

Le web sémantique comme future solution de gestion de l'information

« Ce n'est plus de la recherche. Le web sémantique est sorti des laboratoires, des communautés se sont fondées autour des technologies du web sémantique, des petites et grosses entreprises l'utilisent en interne comme Oracle, IBM ou des institutions comme la BBC. » Pour **Nicolas Chauvat**, fondateur de Logilab, entreprise spécialisée dans la gestion des connaissances et l'informatique avancées, le web sémantique, encore appelé Web 3.0 ou web des données, est aujourd'hui une réalité.

C'est pour affirmer cette réalité et accélérer les développements et adoptions que Logilab organise, pour la première fois en France, le SemWeb.Pro (les 17 et 18 janvier 2011, Paris 14e). Ces conférences, tables rondes et travaux pratiques, plutôt destinés à un public d'experts, abordent tant les technologies et outils dédiés que les cas pratiques mis en oeuvre et les retours d'expérience.

L'origine de l'intégration des données structurées ne date pas d'hier. L'idée d'un web structuré remonte à 1994 avec la création du W3C animé par Tim Berners-Lee, père du web, confirmé en 1998 (avec le début des travaux en 1999). Il faudra quelques années de travail pour définir les premiers protocoles et formats.

Une montée en puissance industrielle

A partir de 2004 apparaissent plusieurs protocoles et formats visant à structurer les données : RDF (Resource Description Framework, langage de base du Web Sémantique), son complément le RDF Schema (qui regroupe processus et outils pour définir les ontologies qui structurent les ressources RDF); l'OWL (Web Ontology Language, qui définit le vocabulaire RDF) et SPARQL (langage fondé sur RDF pour interroger des ressources de données). Autant d'outils qui, s'ils permettent de structurer les données, restent à généraliser dans les outils de gestion de l'information. Les grands groupes ne s'y trompent pas et « nous voyons une montée en puissance industrielle depuis 2007 », assure **Fabien Gandon**, chargé de recherche à l'Inria (Institut national de recherche en informatique et automatique).

Pour lui, six enjeux principaux sont au centre du développement du Web 3.0. D'abord, la confirmation des standards, qui bénéficient des retours de 6 ans d'expérience, et redéfinis dans la RDFa 1.1. Ensuite, l'implémentation massive du Web sémantique à travers les grandes entreprises (Oracle, IBM mais aussi Yahoo même si la question se pose depuis l'intégration de son moteur de recherche dans Microsoft Bing, ou encore Google qui semble faire du «websem» sans l'avouer). « Le déploiement complet nécessite des outils au point », précise le chercheur.

Créer un écosystème autour des données

La mise en ligne des données publiques figure également parmi les scénarios les plus importants pour contribuer à la diffusion de la technologie. Mais la question est forcément politique et sa diffusion risque d'en être d'autant retardée que les intérêts des uns (les citoyens notamment) ne sont pas forcément ceux des industriels (qui souhaitent monétiser cette valeur de l'information).

«Nous avons énormément de mal à faire comprendre que l'on va pouvoir créer un écosystème autour des données et non pas de la valeur sur le traitement en temps réel de l'information», regrette Nicolas Chauvat. Autre enjeu, les compétences : « Un coup de fil que je reçois sur deux concerne une recherche de profil, témoigne Fabien Gandon. Il y a une vraie demande d'ingénieurs et techniciens du web sémantique, et aussi des besoins en décideurs. » Autrement dit, tant que la masse critique de compétences ne sera pas atteinte, point de salut pour le web sémantique.

Un obstacle que la propagation de la dynamique pour aller au-delà de la mise en oeuvre des standards permettra d'atteindre. Ce qui implique évidemment de poursuivre les travaux de recherche notamment à travers le passage à l'échelle des traitements, la prise en compte des nouveaux usages comme la mobilité et les réseaux sociaux, mais aussi la qualité des données et la nécessité de maintenir une interaction aussi simple que possible à travers des interfaces adaptées pour démocratiser l'utilisation du web sémantique.

Il en va de l'avenir des entreprises confrontées à des montagnes de données, ce qui tend à se généraliser passé une certaine taille. « Les technologies du web sémantique sont en train d'apparaître comme une solution de gestion de l'information », conclue le porte-parole de l'Inria.