciés de pertinence sous-tend une hiérarchie préalable des données. À cet égard, l'exemple du PageRank de Google est fort répandu dans la littérature. Le niveau de pertinence d'un contenu est attribué selon le niveau de popularité dudit contenu tel qu'observé par le nombre de liens le partageant sur le Web. Ainsi, plus un contenu a été relié à d'autres contenus sur le Web, qui plus est à d'autres contenus eux-mêmes très reliés, plus ce contenu se verra attribué un PageRank élevé, lire un niveau de pertinence élevé (Brooks 2004; Cardon 2013). Le postulat de ce système hiérarchique réside dans le fait que ce sont les utilisateurs du Web au sens large qui, par l'importance qu'ils accordent à des contenus plutôt qu'à d'autres, « *déterminent une valeur commune* » (Brooks 2004, 3) à ces contenus, lire une valeur populaire.

Le modèle de hiérarchisation des contenus du Web par Google à l'aide de PageRank n'est pas le seul modèle actuel. En effet, la plupart des algorithmes sur le Web utilisent des modèles vectoriels recherchant les similitudes dans les contenus. Ces modèles diffèrent de celui de Google par le fait que la pertinence est accordée aux contenus similaires et non à la structure du Web et à l'intensité d'usage de certains liens plus populaires. (Bianchini, Gori & Scarselli 2005). Autrement dit, le système hiérarchique de PageRank pourrait être associé à celui de la citation dans le domaine scientifique (Bianchini et al. 2005), alors que celui basé sur les similitudes de contenus pourrait être assimilé au fonctionnement des catalogues des bibliothèques publiques accessibles sur les lieux (non-reconnaissance de l'utilisateur et de son historique).

La littérature fait également état du développement de procédés d'identification et de hiérarchisation des données qui sont spécifiques à des contextes particuliers. On consultera à cet égard les travaux de Baumann et Hummel (2005) et de Domms et al. (2011) pour l'industrie de la musique, et ceux de Rosa, Rodrigues et Basso

(2013) sur le recours à la géolocalisation comme outil additionnel en la matière.

Les données utilisées par l'algorithme ne concernent pas seulement les mots-clés saisis par l'utilisateur et les contenus s'y rapportant, mais aussi – et peut-être surtout – le profil de l'utilisateur, c'est-à-dire son historique de navigation et de consommation sur le Web. Cette utilisation des profils des usagers est fort répandue, on y voit même le passage d'une catégorisation [des usagers] par des critères démographiques à une catégorisation par des critères psychographiques où des communautés d'usagers sont créées pour mieux comprendre les caractéristiques non-essentialistes des patrons de consommation basés sur la démographie.

(Cheney-Lippold 2011, 167, notre traduction)

Dit autrement, ce sont davantage les usages du Web que font les consommateurs qui sont utilisés pour mieux comprendre leurs habitudes de consommation, et non plus leurs seuls critères sociodémographiques. Cette catégorisation des usagers du Web par leurs usages réfère à ce que l'on a nommé l'identité algorithmique (Cheney-Lippold 2011, 165).

Ainsi, les données provenant des profils des usagers sont centrales dans l'atteinte de la mise en relation des contenus les plus pertinents à la requête d'un usager, c'est-à-dire le procédé de personnalisation des recommandations (Ochoa, González, Esquivel, Matozzi & Maffucci 2009). À cet égard, de nombreux types de données peuvent être utilisés par les algorithmes, notamment les évaluations des usagers à l'aide des systèmes de notation (5 étoiles, « J'aime » ou « Je n'aime pas », etc.), leur position géographique (Rosa et al. 2013) ou, plus généralement, l'historique de navigation et l'historique de consommation. Qui plus est, l'association du profil d'un usager à des usagers présentant des profils similaires constitue également un type de données (Yom-Tov, Dumais & Guo 2013).

