

Qu'est-ce que la découvrabilité?

La découvrabilité est une notion qui s'abreuve de nombreux concepts très médiatisés souvent mal compris. Le LATICCE, **Laboratoire de recherche sur la découvrabilité et les transformations des industries culturelles à l'ère du commerce électronique** propose d'effectuer un tour d'horizon s'accompagnant d'un lexique afin de mieux comprendre cet enjeu des technologies de l'information et des communications et ses impacts sur l'avenir des secteurs culturels.

«La découvrabilité est la capacité d'un contenu culturel de se laisser découvrir aisément par le consommateur qui le recherche et de se faire proposer au consommateur qui n'en connaissait pas l'existence.»

OBSERVATOIRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC (2017). État des lieux sur les métadonnées relatives aux contenus culturels, Québec, Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, 118 p., [En ligne]. [www.stat.gouv.qc.ca/observatoire], Québec 17 octobre.

De nombreux énoncés de politiques culturelles et mesures de soutien récemment introduits insistent sur l'importance de favoriser la découvrabilité des objets et contenus culturels numérisés. Nous pensons notamment au Plan culturel numérique du Québec, au Plan d'action pour la musique, à la Stratégie numérique du Québec, au cadre stratégique du Canada Créatif, au Fonds stratégie numérique du Conseil des Arts du Canada, au programme Exploration et déploiement numérique du Conseil des arts et des lettres du Québec ou encore au document de définition d'un Plan de découvrabilité pour les séries numériques de format court du Fonds Bell.

En amont et en aval du sommet de mai 2016 sur la Découvrabilité (1), le blogue FMC veille publiait cinq articles sous la plume de Danielle Desjardins (2) qui demeurent fondateurs pour cerner l'enjeu, notamment pour le secteur de l'audiovisuel.

Le LATICCE estime qu'il est actuellement nécessaire que des mesures de la découvrabilité soient développées et ceci notamment afin d'évaluer l'efficacité des efforts d'accroissement de celle-ci engagés par les pouvoirs publics, les acteurs terrains et les producteurs de contenus.

Notre problématique de recherche peut s'illustrer par la question suivante : Est-ce que les services culturels numérisés locaux sont "découvrables" sur les grandes plateformes transnationales qui dominent actuellement l'offre culturelle mondiale/locale? Les processus d'indexation normalisés des *objets culturels et médiatiques numérisés* en amont de la mise à disposition des contenus et les déterminants techniques de leur découvrabilité en aval sont des aspects qui peuvent expliquer le niveau de découvrabilité, comme également les stratégies d'affaires visant à favoriser une offre par rapport à une autre. Ces éléments pourront permettre de cibler des mesures pour accroître la découvrabilité des contenus locaux, un préalable à l'émergence de modèles d'affaires soutenables dans nos industries culturelles.

Pour ce faire, nous développons actuellement une série d'indicateurs, divers procédés de recherche qualitatifs et quantitatifs pour produire, d'ici 2020, un indice de découvrabilité synthétique qui permettra d'établir un palmarès des portails de contenus culturels en regard de leur capacité à **présenter P, rendre visible V et recommander R** des contenus. Nous cherchons à établir quels sont les portails qui favorisent la diversité plutôt que la concentration de l'offre. À l'instar du gouvernement québécois qui établissait en octobre 2017 son *Plan d'action pour la musique*, nous travaillons prioritairement sur le secteur musical pour lequel aucune offre de streaming canadienne ou québécoise n'existe. Nous engageons aussi des efforts pour les secteurs de l'audiovisuel et du livre.

Bien que nous jugeons nos travaux utiles pour une appropriation judicieuse des procédés d'activation de la découvrabilité, nous mettons le lecteur en garde quant au caractère obligatoirement évolutif de notre démarche. Les réseaux numériques connectés sont en rapide mutation, ce sont des cibles mouvantes où les modèles d'affaires et procédés développés par les services sont souvent opaques, où la culture se limite souvent à n'être qu'une lucrative commodité.

Afin de mieux cerner nos travaux, dans le but d'ouvrir la conversation avec les milieux créatifs, les pouvoirs publics et la société civile, nous nous proposons d'abord de présenter sous forme de lexique, une série de concepts avec lesquels le LATICCE entend jongler ou qu'il souhaite aborder par ses recherches. Nous ouvrons les commentaires pour cet article, ainsi que divers espaces de l'actuel Wiki, et invitons les parties concernées par le défi de la découvrabilité à poursuivre la discussion avec nous.

Objet médiatique et culturel numérique (OMCN) | Expression générique utilisée pour nommer l'ensemble des contenus culturels dématérialisés qui circulent en ligne.

Écoute en flux numérique continu (streaming) et **Plateformes numériques d'écoute en ligne** (PNEL) | Il s'agit de la norme actuelle en matière d'écoute des contenus audio et audiovisuels. L'écoute en flux continu s'apparente à un flux radio et requiert une connexion active sur réseau cellulaire ou Wi-Fi. Certains services de *streaming* permettent aussi le téléchargement pour écoute hors connexion.

