



HAL
open science

”Méthodes, techniques et outils. Qu’y aura-t-il demain sous nos moteurs ?”

Olivier Ertzscheid

► To cite this version:

Olivier Ertzscheid. ”Méthodes, techniques et outils. Qu’y aura-t-il demain sous nos moteurs?”. Documentaliste - Sciences de l’Information, 2011, 48 (3), pp.10-11. sic_00634397

HAL Id: sic_00634397

https://archivesic.ccsd.cnrs.fr/sic_00634397

Submitted on 21 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ertzscheid Olivier, "Méthodes, techniques et outils. Qu'y aura-t-il demain sous nos moteurs ?" in Documentaliste, sciences de l'information. Vol. 48, n°3, Octobre 2011, pp. 10-11.

Qu'y aura-t-il demain sous nos moteurs ? Moteurs de recherche et réseaux sociaux occupent une place chaque jour plus prépondérante dans nos accès à l'information et à la connaissance. Ils suscitent également de vives interrogations, notamment dans leur capacité à rendre indexable des informations relevant indistinctement des sphères publiques, privées et intimes des contenus disponibles en ligne. Enfin, inaugurés par le « like » de Facebook, les systèmes de recommandation se multiplient, venant concurrencer ou remplacer l'établissement de liens hypertextes et suscitant de nouveaux usages, de nouvelles métriques. Par ailleurs, la famille documentaire s'est agrandie : les encyclopédies sont devenues collaboratives, d'immenses silos documentaires sont apparus (Youtube, Flickr, etc.), les profils humains sont devenus des objets de « collection ».

CE QUI A REELLEMENT CHANGE.

Question d'économies. Dans le contexte d'une abondance de contenus informationnels, prévaut une économie de l'attention hypothéquée par la capacité à mettre en place une économie de la confiance (Trust economy) avec la foule comme support (crowdsourcing), la modélisation de la confiance comme vecteur, et de nouvelles ingénieries relationnelles comme technologie.

La force des métadonnées. Pour les moteurs mais aussi pour réseaux sociaux, toutes les données sont ou peuvent devenir des métadonnées, qui dessinent des comportements (pas uniquement documentaires) en même temps qu'elles permettent de caractériser la motivation de ces mêmes comportements. Chaque clic, chaque requête, chaque comportement connecté fait fonction de métadonnée dans une sorte de panoptique global.

Le web invisible ne l'est plus vraiment. Le défi technologique, après avoir été celui de la masse des documents indexés, devient celui de la restitution en temps quasi-réel du rythme de publication propre au web (« world live web »). Pour accomplir ce miracle, il faut aux outils de captation de notre attention que sont les moteurs et les réseaux sociaux, une architecture qui entretiennent à dessein la confusion entre les sphères de publication publiques, intimes et privées.

La naissance de l'industrie de la recommandation et des moteurs prescripteurs. La recommandation existe de toute éternité numérique, mais elle est désormais entrée dans son ère industrielle. Moteurs et réseaux sociaux fonctionnent comme autant de prescripteurs, soit en valorisant la capacité de prescription affinitaire des « proches », des « amis » ou des « collaborateurs » de l'internaute (boutons « like » et « +1 »), soit en mettant en avant les comportements les plus récurrents de l'ensemble de leurs utilisateurs.

De nouvelles indexations. Après l'indexation des mots-clés, des liens hypertextes, des images, de la vidéo, des profils, il faut maintenant apprendre à indexer, à mettre en mémoire, la manière dont circule l'information, être capable de rendre compte de cette dynamique virale, capable de prendre en compte la dissémination, l'écho, l'effet de buzz que produisent les innombrables « boutons poussoir » présent sur chaque contenu informationnel pour lui assurer une dissémination optimale.

Navigation virale ou promenade carcérale ? L'approche fermée, propriétaire, compartimentée, concurrentielle, épuisable de l'économie du lien hypertexte que proposent les systèmes de recommandation, ne peut mener qu'à des systèmes de nature concentrationnaire. Des écosystèmes de l'enfermement consenti, en parfaite contradiction avec la vision fondatrice de Vannevar Bush selon laquelle la parcourir, le chemin ("trail")

importe au moins autant que le lien. Les ingénieries relationnelles des systèmes de recommandation – de celui d'Amazon au Like de Facebook – ressemblent davantage à une promenade carcérale qu'à une navigation affranchie parce qu'elles amplifient jusqu'à la démesure la mise en avant de certains contenus au détriment de la mise au jour d'une forme de diversité.

Un nouveau brassage. La recherche universelle, désignant la capacité pour l'utilisateur de chercher simultanément dans les différents index (et les différentes bases de données) proposés par les moteurs de recherche, ouvre un immense champ de possibles pour la mise en œuvre d'algorithmes capables de prendre en charge les paramètres excessivement complexes de la personnalisation, de la gestion des historiques de recherche, de l'aspect relationnel ou affinitaire qui relie un nombre de plus en plus grand d'items, ou encore du brassage de ces gigantesques silos de données. Un brassage totalement inédit à cette échelle.

