

# Lexique

## Lexique

**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 fureteurs 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.

**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.

**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 impliquent un travail spécifique répondant à des normes de catalogage. Utiliser ces technologies a une incidence sur la découvrabilité des objets médiatiques et culturels numériques.

**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 LATICCE 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 génèrera une réponse - souvent sous forme d'affichage d'une page Web - utile ou non à nos recherches.

**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.

**Découvrabilité** | Capacité, pour un contenu culturel, à se laisser découvrir aisément par le consommateur qui le cherche et à se faire proposer au consommateur qui n'en connaissait pas l'existence. (*Observatoire de la culture et des communications du Québec*)

**Découvrabilité** | La découvrabilité, c'est la capacité à mettre en valeur et cibler stratégiquement une proposition artistique dans un espace virtuel et foisonnant. D'abord c'est un terme que ne reconnaissent pas les auto-correcteurs ! La découvrabilité c'est une offre et sa mise à disposition, sa rencontre avec les publics. Les aspects qualitatifs de cette offre jouent un rôle : pertinence, singularité, valeur culturelle intrinsèque, potentiel d'attractivité, caractère temporel. Mais la découvrabilité c'est aussi l'activation des données statistiques, de la sérendipité, de la recommandation, de l'intégration des contenus aux listes de favoris de communautés d'intérêt. (*Jean-Robert Bisailon : "la découvrabilité, un enjeu prioritaire pour les artistes"*)

**Découvrabilité** | La capacité d'une interface utilisateur à laisser découvrir facilement ses caractéristiques sans se référer à un élément extérieur. Ou | Potentiel pour un contenu, un produit ou un service de capter l'attention d'un internaute de manière à lui faire découvrir des contenus autres. (Note : L'emploi, notamment de métadonnées, d'algorithmes de recherche, de mots-clés, d'index, de catalogues augmente la découvrabilité d'un contenu, d'un produit ou d'un service.) (*Office québécois de la langue française*)

**Découvrabilité** | Au départ, la découvrabilité est un terme relevant du domaine juridique en anglais et du domaine informatique, en anglais comme en français. Ce terme désigne la capacité d'un élément – que ce soit une application ou un contenu – à se laisser découvrir facilement. (*Danielle Desjardins, Découvrabilité : vers un cadre de référence commun*)

**Découvrabilité** | Le nouveau monde du contenu mondial en ligne est désormais axé sur la « découvrabilité » — la nouvelle mesure de valeur qui, en plus de refléter le succès critique, reflète celui auprès de l'auditoire. Il s'agit du système de cotes d'écoute ultime, sans intervention d'un interprète comme Nielsen. Le découvrabilité est une méthode qui oriente les spectateurs vers le contenu sans guide télé ni liste de chaînes de télévision. Elle permet en outre de mesurer le succès du contenu dans un monde composé non de 500 mais plutôt de millions de chaînes. (CMPA, Discoverability study)

**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.

**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 WiFi. Certains services de *streaming* permettent aussi le téléchargement pour écoute hors connexion.

**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.

**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.

**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.

**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.

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

**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.

**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.

**Persona** | Le LATICCE 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.

**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.

**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. Le LATICCE a, pour cela, créé trois sous indicateurs associés aux recommandations de contenus soit : la concordance (c) avec le panier de référence du personne étudié (correspondance exacte de certains facteurs avec l'historique d'écoute, tels la liste des artistes interprètes), la pertinence (p) (correspondance avec divers artistes similaires déterminés par filtrage collaboratif) et la nouveauté (n)(présence de propositions de moins de 30 jours).

**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.

\* Voir aussi: [État des lieux sur les métadonnées relatives aux contenus culturels – Glossaire](#) (OCCQ, février 2018)