

Le web sémantique, une opportunité pour le droit et ses acteurs ?*

Je suis très heureux de vous présenter aujourd'hui le concept de web sémantique et plus largement celui d'information sémantique. Je ne le ferai pas à titre d'informaticien, ne l'étant pas. Ni au titre d'avocat possédant des dizaines d'années d'expérience dans le domaine, ne l'étant pas non plus. En revanche, je le ferai en qualité de juriste issu de la génération Y, celle qui n'a pas connu ou presque, le monde sans les nouvelles technologies.

Dans cette présentation nous introduirons la notion de web sémantique, à travers ses origines et son objectif. Puis nous verrons quelles en sont ses applications pratiques, et enfin nous analyserons les opportunités qu'offre cette évolution pour le droit et ses acteurs.

Origines

L'apport extrêmement positif qu'internet a eu jusqu'à présent dans de nombreux domaines pourrait être décuplé par la généralisation de systèmes d'information dits sémantiques¹, mécanismes propres à l'idée de web 3.0. Mais avant de nous intéresser plus en détail à ce concept nous allons présenter brièvement un panorama historique de l'évolution de nos moyens de communication et de la gestion de l'information dans nos sociétés. Cela nous permettra de mieux comprendre les fondements qui sous-tendent ce concept de web sémantique.

Une évolution historique

On peut diviser en quatre grandes périodes, l'histoire de l'évolution de nos moyens de communication et de gestion de l'information².

La première phase est celle de l'oralité, dans laquelle l'information n'avait pour support que

* Gilles de SAINT-EXUPÉRY, g.de.saint.exupery@umontreal.ca, Legal 5.0, 4 avril 2011, Montréal.

¹ Voir aussi sur le sujet: Pompeu CASANOVAS, Joost BREUKER, Aldo GANGEMI, *Law and the semantic web*, Springer, 2005, disponible à <http://tinyurl.com/69o5n9m>, site visité le 5 avril 2011. Et l'article de Miche DUMAIS, <http://www.ledevoir.com/societe/science-et-technologie/59857/technologie-le-web-de-l-avenir-etre-ou-ne-pas-etre-semantique>, site visité le 5 avril 2011.

² Richard SUSSKIND, *The end of lawyers ? Rethinking the nature of legal services*, p.17.

notre mémoire et nos paroles. Dans une deuxième phase, l'écrit commence à apparaître de façon sporadique et limitée, c'est l'âge des scripts. La troisième période qui symbolise la grande avancée vers une société de l'information a débuté par l'invention de l'imprimerie par Gutenberg au XV^{ème} siècle (1450). Les écrits deviennent accessibles au plus grand nombre (reproduction beaucoup plus simple et support plus pratique), mais il fallait encore que les populations apprennent à lire et à écrire, sans quoi ces nouveaux moyens de diffusion de l'information n'avaient que peu d'intérêt.

Pendant cette période deux évolutions se produisent. Une évolution technologique puisque nous avons amélioré notre savoir en matière de reproduction et de stockage des informations, mais aussi en matière de transports. Évolution sociale ensuite puisque cette nouvelle technologie est progressivement rentrée dans les moeurs. Il est alors important de réaliser que ces évolutions ont été progressives et lentes. Le développement d'une telle technologie ne s'est pas fait sans affronter embûches et réticences.

Les débuts d'une nouvelle ère

Aujourd'hui nous sommes entrés dans la 4^{ème} et dernière période, plus communément appelée l'ère des nouvelles technologies de l'information (NTIC), et caractérisée par l'extrême facilité d'accès et de partage de l'information. Et puisqu'on ne cesse de dire que l'histoire se répète, reprenons notre parallèle précédemment développé et appliquons le aux évolutions qui ont lieu actuellement.

D'un côté, une évolution technologique que l'on peut diviser en deux niveaux. Le premier niveau, a été celui de la maîtrise du partage de l'information, appelé le *data processing*. Nous avons développé des outils capables de transporter et de stocker des informations en quantité décuplée par rapport à ce que l'on a jamais connu. À l'heure actuelle nous rentrons dans le deuxième niveau, celui du *knowledge processing*. C'est-à-dire la maîtrise de l'exploitation des données³.