Enfin, les données les plus pertinentes identifiées par l'algorithme seront présentées à l'utilisateur, celles-ci constituant les résultats escomptés au sens de la définition de Gillespie (2012). Ces résultats ont également comme fonction de nourrir l'algorithme, puisque l'utilisation éventuelle de ceux-ci par l'utilisateur constitue un indicateur de pertinence permettant à l'algorithme (et à ses programmeurs) de perfectionner la procédure. À cet égard, une étude a démontré le haut niveau de confiance accordé aux résultats de recherche qui sont présentés par Google à la suite d'une requête de recherche : les chercheurs concluent que « *les sujets sont fortement influencés par l'ordre de présentation dans lequel les résultats de la requête sont présentés et, dans une moindre mesure, par la pertinence du résumé de ceux-ci* » (Pan et al. 2007). Ces résultats laissent poindre un processus où les résul-

tats identifiés comme étant les plus pertinents à une requête seront le plus souvent sélectionnés et, par le fait même, leur niveau de pertinence sera d'autant plus augmenté.

En conclusion de cette section, nous rappelons que les données qui peuvent être identifiées par un algorithme sont, par nature, des données standardisées, c'est-à-dire des données qui sont formatées et prêtes à être utilisées par l'algorithme, et ce, sans intervention humaine. Cela implique, par exemple, que les données erronées, ou absentes, d'un titre de livre sur une plateforme en ligne, feront en sorte que le titre en question ne sera pas sélectionné par l'algorithme.

Les enjeux soulevés par les algorithmes de recommandation

Les enjeux que soulève le recours aux algorithmes de recommandation dans le processus de médiation en ligne des produits culturels sont nombreux. Au-delà du fait que « *les algorithmes ont la capacité à modeler le social et le culturel d'une société jusqu'à des impacts directs sur la vie des individus* » (Beer 2009, 994, notre traduction), il nous a semblé que les enjeux en question pouvaient être agrégés sous l'égide de deux problématiques centrales : l'éventuelle perte d'une diversité culturelle et le cloisonnement des individus dans des communautés d'esprit monolithique.

Diversité culturelle

La menace à la diversité culturelle, c'est-à-dire à l'accès et à la possibilité de découvrir une pluralité de titres, par opposition à une production hégémonique, émane de l'objectif même des algorithmes, du processus de généralisation qui les sous-tend et des critères de pertinence attribués aux données dans le processus d'arrimage des contenus aux requêtes des usagers.

En effet, l'objectif déclaré des algorithmes de recommandation, au-delà des éventuels objectifs commerciaux des plateformes qui les utilisent, demeure la satisfaction des requêtes des usagers. Qu'il s'agisse d'un détaillant en ligne faisant la commercialisation de livres, d'une bibliothèque publique ou d'une plateforme collective de recommandation telle que Babelio, en France, la visée première de ces organisations est de satisfaire au mieux les requêtes de leurs usagers.

Au Québec comme ailleurs, la production littéraire a connu une augmentation importante au cours des dernières décennies. En 2011, au Québec, c'était plus de 6 000 nouveaux titres québécois (Laforce & Milot 2013) qui étaient proposés à la population. La production nationale s'ajoute à la production étrangère importée : en 2007, c'est plus de 25 000 titres étrangers qui avaient été commercialisés au Québec (Lasalle & Gélinas 2007). Devant une telle offre, la volonté de proposer un échantillon réduit d'articles dont la probabilité qu'ils satis-

fassent aux besoins de l'utilisateur soit élevée, notamment par le recours à la procédure « *top-N recommandation task* » (Campochiaro et al. 2009, 648), implique la sélection de titres qui ont préalablement fait leurs preuves, c'est-à-dire des œuvres populaires, nommément les ouvrages à succès (*best-sellers*) du moment.

Aussi, le fait d'avoir recours, en réponse à la pratique des usagers de ne saisir que deux ou trois mots-clés dans les outils de recherche des plateformes, au processus d'expansion sémantique (*query expansion*) (Zhang 2012) donne lieu à un processus de généralisation des requêtes de recherche. Or, la généralisation constitue une « *opération par laquelle on étend à une classe d'objets ce que l'on a observé sur un individu ou quelques cas particuliers* » (Grawitz 2001, 16). Ainsi peut-on proposer que cet arrimage entre les mots-clés des usagers et les classes d'objets en question consiste à associer la requête de l'utilisateur aux œuvres les plus populaires se rapprochant des thèmes de la requête en question.

