Pourquoi s'intéresser aujourd'hui au Web sémantique ? Après tout, depuis ses origines, le Web de Sir Tim Berners-Lee (dont les premiers balbutiements remontent à 1989) a vu se succéder, au fil des conflits commerciaux qu'il a engendrés, des paradigmes dont aucun ne semble laisser la part belle au Web dit "sémantique". Mentionnons, entre autres et en passant : la guerre ouverte que se sont livrée les navigateurs Netscape et Internet Explorer (précédée par la volonté des créateurs de Mosaic de s'emparer littéralement du Web), la bataille des moteurs de recherche dont la firme de Mountain View est sortie vainqueur, du moins en Europe et plus particulièrement en France, la titanomachie actuelle opposant Facebook à Google pour le contrôle des réseaux sociaux, etc. Sans oublier les prodromes de lutte, déjà visibles, autour du *cloud*, des applications mobiles et bientôt des WebTV.

Une telle succession d'acteurs et de modèles pourrait laisser penser que le Web s'inscrit dans une temporalité essentiellement dominée par les révolutions. Pourtant, cette vision tend à masquer l'essentiel. Ce qui, sous cette écologie fluctuante d'acteurs, d'usages et de dispositifs, semble demeurer relativement stable : son *architecture*. Quiconque entend s'intéresser au Web sémantique devra donc porter son regard ailleurs, en quête de stabilité, afin de comprendre ce qui s'apparente bien davantage à une (r)évolution graduelle, préparée de longue date.

Certains récits présentant cette évolution ont acquis une valeur quasi proverbiale. Ainsi, à un Web 1.0, "de documents", aurait succédé un Web 3.0, "sémantique". Mais alors, où situer dans ces conditions le Web social ? S'agit-il d'une invention

1. **http://www.google.com/support/webmasters/bin/answer.py?answer=99170**.
2. **https://developers.facebook.com/docs/opengraph**.

récente, celle que l'on a baptisée, avec le succès que l'on sait, "Web 2.0", synonyme d'ouverture à la contribution des internautes ? Ce serait aller vite en besogne et oublier du même coup que le Web initialement conçu par Tim Berners-Lee comprenait *déjà* la possibilité d'éditer les pages consultées. On parle à cet égard de *Read-Write Web*. Autrement dit, un Web de lecture *et* d'écriture, où nul ne reste cantonné à la place du spectateur. Si ces fonctions ont finalement été abandonnées dans la version grand public du Web, n'oublions pas que les forums, listes de discussions et autres tchats ont très vite ouvert la voie aux échanges. Par conséquent, l'aspect social du Web n'est sans doute pas à mettre au même niveau que ses dimensions documentaires ou sémantiques. De même que le Web 1.0 n'était pas asocial, voire antisocial, en attente de lieux d'échange que seul le développement des technologies AJAX, de mises à jour asynchrones, typiques du Web dit 2.0, lui aurait fournis, de même, il est fort à parier (ce qui n'interdit pas de s'en soucier ni d'œuvrer en ce sens !) que le Web 3.0 ne trahira pas la capacité du Web à constituer un lieu d'échange sans équivalent.

Au plan architectural, ces distinctions n'ont à vrai dire guère de pertinence. Par architecture du Web, il faut ici entendre les principaux standards qui le définissent depuis les origines : les URI, le protocole HTTP et des langages de type HTML ou, aujourd'hui, RDF. Les premières constituent les identifiants du Web (généralement connues pour être ses adresses, les URL, même si le terme n'est plus en usage dans les RFC[1] depuis une bonne quinzaine d'années et ne le fut en fin de compte qu'épisodiquement[2]). Le deuxième n'est autre que le protocole qui gère les échanges client-serveur. Quant aux deux derniers, il s'agit de langages, "documentaire" pour HTML et de description pour RDF.

Il manque toutefois à ce tableau très bref un terme important que le W3C n'a jamais standardisé — et pour cause. Il s'agit des "ressources". On peut s'accorder sur l'unicité du Web et la permanence de son architecture tout en reconnaissant par ailleurs qu'une coupure s'est opérée entre un *Web prémoderne*, encore en gestation, et un *Web moderne*, porté par une conceptualisation qui lui fit rompre les amarres avec les architectures d'échanges de fichiers, voire avec les hypertextes traditionnels (reliant des documents les uns aux autres). Ainsi, entre 1991 et 1994, le Web pouvait-il encore passer pour un Web de documents ou de fichiers auxquels des adresses étaient censées donner accès[3]. À partir de 1994, une intense activité de standardisation (des URI/URL/URN mais aussi du protocole HTTP dans ses versions 1.0 et 1.1) permit de déplacer le curseur de façon décisive. Cette standardisation était en effet portée

1. *Requests for Comments*, documents de l'IETF comprenant de nombreux standards d'Internet et du Web.
2. Entre la RFC 1738 de décembre 1994 et la RFC 2396 d'août 1998.
3. On ne s'étendra pas ici sur l'aspect foncièrement métaphorique que revêt pourtant d'emblée une telle manière de concevoir les choses.

par une vision conceptuelle nouvelle, un "style d'architecture", le style REST[1], dont on peut affirmer sans crainte que la principale caractéristique est d'avoir inspiré la mise en place du Web moderne que nous connaissons (ou, plus justement, habitons). REST ajoutait au Web un élément qui, pourtant, devait toujours, en un sens, lui faire défaut : la ressource. En d'autres termes, tout ce qui est doté d'une identité sur le Web, à savoir tout ce qui est identifié par une URI. Nous n'entrerons pas ici dans les — nombreuses — questions philosophiques posées par cette notion, si ce n'est pour en mesurer l'importance. La ressource a pour caractéristique de n'être jamais accessible sur le Web, contrairement à un fichier par exemple. Elle est *abstraite* (par opposition à concrète). Pour le dire autrement, elle n'est qu'une "ombre"[2], manipulable uniquement à travers ses représentations[3].