De l'autre côté, l'évolution sociale. Avec l'invention de l'imprimerie, les gens ont appris à lire et à écrire, à l'heure des TI, allons nous apprendre à coder ? Certes ces nouvelles technologies entrent dans les moeurs et leur utilisation devient la norme, mais leur maîtrise est loin d'être

³ R. SUSSKIND, *préc.*, note 2, p.7.

assurée. Que l'on soit l'adolescent ou l'avocat, il est nécessaire de perfectionner notre habileté à manier ces technologies tant pour des raisons de sécurité, que d'efficacité.

Pourquoi ce panorama historique? Car il nous semble nécessaire de bien comprendre que nous nous trouvons à un moment charnière de l'évolution, qui n'est que le début d'une nouvelle ère. Comme le rappelle Richard Susskind⁴, le réseau global du web est en place et commence à être bien intégré dans notre société, mais ceci n'est que le début de la révolution numérique. En effet la vraie avancée commence maintenant, d'autant plus dans le domaine juridique. Quel est le lien avec le concept de web sémantique me direz vous ?

L'apparition du concept de web sémantique

Le web incarne l'entrée de notre société dans l'ère de l'information. Nous pouvons maintenant accéder à une quantité d'information en tout temps, sans avoir à nous déplacer dans un lieu spécifique. Mais le corollaire inévitable à cette démultiplication de l'information, est le trop plein d'information. En effet nous sommes maintenant débordés par le contenu informationnel du web, tout comme les entreprises sont débordées par le flot d'information qu'elles ont à traiter. D'ailleurs, nombreuses sont les études qui montrent que les employés d'une entreprise passent globalement un tiers de leur temps à rechercher des informations.

D'où l'apparition du concept de web sémantique que l'on pourrait élargir au concept d'information sémantique, pour les systèmes d'informations des entreprises par exemple. Le web sémantique, c'est l'évolution logique de l'ère des technologies de l'information. Alors qu'il a été nécessaire d'apprendre aux gens à lire et à écrire pour exploiter les documents disponibles grâce à l'invention de l'imprimerie, il est aujourd'hui nécessaire de mettre en place des techniques qui permettent l'exploitation des données, toujours plus nombreuses sur le web et dans les différents systèmes d'informations. Exploitation qui pourra et devra être faite par les machines pour plus d'efficacité, de sécurité et de confiance! Le web sémantique et l'intelligence artificielle sont-ils la solution ? Sur le papier cela fonctionne, mais dans les faits, il va falloir encore s'atteler à la tâche.

Commençons par une définition du terme sémantique. Selon le Grand Robert la *sémantique* est l'*[é]tude des significations de tout système de communication, en particulier d'une langue*

⁴ R. SUSSKIND, *préc.*, note 2, p.22.

naturelle. Ce terme vient du grec *sémantikos*, qui se traduit littéralement par: « qui signifie ». Il s'agit donc d'expliquer un système de communication, de le rendre intelligible. C'est aussi ce à quoi on souhaite parvenir avec le web.

Le web sémantique est surtout et avant tout une théorie, un concept, visant à optimiser le potentiel des technologies de l'information. Cette initiative est issue de l'organisme W3C, et est notamment défendue par le créateur du web, Tim Berner-Lee.

Fonctionnement

En généralisant, le web sémantique, c'est la volonté de mettre de l'ordre dans le web et de permettre ainsi une restitution simplifiée des données stockées et donc de bénéficier au mieux de son potentiel. Il s'agit de rendre les données intrinsèquement intelligentes et du même coup exploitables par des agents logiciels grâce à des métadonnées. Alors que jusqu'à présent le web était principalement conçu comme un espace de stockage, il devient nécessaire d'organiser ce stockage sans quoi l'anarchie qui va y régner pourrait conduire à un désintéressement et une perte de confiance de ses acteurs.

"Le Web sémantique va utiliser la structure pour donner du sens au contenu des pages Web, en créant un environnement où les agents logiciels en parcourant les pages pourront réaliser rapidement des tâches compliquées pour les utilisateurs."⁵

Concrètement, comme l'explique Hervé Tigier⁶ (<http://websemantique.org>):

"Rendre sémantique le web, c'est compléter les balises HTML par des balises porteuses de sens (XML) ; c'est multiplier et organiser les mots-clés dans le document même (sans que le lecteur ne voit de différence) ; c'est substituer à certaines animations Flash des animations SVG dont le texte est lisible pour les moteurs de recherche, etc. À un niveau plus élevé, c'est travailler à l'indexation de toutes les ressources du web, documents et services (norme Resource Description Framework), pour une exploitation optimale."