Enfin, la hiérarchisation systématique des niveaux de pertinence des données telle qu'elle est programmée au cœur de l'algorithme suggère, en lien avec l'objectif de satisfaction des usagers, que le niveau de consultation ou de transaction (emprunt ou achat) d'un titre constitue un critère dont la pertinence est très élevée. Par ailleurs, Gillespie a identifié le fait que « *les critères avec lesquels les algorithmes déterminent ce qui est pertinent de ce qui ne l'est pas, le fait que ces critères sont méconnus du public, et comment ces critères peuvent influencer sur les choix politiques* » (2012, 2, notre traduction) comme étant un enjeu en soi. Qui plus est, ce dernier a également élargi l'enjeu de la pertinence aux « *schémas d'inclusion : des intentions derrière la décision d'inclure et d'exclure un indicateur, et ce qui fait qu'une donnée est réputée formatée et prête à être utilisée par un algorithme* » (Gillespie 2012, 2, notre traduction).

Nous sommes d'avis que, dans un contexte de forte abondance, l'objectif de satisfaction des usagers nécessite l'implantation de processus systématisés tels que l'expansion sémantique et l'établissement de la pertinence par la popularité. Toutefois, cette visée centrée sur l'utilisateur et les techniques utilisées pour y parvenir nous semble constituer une menace à la diversité culturelle, notamment et principalement par la recommandation des œuvres les plus populaires au détriment d'œuvres méconnues.

Cloisonnement des individus

S'appuyant sur les nombreux travaux provenant de diverses disciplines, notamment ceux de la psychologie sociale et de la communication, faisant état de la tendance générale des individus à rechercher des informations qui les confortent dans leurs points de vue, plusieurs auteurs ont suggéré que la structure du Web en général, et les algorithmes en particulier, concourent au cloisonnement des individus.

De la formation de « *like-minded spheres* » et de « *one-dimensional mentalities* » (Dahlgren 2009, 164) au « *filter bubble* » (Yom-Tov et al. 2013, 145), un phénomène émanant de l'exposition d'un individu à une gamme plus étroite de points de vue, en passant par le « *marketing discrimination* » (Beer 2009, 990), un processus de catégorisation fine des consommateurs, le cloisonnement des usagers du Web semble être une résultante importante de la conjugaison du Web, des algorithmes et de notre tendance à rechercher le connu et le reconnu.

Gillespie en a également traité récemment en se questionnant jusque sur la modification même des pratiques des usagers face aux impacts des algorithmes, ainsi que sur la formation de communautés « *calculated publics* » (Gillespie 2012, 2).

Ainsi, le recours de plus en plus intense aux données des usagers dans une visée de personnalisation sous-tendue par l'objectif de satisfaction fine de leurs besoins participe à la formation d'une identité algorithmique (Cheney-Lippold 2011). Cette nouvelle identité, basée sur les usages que les individus font du Web, est associée à des usagers dont l'identité est similaire, formant ainsi des communautés partageant les mêmes valeurs et les mêmes visions.

Ce cloisonnement des individus dans des identités numériques et dans des communautés nichées constitue une entrave à l'émancipation individuelle, à la découverte, et plus généralement au processus de socialisation et de participation culturelle des individus.

Conclusion

Nous avons souhaité ici, à l'aide d'une réflexion théorique, analyser la médiation en ligne du livre à l'aide des algorithmes de recommandation. Nous avons adopté une approche communicationnelle, celle-ci portant un regard sur le processus de médiation *per se*. À l'aide d'une revue de la littérature, nous avons tenté de circonscrire le concept d'algorithme de recommandation ainsi que d'en faire la description du fonctionnement en tant que système de médiation. Cela nous a permis d'identifier les enjeux centraux que soulève le recours aux algorithmes de recommandation dans le processus de médiation en ligne du livre, c'est-à-dire l'enjeu de la diversité culturelle et celui du cloisonnement des individus dans des identités et des communautés numériques.

Nous souhaitons, en guise de conclusion, et en cohérence avec la méthode de l'analyse de pertinence émanant de notre méthodologie, proposer des pistes de recherche pouvant nous aider à mieux comprendre la médiation du livre à l'ère du numérique.