Déterminants techniques | Dans un univers où la découvrabilité est très souvent influencée par les comportements de navigation de l'internaute et les traces que celui-ci laisse derrière lui, nous considérons que des déterminants techniques, des biais industriels ou informatiques sont à l'oeuvre. Il s'agit de facteurs qui vont moduler le comportement des algorithmes de recommandation de contenus. Voir les sections Adresse MAC, adresse IP et cookies / Modèles d'affaires et d'accès / Profilage de goût et recommandation, etc.

Présence (P) | Première routine de collecte automatisée du LATICCE procédant par requêtes aux API des plateformes et consistant à vérifier la présence sur un service donné, d'un titre tiré d'une liste de référence.

Visibilité (V) | Seconde routine de collecte automatisée du LATICCE procédant par requêtes aux API des plateformes, actuellement en cours d'élaboration. Les indicateurs de visibilité sont conditionnés par une série de déterminants techniques à l'étude. Voir Déterminants techniques.

Profilage de goût et recommandation (R) | Les plateformes offrant des produits culturels cherchent le plus possible à établir une corrélation entre ce que l'internaute cherche et ce qu'elles ont à lui offrir. Ainsi par diverses méthodes, elles profilent les goûts de leurs abonnés ou des internautes qui fréquentent anonymement leurs services. On moucharde les écoutes, les recherches et les listes d'écoute créées par les usagers, on questionne les internautes quant à leurs préférences. Les mécaniques de recommandation se fondent sur le **filtrage collaboratif**, l'**analyse du langage naturel** (métadonnées descriptives, analyse textuelle de blogs et de sites spécialisés)(3), les **annotations Web sémantiques**, la **curation humaine**. À la demande ou à fréquence définie, on recommande aux consommateurs des listes d'écoute de nouveautés, de classiques, d'ambiances diverses, modulées selon l'activité ou le moment de la journée. La lecture des recommandations constitue à terme la troisième routine de collecte automatisée du LATICCE.

Liste d'écoute (playlist) | Les plateformes accompagnent l'expérience d'écoute et de consultation et le font très souvent en créant des listes d'écoute, des programmes de contenus. Les *playlists* peuvent être générées par des humains agissant à titre d'éditeurs, par des algorithmes, par les usagers et abonnés eux-mêmes et par diverses combinaisons de ces interventions. Les services peuvent vous proposer une seule liste à fréquence régulière, comme elles peuvent dans certains cas en produire de très nombreuses sans respecter une périodicité spécifique.

Adresse MAC, adresse IP et cookies | Nous cherchons à identifier les déterminants techniques qui peuvent influencer le profilage de goût des internautes abonnés aux services d'écoute en flux numérique. Dans quelle mesure l'appareil (reconnu par l'adresse MAC ou Media Access Control de sa carte SIM ou réseau), son système d'opération ou la localisation de l'abonné (souvent déterminée par l'adresse IP Internet-Protocol associée au serveur de connexion de l'internaute) ou encore les cookies enregistrés dans l'historique de navigation de vos navigateurs Chrome, Firefox, Safari ou autre, sont pris en compte lors de la constitution des recommandations. Cette problématique de recherche est en cours et n'est pas encore résolue.

Métadonnées | Les métadonnées sont des informations servant à décrire, enrichir et trouver les objets culturels et médiatiques numérisés mis à disposition sur Internet. Elles existent sous de nombreux types à la fois sectoriels ou trans-sectoriels, inter-opérables ou pas, normés ou pas. Il faut distinguer les métadonnées qui permettent l'indexation des contenus *en amont* de leur transmission dans la chaîne de valeur, des données d'usage moissonnées *en aval* sur la base des comportements des internautes à l'égard des contenus : achat, écoute complète ou partielle, géolocalisation du consommateur etc.

Algorithmes et intelligence artificielle | L'algorithme est une suite d'instructions permettant d'obtenir un résultat. Ils sont utilisés par les logiciels de recommandation. En intelligence artificielle, l'apprentissage automatique (machine learning) est quant à lui une routine qui permet à un programme informatique ou un algorithme de s'améliorer lui-même. On peut présumer qu'une certaine part d'apprentissage automatique est à l'oeuvre dans les processus de raffinement des outils de recommandation ou de profilage de goût.

Modèles d'affaires et d'accès | Les offres en ligne et leurs conditions d'accès se déclinent selon de nombreuses variantes. Ces variantes jouent sur la convivialité du service, sa souplesse à l'égard des besoins de l'abonné et à terme sur la découvrabilité des contenus qui s'y trouvent répertoriés. Que le modèle soit fondé sur la gratuité ou sur une mensualité plus ou moins élevée pourra avoir une incidence sur la qualité de la prestation aux abonnés, sur les outils de découvrabilité ou de recommandation à sa disposition.