⁵ Tim BEERNER-LEE, "The semantic web", disponible en version française à <http://www.urfist.cict.fr/archive/lettres/lettre28/lettre28-22.html>, site visité le 5 avril 2011.

⁶ Hervé TIGIER, disponible à <http://websemantique.org/GuideWikiDifferences>, site visité le 5 avril 2011.

Pour imager ces dires, le web 1.0 était l'équivalent d'une grande pièce dans laquelle chacun avait la liberté d'y déposer des données, sans système de classification, ni personne référente pour aider à la recherche. Le développement du web 2.0 va permettre aux utilisateurs de cette grande pièce d'échanger entre eux leurs données et donc améliorer à la fois leur classification, et leur restitution, mais la quantité de données augmentant considérablement il est apparu nécessaire de mettre en place un système automatisé de gestion, car il s'est avéré que la restitution de données était trop aléatoire. C'est là qu'intervient le web 3.0. La grande pièce dans laquelle chacun déposait ses données de manière anarchique va se voir doter d'un système de classification uniforme autogéré qui va faciliter les recherches. Plutôt que d'essayer d'améliorer les moteurs de recherche, on cherche à rendre les données intelligentes pour qu'elles soient plus efficacement exploitables par des *agents logiciels*.

Globalement le concept d'informations sémantique va apporter plus de clarté au web:

Web 1.0	Web 3.0
quantitatif	qualitatif
opaque	transparent
anarchique	indexé
(in)compréhensible par les humains	compréhensibles par les humains et les logiciels

En pratique quelle est la différence entre un moteur de recherche classique et un moteur de recherche sémantique ?

La différence entre ces deux types de moteur de recherche est leur façon d'effectuer les recherches. Le moteur de recherche classique, de type Google, se contente de scanner par mots clefs le contenu des pages web, tandis qu'un moteur de recherche sémantique, va détecter à la fois le sens de la demande, et celui du contenu des pages⁷.

Comme l'explique Beener-Lee⁸:

⁷ Michael ARRINGTON, disponible à <http://fr.techcrunch.com/2008/05/12/premiere-presentation-de-powerset-le-moteur-de-recherche-semantique/>, site visité le 5 avril 2011.

⁸ T. BEERNER-LEE, disponible à <http://www.moovie.fr/index.php/2009/02/17/google-et-les-rseaux-sociaux-actuels-pourraient-tre-dpasss-par-le-web-smantique-tim-berners-lee/>, site visité le 5 avril 2011.

“Imaginez si deux choses totalement distinctes, par exemple votre compte en banque et votre agenda, puissent parler le même langage et se partager des informations. Vous pourriez ainsi faire glisser les informations de votre compte en banque vers votre agenda et vous pourriez savoir à quelles dates précises vous avez dépensé cet argent (...) Avec le web sémantique, c’est comme si l’on attribuait à toutes les données une longitude et une latitude sur une carte, et que n’importe qui pourrait les mélanger (“mash”) et les utiliser afin de créer autre chose”

Mettons maintenant en image ces explications avec des résultats de recherche obtenus avec ces deux types de moteur. Si je pose à Google la question "Quel est l'âge de Jean Charest ?", j'obtiens les réponses suivantes (voir Slide). On voit bien que je n'ai pas directement la réponse à la question, il va falloir que je me rende sur wikipédia, que je regarde sa date de naissance et que je fasse le calcul. Si je pose la même question à un moteur de recherche sémantique, comme Wolfram|Alpha, j'obtiens directement la réponse (voir Slide).

Autre exemple si je pose à Google la question "Where am I ?", j'obtiens les réponses suivantes sur Google. Si je fais la même chose dans Wolfram, j'obtiens grâce à la géolocalisation ma situation.

Autre exemple de moteur de recherche sémantique, Qwiki. Ce moteur a la particularité de mettre en relation différents types d'informations multimédias, textes, photos, vidéos pour rendre la réponse beaucoup plus interactive et agréable. Pour exemple lorsque je tape montréal dans google, voilà ma page de résultat. La même chose dans Qwiki et j'obtiens un topo vidéo, audio sur la ville de montréal.

Ces exemples sont bien sur symboliques, mais ils permettent de mieux comprendre la manière dont sont destinés à fonctionner les futurs moteurs de recherche. A ce propose Beerner-Lee a même avancé dans un article en 2008 que Google pourrait ne pas rester n°1, à cause du développement de tels mécanismes⁹.