Notre première proposition à cet égard prend sa source dans la dichotomie des modèles de commercialisation des produits culturels. Deux modèles sont opérés par les acteurs, à savoir le modèle culturel, où la culture est promue, protégée et démocratisée, et le modèle de marché, où la culture est marchandisée, libéralisée

et économiquement instrumentalisée (Audley 1994). En ligne, ces modèles sont également présents dans la typologie des algorithmes : les algorithmes de contenus (*content filtering*) et les algorithmes de relations (*collaborative filtering*). Or, il a été démontré que le recours à un seul type d'algorithme, le plus généralement les algorithmes de relations, s'avérerait moins efficace que le recours à des algorithmes hybrides, c'est-à-dire des algorithmes procédant simultanément à l'analyse des similitudes des contenus et des relations entre ces contenus et les profils des usagers (Baumann & Hummel 2005; Ochoa et al. 2009; Dooms et al. 2011a). Des recherches portant sur la forme que pourrait prendre ce nouveau type d'algorithme dans le domaine du livre, notamment sur les plateformes de prêts de livres numériques, nous semblent d'un intérêt certain. Et de façon plus générale, une recherche documentant les modèles hybrides de commercialisation des produits culturels, c'est-à-dire des modèles combinant le modèle de marché et le modèle culturel, ou encore des modèles visant à la fois la satisfaction des usagers et la mise en valeur de la diversité de la production littéraire, nous apparaît pertinente. Pour appuyer cette proposition, citons Dan Franklin, éditeur de contenu numérique (*digital publisher*) chez Penguin Random House :

We are becoming more consumer-centric. But we have to be careful, you can't afford to lose that really interesting, creative, risk-taking element. There's a sense in which we don't know what we want, and publishers [et qui plus est, les bibliothécaires] need to try and help readers learn what they want. Data and creativity need to be blended together.

(Jones 2013)

Notre deuxième proposition de recherche concerne l'enjeu de la diversité culturelle. Des recherches faisant état des résultats de recherche obtenus sur différentes plateformes de médiation du livre en ligne en réponse à l'utilisation de groupes de mots-clés pourraient nous aider à prendre la mesure de la tendance des algorithmes à suggérer les titres les plus populaires du moment. De la même manière, des études comparatives pourraient être menées selon les types de mots-clés les plus utilisés dans les outils de recherche et les résultats qu'ils produisent (auteur, thème, mots présents dans le titre, etc.)

Une recherche procédant à des analyses des résultats de recherche obtenus sur les plateformes de médiation en ligne de livres, selon des profils individuels différents, constitue notre troisième proposition de recherche. Un tel projet pourrait contribuer à mieux comprendre la formation des identités algorithmiques et des communautés de consommateurs.

Aussi, un projet visant à produire une typologie des différents types de données et métadonnées utilisées par les algorithmes de recommandation dans le domaine culturel serait pertinent. Un tel projet pourrait notam-

ment faire la description des formats minimaux nécessaires à leur traitement par les algorithmes. De la même manière, un projet proposant une typologie des types d'algorithmes de recommandation pourrait permettre une meilleure compréhension des fonctionnalités, des natures et des visées possibles.

Enfin, une recherche plus générale sur les rôles des différents acteurs de l'écosystème du livre, notamment les bibliothécaires, en matière de médiation du livre à l'ère du numérique, pourrait concourir à l'identification d'enjeux spécifiques aux différentes professions, au moment où celles-ci sont en évolution. Depuis les années quatre-vingt-dix au Québec, la médiation telle qu'elle a été envisagée par les politiques de démocratisation de la culture, notamment par sa visée générale d'intéresser le public, s'est traduite le plus souvent par des stratégies centrées sur la satisfaction des besoins des usagers. L'ère du numérique va-t-elle faire évoluer cette visée? ■

