

Les métriques du Web

L'absence de clôture tangible dans l'environnement numérique produit un effet de nivellement qui dissimule les opérations de tri et de classement des informations. La hiérarchisation des contenus constitue pourtant un enjeu stratégique majeur, à la fois nerf de la concurrence que se livrent les plateformes et vecteur d'autant de philosophies qu'il y a de prétendants au contrôle du Web.

Louise Merzeau, *Sciences de l'information et de la communication, Paris Ouest Nanterre-La Défense*

Pour l'internaute, savoirs, actualités, conversations semblent se juxtaposer sur la toile, dans un apparent désordre qu'on oppose volontiers à l'ordonnement cartésien de nos bibliothèques. Pour les industries de l'accès en revanche, la visibilité des contenus est tout sauf aléatoire, car c'est leur capacité à filtrer et classer l'information qu'elles monétisent auprès des annonceurs. Peu perceptibles en tant que telles, souvent même couvertes par le secret, ces procédures de tri sont l'élément névralgique de la lutte pour la captation et la capitalisation de l'attention. Véhiculant des conceptions divergentes du Web, elles en signalent aussi l'évolution.

LES QUATRE MODÈLES DE CLASSEMENT

Des années 1990 à aujourd'hui, l'accès aux contenus en ligne s'est opéré selon différentes logiques que l'on peut ramener à 4 modèles⁽¹⁾.

Éditorialisation

Le premier coïncide avec l'époque où l'on pouvait encore répertorier le web pour en extraire manuellement ce qui était considéré comme digne d'intérêt. C'est le modèle des annuaires (Yahoo!) et des encyclopédies (Wikipédia), où la hiérarchisation des informations procède d'une *pratique éditoriale* sur la base de règles négociées par une communauté.

Autorité

Le deuxième s'est imposé à mesure que l'étendue du maillage à indexer prenait des proportions dépassant les capacités humaines de tri. On a alors confié à des algorithmes le soin d'analyser le graphe du web, pour *calculer l'autorité* des sites en fonction des hyperliens qui pointent vers eux. Inspirée de la scientométrie, cette méthode valorise encore l'intelligence collective des éditeurs, tout en automatisant et en rentabilisant son évaluation (PageRank de Google).

Audience

La massification du Web a ensuite favorisé l'importation de métriques issues des industries de programmes. Dans un système où le nombre de visiteurs l'emportait désormais largement sur celui des

publiants, la *mesure d'audience*, classant les contenus en fonction du nombre de pages ou de vidéos vues (YouTube), trouvait toute sa pertinence.

Affinité

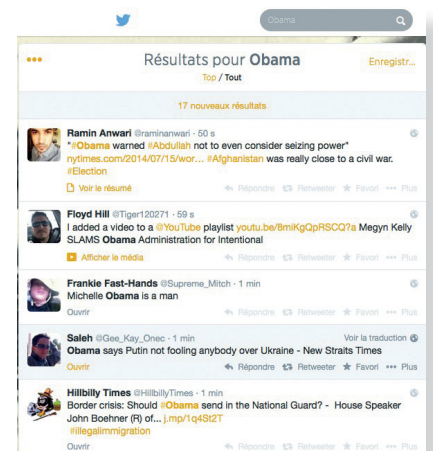
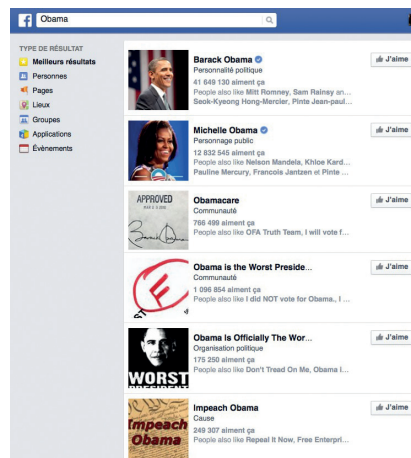
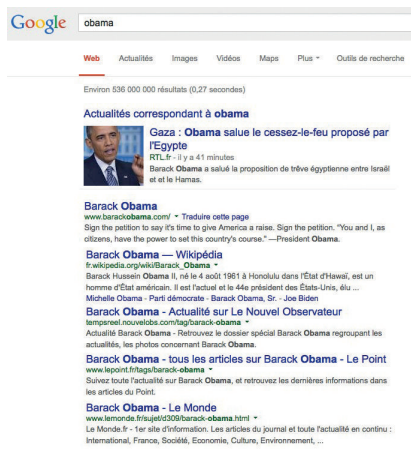
L'arrivée des réseaux sociaux a cependant de nouveau changé la donne : en ouvrant aux non-éditeurs un espace de publication, elle a généré un volume de données conversationnelles qui permettaient de réorganiser les accès, en personnalisant les filtrages et les hiérarchisations. La logique de classement est alors devenue *affinitaire*, en diminuant l'importance des contenus au profit des relations (Facebook) et de la synchronisation (Twitter).