Le mirage des nuages. De rachats successifs en monopoles établis, l'extrême mouvement de concentration qui touche la médiasphère internautique fait débat. D'un immense écosystème ouvert, le web mute aujourd'hui en une succession de ce que Tim Berners Lee nomme des « walled gardens », des jardins fermés, reposant sur des données propriétaires et aliénant leurs usagers en leur interdisant toute forme de partage vers l'extérieur. L'enjeu n'est alors plus simplement celui de l'ouverture des données, mais celui de la mise en place d'un méta-contrôle, un contrôle accru par la migration de l'essentiel de nos données sur les serveurs des sociétés hôtes, grâce à la banalisation du cloud computing : l'essentiel du matériau documentaire qui définit notre rapport à l'information et à la connaissance étant en passe de se retrouver entre les mains de quelques sociétés marchandes

Et tout ça pour quoi ? Il s'agit de porter à visibilité égale des contenus jusqu'ici sous-utilisés ou sous-exploités, pour augmenter leur potentiel marchand en dopant de la sorte le rendement des liens publicitaires afférents. Un objectif qui ne peut être atteint sans la participation massive des internautes.

Le web personnel. La pertinence et la hiérarchisation d'un ensemble de contenus hétérogènes n'a de sens qu'au regard des intérêts exprimés par chacun dans le cadre de ses recherches précédentes. La condition *sine qua non* de la réussite d'une telle opération est le passage au premier plan de la gestion de l'historique des recherches individuelles.

Algorithmie ambiante. A la manière de l'informatique « ambiante » qui a vocation à se diluer dans l'environnement au travers d'interfaces prenant la forme d'objets quotidiens, se dessinent les contours d'une algorithmie également ambiante, c'est à dire mettant sous la coupe de la puissance calculatoire des moteurs, la moindre de nos interactions en ligne.

Marchands de documents. Derrière cette algorithmie ambiante on trouve la volonté déterminée d'optimiser encore davantage la marchandisation de toute unité documentaire recensée, quelle que soit sa sphère d'appartenance d'origine (publique, privée, intime), sa nature médiatique propre (image, son, vidéo, page web, chapitre de livre, etc...), sa granularité (un extrait de livre, un billet de blog, un extrait de vidéo ...) et son taux de partage sur le réseau (usage personnel uniquement, usage partagé entre « proches », usage partagé avec l'ensemble des autres utilisateurs du service).

La recherche prédictive. Les grands acteurs du web disposent aujourd'hui d'une gigantesque « base de donnée des intentions » (concept forgé par John Battelle), construite à l'aide de nos comportements d'achats, de l'historique de nos requêtes, de nos déplacements (géolocalisation), de nos statuts (ce que nous faisons, nos centres d'intérêt), de nos « amis » (qui nous fréquentons). Une base de donnée des intentions qui va augmenter la « prédictibilité » des recherches. Et donc augmenter également leur valeur transactionnelle, leur valeur marchande.

Recherche de proximité et moteurs de voisinage. A l'aide de moteurs comme Intelius.com ou Everyblock.com, il est possible de tout savoir de son voisin numérique, depuis son numéro de sécurité sociale jusqu'à la composition ethnique du quartier dans lequel il vit, en passant par le montant du bien immobilier qu'il possède ou l'historique de ses mariages et de ses divorces. Toutes ces informations sont – aux Etats-Unis en tout cas – disponibles gratuitement et légalement. Ne reste plus qu'à les agréger et à faire payer l'accès à ces recoupements numériques d'un nouveau genre. Surveillance et sous-surveillance s'affirment comme les fondamentaux de cette nouvelle tendance du « neighboring search ».

Pourquoi chercher encore ? Le nouvel horizon de la recherche d'information pose trois questions très étroitement liées. Demain. **Chercherons-nous pour retrouver** ce que nous ou nos « amis » connaissent déjà, permettant ainsi aux acteurs du secteur de vendre encore plus de « temps de cerveau disponible » ? **Chercherons-nous simplement pour acheter**, pour consommer et pour affiner le modèle de la base de donnée des intentions ? **Ou pourrions-nous encore chercher pour découvrir** ce que nous ne savons pas (au risque de l'erreur, de l'inutile, du futile) ?

Le web était un village global. Son seul cadastre était celui des liens hypertexte. Aujourd'hui, les systèmes de recommandation risquent de transformer le village global en quelques immeubles aux incessantes querelles de voisinage.

Un web hypermnésique et des moteurs omniscients. Aujourd'hui le processus d'externalisation de nos mémoires documentaires entamé dans les années 1980 avec l'explosion des mémoires optiques de stockage est totalement servicialisée et industrialisée. L'étape suivante pourrait ressembler à l'hypermnésie dont souffre le Funès de J.L. Borgès. c'est l'hypermnésie. Celle dont souffre Funès dans la nouvelle de Borges. Mais cette hypermnésie est aujourd'hui calculatoire, algorithmique, ambiante. Elle est massivement distribuée, ce qui lui confère cette impression de dilution, de non-dangerosité. Mais quelques acteurs disposent des moyens de l'activer et de tout rassembler. Pour l'instant ce n'est que pour nous vendre de la publicité, du temps de cerveau disponible. Mais que deviendrait cette arme hypermnésique entre les mains d'états ? Nous avons tendance à oublier l'importance de se souvenir puisqu'il est devenu possible de tout se remémorer.

Des enjeux de politique ... documentaire. La deuxième question c'est celle de l'écosystème informationnel que nous souhaitons pour demain. Ne rien dire aujourd'hui, c'est consentir. Il est aujourd'hui absolument nécessaire d'ouvrir un débat autour de l'écosystème non plus simplement documentaire mais politique que les moteurs représentent, il est vital de s'interroger sur la manière dont cet écosystème documentaire conditionne notre biotope politique et social ... Or toutes ces questions sont par essence documentaires, ce sont les questions que posent une macro-documentation du monde. Voilà pourquoi le rôle des professionnels de l'information est et sera absolument déterminant.