Listes de référence / Nettoyage des données / Codage manuel | On ne peut pas attendre d'un algorithme qu'il effectue un travail convaincant et pertinent si les données que nous soumettons à son traitement sont incomplètes, fausses, contiennent des fautes ou ne sont pas transmises selon certains protocoles, normes et standards. Nous devons par conséquent circonscrire nos listes de référence, les nettoyer, les coder avant de les soumettre à nos procédés de forage et d'interrogation des plateformes.

API et requêtes automatisées | L'API est une interface logicielle applicative (programme informatique) qui permet de récupérer des informations par requêtes répétées dans les bases de données qui en offrent la possibilité. C'est le cas pour iTunes, YouTube ou Spotify. Le LATICE a recours aux API des plateformes et développe aussi ses propres requêtes émulant le comportement des abonnés (Google Play, Netflix). La requête une fois transmise générera une réponse - souvent sous forme d'affichage d'une page Web - utile ou non à nos recherches.

Persona | Le LATICE étudie la perspective de créer des persona ou portraits robots en matière de consommation culturelle, profils d'écoute subjectivement normés, qui seront utilisés lors du recours à certains types de requêtes aux API afin de générer des résultats longitudinaux contrôlés.

Conditions légales d'utilisation | Les abonnés aux divers services doivent toujours adhérer aux conditions légales d'utilisation des plateformes pour se prévaloir du produit. Ces contrats d'adhésion rarement lus contiennent des dispositions régissant la latitude que les services se donnent en matière d'exploitation des données personnelles. Elles sont d'un intérêt manifeste lorsque des opérations de profilage ont cours, ce qui est le cas en ce qui a trait à vos goûts en matière de fréquentation culturelle.

Optimisation de la recherche (SEO) et amplification par le truchement des médias sociaux-numériques | Ces approches habituellement associées aux opérations marketing sont fréquemment assimilées aux stratégies d'activation de la découvrabilité. L'utilisation d'annotations sémantiques ou de métadonnées d'indexation optimales, une activité soutenue sur Facebook ou Twitter ont une incidence sur la découvrabilité, mais nous formulons l'hypothèse à l'effet que ces actions doivent être complétées par d'autres efforts tels l'ouverture des données, leur sémantisation ou encore l'emploi de standards descriptifs rigoureux pour faciliter le travail des algorithmes et des moteurs de recommandation.

Annotations sémantiques et sémantisation des bases de données relationnelles | Ce travail est assimilable aux concepts du Web sémantique et du Web des données : réseaux Internet étendus à l'ensemble des données ouvertes exposées en ligne pouvant être lues, interprétées et présentées par les machines. Ajouter des annotations ou sémantiser des contenus implique un travail spécifique répondant à des normes de catalogage. Utiliser ces technologies a une incidence sur la découvrabilité des objets médiatiques et culturels numériques. Voir à ce titre l'article *Découvrabilité et métadonnées : nous sommes nuls en documentation de contenu* (4).

Ontologies / Référentiels / Thésaurus / Triple stores et triplets | Ces termes issus du champ des sciences de l'information forment les assises du Web sémantique et sont aussi utilisés en intelligence artificielle. Ils renvoient aux notions qui permettent aux machines de tirer du sens des données soumises aux algorithmes de calcul informatique.

Jean-Robert Bisailon - codirecteur du LATICE, avril 2018.

(1) Sommet de la découvrabilité / Les contenus à l'ère de l'abondance | <http://decouvrabilite.ca/> | 10 et 11 mai 2016

(2) Danielle Desjardins (2016), Découvrabilité : Vers un cadre de référence commun <https://trends.cmf-fmc.ca/fr/blog/decouvrabilite-vers-un-cadre-de-reference-commun> / Découvrabilité Volet 2 : Le parcours de l'auditoire <https://trends.cmf-fmc.ca/fr/research-reports/decouvrabilite-volet-2-le-parcours-de-lauditoire> / Surfer sur la vague de la quatrième révolution industrielle <https://trends.cmf-fmc.ca/fr/blog/surfer-la-vague-de-la-quatrieme-revolution-industrielle> / Les machines qui prédisent ce que vous voulez voir <https://trends.cmf-fmc.ca/fr/blog/les-machines-qui-predisent-ce-que-vous-voulez-voir> / Économie de l'attention, la conquête des cerveaux <https://trends.cmf-fmc.ca/fr/blog/economie-de-lattention-la-conquete-des-cerveaux>

(3) Brian Withman (2012), Comment fonctionne et ne fonctionne pas la recommandation musicale <https://fr.scribd.com/document/136235243/Brian-Withman-Comment-fonctionne-et-ne-fonctionne-pas-la-recommandation-musicale>

(4) Josée Plamondon (2018), Découvrabilité et métadonnées : nous sommes nuls en documentation de contenu <https://joseeplamondon.com/decouvrabilite-et-metadonnees-nous-sommes-nuls-en-documentation-de-contenu/>

Mon gilet de flottaison répond aux normes et il est repérable...