Sources consultées

- Adomavicius, Gedemias & Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions* 17 (6) : 734-749.
- Audley, Paul. 1994. Cultural industries policy: Objectives, formulation, and evaluation. *Canadian Journal of Communications* 19 (3). <jc-online.ca/index.php/journal/article/view/823/729> (consulté le 8 mai 2014).
- Baumann, Stephan & Oliver Hummel. 2005. Enhancing music recommendation algorithms using cultural metadata. *Journal of New Music Research* 34 (2) : 161-172.
- Beer, David. 2009. Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11 (6) : 985-1002.
- Bertrand, Anne-Marie. 1995. *Les bibliothécaires face au public*. Paris : BPI.
- Bianchini, Monica, Marco Gori & Franco Scarselli. 2005. Inside PageRank. *ACM Transactions on Internet Technology (TOIT)* 5 (1) : 92-128.
- Brooks, Terrence A. 2004. The nature of meaning in the age of Google. *Information Research* 9 (3) : 1-18.
- Campochiaro, Elica, Riccardo Casatta, Paolo Cremonesi & Roberto Turrin. 2009. Do metrics make recommender algorithms? In *Advanced Information Networking and Applications Workshops*. Braford, UK : IEEE, 648-653.
- Cardon, Dominique. 2013. Dans l'esprit du PageRank : une enquête sur l'algorithme de Google. *Réseaux* 1 (177) : 63-95.
- Chaumier, Serge & François Mairesse. 2013. *La médiation culturelle*. Paris : Armand Colin.
- Cheney-Lippold, John. 2011. A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society* 28 (6) : 164-181.
- Colbert, François. 2012. *Le marketing des arts et de la culture* (2^e éd.). Montréal : Gaëtan Morin éditeur.
- Dahlgren, Peter. 2009. *Media and Political Engagement. Citizens, Communication, and Democracy*. New York : Cambridge University Press.
- Davallon, Jean. 2000. *L'exposition à l'œuvre*. Paris : L'Harmattan.
- Dooms, Simon, Toon De Pessemier & Luc Martens. 2011a. A user-centric evaluation of recommender algorithms for an event recommendation system. In *Proceedings of the RecSys 2011*. Ghent : Ghent University, 67-73.
- Dooms, Simon, Toon De Pessemier & Luc Martens. 2011b. An online evaluation of explicit feedback mechanisms for recommender systems. In *Proceedings of the 7th International Conference on Web Information Systems and Technologies (WEBIST-2011)*. Ghent : Ghent University, 391-394.
- Gillespie, Tarleton. 2014. The relevance of algorithms. In *Media Technologies*, sous la direction de Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot. Cambridge : MIT Press, 167-194.
- Grawitz, Madeline. 2001. *Méthodes des sciences sociales*. Paris : Dalloz.
- Jones, Philip. 2013. 'Ditch old ways', Frankfurt hears. *The Bookseller*. <<http://www.thebookseller.com/news/ditch-old-ways-frankfurt-hears.html>> (consulté le 27 mars 2014).
- Laforce, Mireille & Guylaine Milot. 2013. *Statistiques de l'édition au Québec en 2011*. Montréal : Bibliothèques et Archives nationales du Québec.
- Lasalle, Michel & Renée Gélina. 2007. *Étude sur la mise en marché des nouveautés par le système de l'office au Québec*. Montréal : Table de concertation interprofessionnelle du milieu du livre.
- Leroux, Éric, Yvon Lemay, Pierrette Bergeron, Christine Dufour, Dominique Gazo, Yves Marcoux, Sabine Mas & Jean-Michel Salaün. 2009. Les professions et les institutions. In *Introduction aux sciences de l'information*, sous la direction de Jean-Michel Salaün & Clément Arsenaault. Montréal : Presses de l'Université de Montréal, 15-29.
- Luckerhoff, Jason & Daniel Jacobi. 2014. L'étude communicationnelle de la culture : le cas des publics des musées d'art. In *Introduction à la communication sociale*, sous la direction de Stéphane Perreault & Yvon Laplante. Trois-Rivières : SMG.
- Ochoa, Alberto, Saul González, Cynthia Esquivel, Girolama Matozzi & Abigail Maffucci. 2009. Musical recommendation on thematic Web radio. *Journal of Computers* 4 (8) : 742-746.
- Pan, Bing, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay & Laura Granka. 2007. In Google we trust: User's decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication* 12 (3) : 801-823.
- Québec (Province). 2014. *Loi sur le développement des entreprises québécoises dans le domaine du livre*. Québec : Éditeur officiel du Québec.
- Raiche, Gilles & Monique Noël-Gaudreault. 2008. Article de recherche théorique et article de recherche empirique : particularités. *Revue des sciences de l'éducation* 34 (2) : 485-490.
- Rosa, Pedro M. P., Joel J. P. C. Rodrigues & Filippo Basso. 2013. A weight-aware recommendation algorithm for mobile multimedia systems. *Mobile Information Systems* 9 (2) : 139-155.
- Yom-Tov, Elad, Susan Dumais & Qi Guo. 2013. Promoting civil discourse through search engine diversity. *Social Science Computer Review* 32 (2) : 145-154.
- Zhang, Yu. 2012. Research of Web search based on cultural algorithm new framework. *Procedia Engineering* 29 : 3641-3645.