

# Introduction

Par Michel SCHMITT

Conseil général de l'Economie (CGE)

Il n'est pas si loin, le temps des disquettes informatiques souples qui avaient une capacité de 128 kilo-octets... Ceci fait sourire les anciens, car aujourd'hui, Internet et le *cloud* ont balayé ces dispositifs, en nous donnant la sensation que nous disposons d'un espace infini pour y déposer nos contenus. De surcroît, une infinité de contenus sont à notre disposition : des sites nous proposent des bibliothèques de films, de musiques ou de livres en nombre vertigineux.

Cette profusion de ressources sur la toile a modifié nos comportements. Recherche-t-on une information, une définition ? Wikipédia sur son *smartphone* que nous avons à portée de (ou dans) la main. La réponse est instantanée, alors que se déplacer jusqu'à sa bibliothèque, chercher son dictionnaire préféré et tourner les pages prend bien plus de temps. Et puis, l'édition de ce dictionnaire date de quelques années, rendant son information parfois un peu datée, et, le comble, le texte n'est pas indexé. Cette anecdote montre que la quantité<sup>(1)</sup>, la mise en réseau et l'immédiateté ont changé notre rapport aux connaissances et ont forgé de nouveaux usages.

Par ailleurs, nous participons tous à accumuler les connaissances sur la toile, au travers des réseaux sociaux, des encyclopédies participatives, par exemple. L'une des qualités paradoxales de cet infini est sa taille humaine puisque quelques clics permettent d'accéder à tout document. Comme toujours, les choses ne sont pas si simples, et l'objet de ce numéro des *Annales des Mines* consacré aux « Usages et archivages des contenus du Web » est d'examiner comment notre relation aux contenus se modifie progressivement, comment cet infini s'organise, se structure et crée une nouvelle économie, faisant émerger de nouveaux concepts et de nouveaux métiers. Se pose également la question de garder une trace de ces contenus.

## Les usages, ou notre relation aux contenus

La révolution numérique a profondément modifié notre accès à la connaissance et de nombreux usages en ont été transformés ou ont émergé. En cette période de confinement, tous les parents pensent à assurer la continuité de l'éducation de leurs enfants. Les MOOC permettent l'accès aux enseignements de manière simple, sans se déplacer, et la classe, géographiquement éclatée, se reforme grâce à un réseau social dont le professeur devient modérateur. Le livre numérique, qui a cependant quelques difficultés à s'imposer, rend obsolètes les mètres linéaires d'étagères ou la lourde caisse pour partir en vacances. De même, les jeux en réseau ont transformé le quotidien de nos jeunes... et des moins jeunes également.

Côté producteur de contenus, la course à la visibilité est engagée. De nouveaux métiers – dans le sens d'une source de revenus – sont apparus, tels les *youtubeurs*, les *influenceurs*. Des sociétés de service proposent de faire mieux référencer les sites Web. Les *cookies* permettent d'adapter les propositions de contenus aux utilisateurs. Ainsi, deux personnes faisant la même recherche n'obtiendront plus le même résultat, résultat qui sera influencé par le comportement passé sur la toile. Ce qui frappe,

---

(1) Il ne faudrait pas en conclure que l'on trouve tout sur le Web ! J'avais en projet de faire en 2018 une leçon inaugurale de probabilités à mes étudiants aux Mines de Paris. On trouve, certes, des introductions au sujet, mais pas une leçon digne de ce nom sur les sites Web des grandes universités. J'ai donc dû prendre mon bâton de pèlerin pour la construire... et elle se retrouvera sur la toile prochainement.

c'est que le monde se rétrécit grâce au Web, rend proches les unes des autres les personnes, et ce rapprochement est tel que, dans cet univers virtuel, nous ne voyons plus notre voisin !

## **Structurer, ou comment s'y retrouver**

Côté utilisateur, comment s'orienter dans cette jungle de contenus ? C'est comme si l'on devait faire le point dans l'océan des données. Autrefois, les hauturiers, afin de naviguer en haute mer (par opposition aux caboteurs qui restaient toujours en visibilité des côtes), ont dû inventer de nouveaux outils, et c'est ainsi que sont apparus les sextants, les éphémérides et les horloges précises qui ont permis de faire le point indépendamment de repères terrestres. Les outils pour se diriger sur le Web sont les moteurs de recherche qui nous facilitent bien la tâche. Les premiers moteurs cherchaient des occurrences de mots dans les pages Web. Ils emploient maintenant des formules et des algorithmes plus complexes, certains issus de l'intelligence artificielle, pour parcourir, indexer, sélectionner et trier les résultats qu'ils présentent. Nous assistons en parallèle à des évolutions de la toile dans de multiples directions : le Web social (Web 2.0) qui nous connecte toujours plus, le Web sémantique (3.0) qui rend le Web plus accessible aux machines et applications, le Web des objets qui inclut jusque dans nos applications les objets connectés aux réseaux, etc. Et dans cette expansion galopante, nous examinerons plus particulièrement le Web sémantique, qui a pour rôle de garder l'ensemble des ressources que nous versons au Web universellement compréhensibles et utilisables. Il est cependant à noter que ces évolutions ne disent rien sur la qualité des contenus eux-mêmes<sup>(2)</sup>.

## **Aspects économiques et émergence des GAFAM**

Derrière tout cela, se jouent des enjeux économiques colossaux. Les plus grandes capitalisations boursières sont le fait des géants de l'Internet, les GAFAM... Nous regarderons dans ce numéro les stratégies des diffuseurs de contenus, en tête desquels Netflix, YouTube et la manière dont le droit d'auteur évolue, puisque dupliquer un contenu est techniquement le fait d'un clic... Son coût marginal est quasi nul. Sans oublier que plus de la moitié du flux de données vers l'utilisateur est constituée de vidéo...

## **Archiver le Web ou le Web comme archive ?**

Qui n'a pas été agacé par la disparition d'une page Web ? Si l'on regarde le Web comme une immense bibliothèque de contenus, se pose la question de son instabilité et, par suite, de son archivage. Plusieurs projets ont vu le jour, mais ils concernent avant tout le patrimoine – les livres, les objets, les monuments, etc. – et l'on est plus dans une démarche d'archiviste, dont le métier se transforme. Nous sommes toujours dans l'angoisse d'une perte irrémédiable, comme autrefois face à l'incendie d'une bibliothèque ou d'un monument. Là encore se posent de nouvelles questions : que signifie archiver un livre ou un écrit ? Comment le rendre accessible ? Que devient le dépôt légal lorsque l'écrit n'est plus que numérique ?

De plus, de nouveaux types de contenus ont vu le jour. Par exemple, que vont devenir les contributions dans les réseaux sociaux ? Chaque minute, 50 000 nouvelles photos sont postées sur Instagram, 500 000 nouveaux tweets sont diffusés. *Quid* des logiciels qui ont été développés ? Voici quelques angles d'attaque que nous examinerons dans ce numéro.

---

(2) L'immédiateté de l'accès est l'une des caractéristiques de ces contenus numériques. L'édition scientifique est un bel exemple du bouleversement des usages. Ecrire un article scientifique, le soumettre à une revue, le faire *reviewer* (relire), le corriger et enfin l'insérer dans la revue... : ce processus peut prendre plusieurs années. Il est censé garantir la qualité de la recherche, et c'est sur lui que repose la bibliométrie. Cependant, ceci est en contradiction avec l'immédiateté d'Internet, et puisque chacun peut déposer ce qu'il veut – neutralité du Net oblige –, les archives ouvertes ont fleuri. La contrepartie est que c'est l'utilisateur qui doit juger de la qualité.