CONCURRENCE ET HYBRIDATION

Aujourd'hui, ces modèles coexistent dans l'espace numérique, chaque plateforme calculant sa propre pertinence avec une logique qui lui est propre. Selon qu'on interroge Wikipédia, Google, YouTube, Facebook ou Twitter, la même requête ne retournera pas le même classement, donnant du même coup une image du web qui masque les autres représentations possibles (*voir doc. 1*).

Si l'on peut encore les distinguer, ces modèles tendent cependant à s'hybrider, comme en témoigne l'évolution du PageRank de Google. À la mesure de l'autorité des sites se sont en effet ajoutés des critères affinitaires, pondérant de plus en plus l'évaluation des hyperliens par celle des graphes sociaux : le contenu le plus pertinent n'est pas tant celui qui est le plus cité que celui qui est le plus compatible avec mon profil. Cette évolution atteste la personnalisation croissante de l'information, signant la victoire progressive du « *like* » sur le lien et la transformation d'un espace documentaire en un web identitaire.

Mais, parallèlement, le moteur de recherche a également éditorialisé sa page de résultats, en y insérant l'infobox construite à partir des données prélevées dans Wikipédia (*voir doc. 2*). Ce faisant, il ne se contente plus d'indiquer des références, il répond directement à la question qu'on lui soumet. De même, les brèves qu'il reprend des sites d'actualité dispensent de plus en plus les internautes d'aller et consulter l'information qu'ils cherchent.



Doc. 1 : Les différentes réponses de Google, Facebook et Twitter à la requête « Obama » en juillet 2014.

Le principe de la mesure d'audience a quant à lui contaminé le modèle du web social, à travers la multiplication d'outils permettant de surveiller ses propres scores. Menée dans un registre semi-public, la conversation sert alors, en même temps, à maintenir sa popularité.

NEUTRALITÉ DES ALGORITHMES OU ÉDITORIALISATION ?

Dans les premiers temps du Web, les classements se revendiquaient comme tels, à l'instar des palmarès publiés dans les médias traditionnels. Aujourd'hui, les portails comme Wikio ont disparu, laissant l'ordonnement du monde numérique aux acteurs qui en monopolisent les points d'entrée. Le risque est alors de laisser ces acteurs naturaliser la mise en ordre qu'ils appliquent au web, sous couvert d'une prétendue neutralité des algorithmes. Ce serait oublier que, derrière toute métrique, il y a un projet d'éditorialisation, qui configure les distances informationnelles et façonne politiquement le web. Qu'elle soit celle, élitiste, des citations, celle, massive, de la popularité ou celle, relationnelle, des affinités, la pertinence n'est jamais neutre : c'est un choix technique et idéologique.



Doc. 2 : Infobox Wikipédia sur « Obama » en juillet 2014.

(1) Cette typologie est empruntée à Dominique Cardon, *Réseaux* n° 177, 2013/1, « Politique des algorithmes. Les métriques du Web ».