Le Web repose donc sur des unités plus abstraites que les documents, gagées sur la notion de "continuité réticulaire" (Justin Erenkratz). La page d'accueil du *Monde*, par exemple, ne saurait correspondre à un fichier HTML ponctuel (il y a fort à parier qu'elle soit en outre beaucoup plus complexe !). Elle conserve son identité en tant que ressource, bien que ses représentations ne cessent — seconde après seconde ! — de se modifier. On peut en ce sens la comparer à une règle identique à elle-même, quand bien même son application livrerait des résultats variables. Cette continuité réticulaire n'est rien d'autre que l'équivalent dynamique de ce que le document, en tant qu'inscription, est censé attester. Ce sont des opérateurs de confiance dans les deux cas[4]. D'une nature cependant fort différente, la régularité, notion dynamique et générative, se substituant à la trace "hétérothétique"[5], dont la fonction est d'affirmer autre chose qu'elle-même (afin, historiquement, de remplir le rôle de preuve).

---

1. Pour *REpresentational State Transfer*, dont les principes sont exposés dans la thèse de Roy Fielding (*Architectural Styles and the Design of Network-based Software Architectures*) et dans un article publié deux ans plus tard en collaboration avec Richard N. Taylor (*Principled Design of the Modern Web Architecture*). Tous deux sont disponibles en ligne.
2. L'expression se trouve telle quelle dans l'article qu'écrivit en 2002 Roy Fielding avec Richard N. Taylor (*cf. Principled Design of the Modern Web Architecture*, § 7.1.2).
3. Au sens que ce terme a pu acquérir *via* REST et le protocole HTTP : les données envoyées/générées à la suite d'une requête formulée par un client.
4. Rappelons ici que la couche la plus haute du cake du Web sémantique n'est autre que la confiance. Loin d'être cantonnée à son sommet, elle imprègne en réalité toutes les strates du Web.
5. Par opposition à la constitution autothétique du numérique selon Bruno Bachimont : "Une inscription numérique est, comme nous l'avons dit, une inscription dans un langage formel qui, par définition, est manipulable. Le support de l'inscription numérique est le calculable. Cela signifie qu'une inscription numérique est quelque chose qui, potentiellement, a toujours été manipulée. Une inscription numérique ne porte pas la trace des calculs qui l'ont produite. Une telle inscription n'a pas de mémoire." (*Cf. Du texte à l'hypotexte : les parcours de la mémoire documentaire*, **http://www.utc.fr/~bachimon/Publications_attachments/Hypotexte.pdf**).

En définitive, la ressource, "concept"[1] ou "objet" — au sens que ce terme a pu acquérir en philosophie –, peut être "n'importe quoi"[2]. Abstraite, il ne s'agira jamais de l'atteindre autrement qu'au travers du prisme concret, donc manipulable, de ses représentations. Les possibilités qu'ouvre cette notion sont immenses. Dès lors que son centre de gravité se déplace des documents et autres fichiers concrets vers ces ombres conceptuelles socialement constituées, le Web dispose d'un moyen de prendre pour cible aussi bien des objets dits "en ligne" que des objets qui ne transitent pas en tant que tels sur le réseau (personnes et objets concrets, concepts abstraits, etc. ; en somme, tout ce qui peut être visé par un nom propre/une URI, cela allant jusqu'à inclure des fictions ou des éléments purement déclaratifs). Autrement dit, *le Web peut nous permettre de parler de n'importe quoi*.

En fin de compte, le Web sémantique ajoute simplement une brique à cet édifice sous la forme d'un langage de représentation des connaissances adapté à de pareilles ressources, fussent-elles "informationnelles" ou "non informationnelles". Ce langage n'est autre que RDF. Allons jusqu'à dire qu'il accomplit le programme à l'origine du Web en venant compenser les limites du langage HTML (de structuration de l'information documentaire).

Pourquoi s'intéresser aujourd'hui au Web sémantique ? Pour la simple et bonne raison qu'il accomplit le Web moderne selon l'esprit, si ce n'est selon la lettre. De ce point de vue, l'événement éditorial que constitue la traduction en français du livre de Tom Heath et Christian Bizer aidera à n'en pas douter le lecteur à s'orienter dans les méandres du Web sémantique mais également — et surtout ! — à prendre connaissance des derniers développements d'une histoire longue d'au moins deux décennies.

Alexandre Monnin, responsable recherche web et métadonnées à l'IRI

Votre avis

L'avis de nos lecteurs étant toujours le bienvenu, n'hésitez pas à nous faire part de vos commentaires. Pour cela, rendez-vous sur la page dédiée à cet ouvrage sur le site web Pearson (**http://pearson.fr**) dans la librairie en ligne et cliquez sur le lien Réagir.

1. C'est bien ainsi que Fielding la caractérise.
2. "A resource can be anything that has identity", in *IETF, RFC 2396 (1998) : Uniform Resource Identifiers (URI): Generic Syntax*.