

Le Web 3.0 ou le Web sémantique, quel futur ?

Sur le Web, il ne suffit plus de gérer des mots (clés ou non) ; il faut désormais gérer des vocabulaires. Le Web 3.0 peut constituer la solution s'il permet de connecter la façon dont une personne a exprimé une idée avec le mode d'expression d'une autre personne. Pour franchir cette étape, une nouvelle révolution technologique est nécessaire.

Par Patrick Constant, Président Pertimm France*

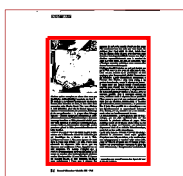
Le Web 3.0 a été défini, il y a déjà plusieurs années par Tim Berners Lee, en référence au Web sémantique. Pour définir ce dernier en quelques mots, nous dirons que c'est « simplement la capacité de stocker du "sens" dans les pages Web et de "traiter" ce sens ».

Mais cette définition n'est sans doute pas suffisante. En effet, qu'est-ce que le « sens », ou la « signification », mots qui ont la même racine que « sémantique » ? Cette question – de prime abord naïve – est totalement justifiée. En effet, les mots, les phrases, les images, en bref, le contenu d'une page ne sont-ils pas déjà remplis de sens ?

Et c'est bien le cas pour chacun d'entre nous : nous parvenons tous à lire et à comprendre – au moins en partie – une page Web. En fait, et sauf le cas d'une page vide, chaque page a sa signification propre, signification donnée par l'auteur. Et même dans le cas d'une page remplie de lettres disposées aléatoirement, il y a tout de même le sens du vide ou de l'aléatoire, le sens de ce qu'est une page Web, de ce qu'est le Web, ce qui permet de comprendre qu'on a en face de nous une page Web

vide, ou remplie aléatoirement de lettres, etc. De plus, la signification même d'une page sera différente en fonction de son lecteur. Pour quelle raison ? Simplement parce que le processus consistant à établir une signification à partir d'un texte a lieu dans la tête d'une personne, jamais dans un ordinateur. L'ordinateur transforme seulement les données dans d'autres formes, parfois plus lisibles (résumé de texte, liste de résultats de recherche, etc.) parfois moins lisibles (codes exécutables, etc.). Il s'agit là d'un point fondamental du Web sémantique. En un sens, le Web est totalement sémantique depuis le début, puisque chacun y trouve toute la signification qu'il veut. Ainsi, parler de Web sémantique, c'est parler en tautologie.

Alors, la question du Web sémantique est-elle donc vaine ? Dans le sens indiqué précédemment oui, absolument ! Cependant, on peut poser la question différemment : comment améliorer les programmes travaillant sur le Web afin qu'ils vous permettent de mieux « comprendre » les données qu'ils vous présentent, et qu'ils « démontrent » une meilleure capacité à « transformer » (pour ne pas dire comprendre) votre demande en des



Patrick Constant

résultats qui ne sont plus au niveau des mots que vous tapez, mais davantage au niveau du sens ?

Un exemple pour illustrer notre propos est celui du libraire. Vous cherchez un livre, et vous allez donc rencontrer votre libraire favori. Plus vous décrivez le livre recherché, plus vite le libraire trouvera la réponse à votre recherche. Ce processus d'échange et de compréhension – le propre d'un échange humain – résume finalement exactement et complètement la notion de Web sémantique : plus vous vous exprimez face au Web, plus celui-ci doit pouvoir vous répondre précisément et intelligemment. Ainsi, le Web sera d'autant plus sémantique que les processus gérant le Web pourront fonctionner de cette manière.

N'est-ce pas déjà le cas ? L'exemple le plus prédominant sur le Web est le moteur de recherche, qui est l'archétype du « libraire » sur le Web. Qu'observons-nous ? Plus nous tapons de mots dans la petite case dédiée à cet effet, moins nous obtenons de réponses. Mais est-ce que ces réponses sont meilleures ? Non. L'expérience est très commune. On mesure d'ailleurs que le nombre de mots moyens tapés est de l'ordre de 2,5 et que ce nombre reste inchangé depuis plusieurs années (il est passé de 1,5 avec Altavista à 2,5 pour les moteurs actuels, et s'est stabilisé). Pourquoi cette stabilisation ? Parce que la plupart des

moteurs de recherche actuels cherchent des pages Web contenant tous les mots que vous tapez. En pratique, plus vous tapez de mots, moins vous avez de chance de les trouver tous dans une page donnée. Ainsi, la probabilité de trouver « La Page » diminue exponentiellement avec le nombre de mots que vous tapez. Au-delà de trois mots, vous n'êtes que rarement satisfait, et nous venons d'en donner les raisons théoriques.

D'ailleurs, il suffit d'ajouter un mot qui n'existe pas sur le Web pour ne plus avoir de réponses du tout. C'est un peu comme si, en prononçant un mot magique, votre libraire restait pétrifié ! Les moteurs de recherche actuels fonctionnent exactement à l'inverse d'un libraire : plus vous parlez à un moteur de recherche, moins la réponse est bonne. Il existe donc une grande marge d'amélioration possible pour obtenir des moteurs de recherche qui seraient plus « sémantiques », simplement en les faisant fonctionner comme le libraire. Avec ce nouveau modèle, plus la personne saisirait de mots, meilleure serait la réponse. Est-ce possible ? Nous pensons que c'est non seulement possible, mais que ces moteurs constitueront la nouvelle génération de moteurs de recherche. Ils pourraient être, à juste titre, qualifiés de « sémantiques » car leur fonctionnement se rapprocherait fortement d'un libraire, et non, comme c'est le cas actuellement, selon un fonctionnement inverse.

La société Pertimm a développé ce type de technologie, qui est déjà opérationnelle sur des Intranets de plusieurs millions de pages. Notre challenge actuel est de faire la même chose sur les milliards de pages de l'Internet. Les utilisateurs disposeront alors d'un véritable Web sémantique, motorisé par des outils sémantiques.

Qu'en est-il alors du Web 3.0 ? Le Web 3.0 est simplement la connexion entre votre manière d'exprimer une idée, un thème, et la manière dont une autre personne a exprimé cette même idée. Ainsi, le Web 3.0 nécessite une révolution technologique : on ne peut plus se contenter de gérer des mots (clés ou non), il faut gérer des vocabulaires. ■

* www.pertimm.com. La société a reçu en 2004 le prix Label Start-up innovante de l'Anvar.