Introduction

Michel Schmitt,
Conseil Général de l'Économie (CGE)

Once upon a time, there were floppy disks with a capacity of 128 kilobytes…. This puts a smile on the face of the older among us, since the Internet and cloud have blown away such objects while imbuing us with the sensation that the space for storing contents is sky high… after all, the contents at our disposal are infinite. Websites offer libraries of films, music and books, all in an astonishing quantity.[1]

This profusion of resources on the Web has changed our behavior patterns. Looking for information or a definition? Wikipedia on the smartphone, which is within reach if not already in our hands. The answer is instantaneous, whereas going to a library or looking for your favorite dictionary and thumbing through it takes much more time. Besides, since the dictionary is an edition dating back a few years, its information might be outdated and, to clap the climax, the information found has not been indexed. As this anecdote shows, the quantity and immediacy of the information placed on the Net have changed our relation to knowledge and spawned new uses.[2]

Coincidentally, we are all contributing to this accumulation of knowledge via user input on the social networks and to online encyclopedias. A paradoxical aspect of this infinity is its human scale, since a few clicks suffice to access any document. As usual of course, things are not so simple. This issue of *Annales des Mines — Enjeux Numériques* devoted to *Web contents: Uses and archives* examines how our relation to online contents is gradually changing, how this infinity is organized and structured, how it is creating a new economy with new jobs, how it has forged new concepts… Another question raised herein is about how to keep track of all these contents.

Using the Web: Our relation to its contents

The digital revolution has deeply altered our access to knowledge, several uses having emerged or been transformed. During the period of "shelter-in-place", as parents have been trying to see to the continuity of their children's education, massive open online courses (MOOC) have provided easy access to lessons without going anywhere; but this geographically dispersed class is organized via a social network site with the professor as moderator. Despite the hard time they have had winning acceptance, e-books have made linear meters of shelf space obsolete, along with the heavy suitcase to be dragged on vacation. Games on the networks have transformed the everyday lives of our young people… and, too, of people more than young!

Among content-producers, the race is on for visibility. New jobs (and sources of income) have arisen: "YouTubers" and "influencers". Companies are offering services for better referencing websites. Cookies are being used to adapt the contents to users. As a consequence, two persons making the same search will not obtain the same results; the findings will be influenced by the individual's past actions on the Web. The Web is shrinking the world; people are coming closer to each other. This convergence on a virtual realm is so powerful that we no longer see our neighbors.

---

[1] This article has been translated from French by Noal Mellott (Omaha Beach, France).

[2] Do not, however, draw the conclusion that everything can be found on the Web. In 2018, I was planning an inaugural lecture on probability. Introductory courses on this subject can be found on the websites of major universities — but not information worthy of being called a lesson. I had to take up the assignment to make one, which will soon be uploaded.

Structuring the Web: How to find our way around?

For users, the problem is to find their way around in this jungle of contents, this ocean of data. Sailors on the deep sea (unlike coasters that always remained within sight of the shore) had to invent new instruments (sextants, ephemeris, and accurate clocks) for taking their bearings independently of landmarks. The instruments for making it easier to find one's way on the Web are search engines. Whereas the first search engines looked for the occurrences of words on web pages, current engines use very complicated formulas and algorithms, some of them derived from artificial intelligence, to browse, index, select and sort the results they present.

In parallel, the Web has steered a course by tacking about: the "social Web" (Web 2.0) for increasing our "connections" with others, the "semantic Web" (3.0) for opening the Web to machines and applications, the Web of "things" for connections via applications on board connected devices, etc. In this rapidly expanding universe, this issue takes a closer look at the semantic Web, which has the role of making all the resources we contribute to the Web universally comprehensible and useable. Note, however, that these trends tell us nothing about the quality of the contents.[3]

Economic considerations and the upsurge of monopolies

The economic stakes behind all this are enormous. The biggest market capitalizations are due to the giants of the Internet: the Big Five — GAFAM: Google, Apple, Facebook, Amazon and Microsoft. Herein, we take a look at the strategies adopted by the distributors of online contents, chief among them Netflix and YouTube, and also at changes in copyright law. After all, duplicating contents is a click away, and the marginal cost of duplication is next to naught. Let us also not forget that more than half the streams of data toward users are… videos.

Archiving the Web, or the Web as an archive?

Who has not been annoyed because a web page has vanished? Seeing the Web as an immense library of contents, we have questions about the stability of this library and, then, about its archives. Several programs have been set up, but mainly for our cultural heritage (books, objects, monuments, etc.). This is not the approach of archivists. Besides, their job is changing. We are haunted by an anxiety about irrecoverable losses, as used to happen when a library was burned down or a monument destroyed. Once again, questions crop up. What does it mean to archive a book or piece of writing? How to make it accessible? What about the obligation of a legal deposit for documents that are uniquely electronic?

Meanwhile, new types of contents have thrived. For instance, what will become of the posts on social networks? Every minute, 50,000 new photos are posted on Instagram, 500,000 new tweets are circulated. And what about software? These are a few angles from which this special issue will tackle these many questions.

---

[3] The immediacy of access is one of the characteristics of Web contents. Scientific publications illustrate this disruption in uses. Writing a scientific article, submitting it to a journal for review, correcting it and, finally, having it published… all this can take years. This long process is intended for warranting the quality of the research presented; and bibliometrics is based on it. Nonetheless, it is incompatible with the immediacy of the Web. Since anyone can upload what he/she wants (Net neutrality), open archives have thrived. As a counterpart however, users are responsible for judging the quality of the contents they retrieve.