

1993

Le Web

une utopie à défendre

Conçu initialement au Cern pour faciliter le travail des chercheurs, le Web a largement dépassé ce cadre. Aujourd'hui omniprésent dans notre quotidien, il continue d'évoluer. Une évolution que les citoyens doivent contrôler, faute de quoi les principes fondateurs de ce réseau seront dévoyés.

En 1993, le laboratoire européen de physique des particules, le Cern, renonce à ses droits d'auteur sur ses logiciels du World Wide Web. Le grand public découvre alors la «toile d'araignée mondiale» – un réseau de textes, tableaux, images, sons, vidéos, etc. connectés les uns aux autres. Ce réseau a été conçu au Cern quelques années plus tôt, vers 1990, par deux informaticiens, le Britannique Tim Berners-Lee et le Belge Robert Cailliau, pour faciliter l'échange d'informations entre les chercheurs du monde entier. Le Web s'appuie sur un réseau d'ordinateurs déjà ancien à l'époque, puisqu'il datait de 1969: le réseau Internet.

Initialement, le Web est une simple collection de textes en langues naturelles auxquels on accède par le réseau Internet. Ces textes sont exprimés dans un format nommé HTML, qui permet de les mettre en page et de les structurer, par exemple en définissant des titres et des sous-titres. Mais, et c'est le point essentiel, le format HTML permet aussi de définir des ancres sur lesquelles le lecteur peut cliquer et qui conduisent à d'autres pages HTML. Il n'est donc plus nécessaire de connaître la localisation d'un texte pour y accéder: il suffit de cliquer sur l'ancre correspondante dans un autre texte. Le lecteur peut alors partir d'une adresse qu'il connaît, par exemple www.bnf.fr, et butiner de page en page au gré de son humeur.

L'idée est simple, mais la prouesse est inouïe: en 2017, quatre milliards de personnes sont connectées au réseau Internet et peuvent ainsi accéder à tous les contenus du Web. Il faut

imaginer les millions de lignes de programme des serveurs web qui publient des données, des serveurs internet qui routent ces données à travers le réseau Internet et des navigateurs qui permettent d'y accéder. Tout cela fonctionne en dépit de l'hétérogénéité des matériels et des logiciels utilisés par l'ensemble des utilisateurs.

Le volume des données numériques doublerait tous les 18 mois. Le Web est la partie émergée de cette *terra data*, celle visible depuis n'importe quel ordinateur, téléphone, tablette... connecté au réseau. Le Web met à portée de clics pratiquement toutes les connaissances, toute la culture, toute la vie sociale de la planète.

Dans les années qui ont suivi sa naissance, le Web a mené rapidement à d'autres inventions. Tout d'abord, il s'est enrichi d'autres formats permettant de publier non seulement des textes, mais aussi des images, des sons, des vidéos ou des données structurées.

NAVIGUER SUR LA TOILE

Sont apparus ensuite les moteurs de recherche qui aident à découvrir de l'information dans cet océan de données. L'utilisateur choisit quelques mots-clés et le moteur de recherche trouve en quelques instants les pages qui contiennent ces mots. En ce sens, ce n'est

LES AUTEURS



SERGE ABITEBOUL
chercheur à l'Inria,
membre de l'Académie
des sciences



GILLES DOWEK
chercheur à l'Inria



rien de plus qu'un index, mais un index sur une collection gigantesque de pages qui évoluent en permanence. Une prouesse algorithmique quand on considère les échelles: aujourd'hui, le moteur de recherche Google indexe 30000 milliards de pages et, chaque jour, 20 milliards de sites sont indexés et 3,3 milliards de requêtes sont effectuées.

Comment trouver, parmi les éventuelles centaines de millions de pages qui correspondent à une requête, les quelques pages qui contiennent l'information pertinente pour le lecteur? C'est la seconde grande fonction du moteur de recherche, le classement, qui fait émerger les pages «intéressantes». Dans le cas du moteur de Google Search, le principe utilisé historiquement pour le classement est celui de la «popularité sur le graphe du Web», c'est-à-dire la probabilité pour un flâneur qui se baladerait au hasard sur la Toile de se trouver sur une page particulière. L'algorithme PageRank permet de calculer cette popularité. Aujourd'hui, de nombreuses autres informations sont utilisées pour le classement, notamment les clics des internautes, qui votent ainsi avec leurs doigts. Le moteur de recherche sait ainsi faire surgir, de la masse du Web et des choix des internautes, une intelligence collective pour évaluer, classer, filtrer les informations.

Dans les années 2000, le développement de nouvelles techniques, désignées sous le terme de Web 2.0, a permis à tout un chacun non seulement d'accéder à des informations publiées par des autorités – administrations, universités, entreprises... – mais aussi de publier lui-même des informations. Ce mouvement a conduit quelques années plus tard à des réseaux sociaux entre humains.

Au fil des ans, il est devenu évident qu'au-delà d'une collection universelle de documents, le Web offrait une gamme infinie d'applications à inventer. Nous avons vu arriver le Web de la 3D avec ses mondes virtuels, le Web des objets. Peut-être l'évolution la plus fondamentale sera-t-elle le passage d'un Web des humains à un Web où humains et machines coexistent, en s'appuyant sur les services web, les données semi-structurées, et les connaissances. À chaque fois, l'existence d'un langage standardisé est un composant essentiel du progrès.

Si le Web a été conçu initialement comme outil pour la recherche, il est devenu quasiment impossible aujourd'hui de vivre sans: de trouver du travail, de travailler, de se loger, de gérer ses comptes bancaires, presque d'avoir des amis, etc. On peut cependant regretter le Web des débuts, libre, gratuit, universel, ouvert et anarchique. Le Web évolue vers des espaces plus fermés, notamment sous la pression de la monétisation des contenus ou des biais des systèmes de recommandation, vers des espaces plus hostiles avec ses rumeurs et fausses nouvelles, ses sites racistes ou criminels... Aujourd'hui, plus que jamais, le Web est une utopie à défendre: c'est à nous de décider ce que sera le Web demain. ■

BIBLIOGRAPHIE

S. Abiteboul et G. Dowek,
Le Temps des algorithmes,
Le Pommier, 2017.

J. Gillies et R. Cailliau,
**How the Web was born :
The story of the World
Wide Web**, Oxford
University Press, 2000.