⁹ T. BEERNER-LEE, disponible à

http://technology.timesonline.co.uk/tol/news/tech_and_web/article3532832.ece, site visité le 5 avril 2011.

Seulement il faut se rappeler que nous ne sommes qu'au début de cette nouvelle phase qu'est le *Knowledge management* et comme l'explique Karl Dubost, qui participe activement à la promotion du web sémantique:

"Le web sémantique est à ses balbutiements, mais les premières pierres sont posées."¹⁰

Il est ainsi nécessaire de faire preuve de modération, car si en théorie le web sémantique semble être la meilleure solution pour faciliter la navigation et la restitution d'informations sur le web ou sur tout autre système d'information, en pratique, sa mise en place est complexe. La quantité d'information déjà disponible sur le web ou sur les intranets d'entreprises rend très difficile la mise en place d'une gestion de l'information sémantique à la main et le succès de l'opération trop aléatoire. Il faut d'ores et déjà espérer que le développement de l'intelligence artificielle puisse nous aider à effectuer ce travail fastidieux.

C'est ici qu'intervient l'intelligence artificielle, qui une fois couplée au concept d'information sémantique, permet aux machines de traiter des données, ou des informations y compris dans un langage naturel. Pour preuve, les récents exploits de Watson, l'ordinateur d'IBM le "plus intelligent" au monde sont de bon augure. En effet pour ceux qui n'auraient pas suivi cette affaire, lors d'une émission du jeu Jeopardy, Watson, l'ordinateur, a battu haut la main les deux champions américains de l'émission. Cet ordinateur a été conçu dans le but d'être capable de répondre à des questions dans le langage naturel grâce à la sémantique. On voit donc bien que ces concepts ne vont surement pas passer inaperçus dans les années à venir !

Autre difficulté. Pour que la mise en place de systèmes sémantiques fonctionne correctement, surtout sur le web, il faut définir des standards sémantiques de métadonnées, ou à tout le moins des grilles d'équivalences entre les standards au niveau international. La tâche est donc titanesque.

Enfin, le principal danger pointé par les spécialistes au sujet de système d'information sémantique, est l'efficacité accrue du contrôle des informations disponibles. En effet une

¹⁰ Karl DUBOST, disponible à <http://www.w3.org/2000/Talks/1019-rdf-poly/slide31-0.html>, site visité le 5 avril 2011.

meilleure recherche signifie une meilleure surveillance des systèmes d'information et donc une éventuelle censure de l'information. Mais ne poussons pas la paranoïa.

Le web sémantique et son impact sur le droit et ses acteurs

Je voudrais pouvoir être plus précis en matière d'applications concrètes, pratiques de ce concept dans le monde juridique, mais si le concept n'est pas très avancé chez les spécialistes de TI ou chez les informaticiens, il l'est encore moins chez les juristes. Ce qui semble certain c'est le potentiel que de tels systèmes promettent d'offrir aux différents acteurs du droit, du juge à l'avocat en passant par les simples citoyens.

Tout d'abord, ce n'est pas tant le développement du web sémantique qui va avoir un impact sur les acteurs du droit, mais plutôt l'implantation de système d'information sémantique dans les cabinets ou dans les tribunaux qui va permettre d'améliorer l'efficacité du travail des acteurs en questions. En effet la recherche de l'information pertinente (lois, jurisprudences, règlements, contrats...) pour les acteurs du droit est une tâche fastidieuse, longue, parfois aléatoire que ces nouveaux mécanismes vont faciliter, mais ils vont aussi surtout permettre une meilleure gestion et un contrôle renforcé des informations.

Quel est l'avenir du métier d'avocat dans ce contexte ? Loin de parler de la *mort* de l'avocat, dans son livre *The end of lawyer*, Susskind¹¹, défend l'idée que l'avocat des prochaines décennies sera non seulement plus performant, mais aussi un meilleur gestionnaire. Car les systèmes dits sémantiques peuvent non seulement aider à la recherche juridique, mais aussi à la gestion d'un cabinet, de sa clientèle, etc. Pour Susskind, ce changement n'interviendra pas nécessairement sous l'impulsion des cabinets d'avocats, réticents à investir encore et toujours dans la technologie, mais plutôt lorsque les clients des cabinets feront pression sur ces derniers. Nous dirons que la réussite de la conférence est l'exception qui confirme la règle, à voir le nombre de participants intéressés par ces sujets ! La balle est en tout cas dans le camp des avocats qui adoptent les technologies, preuve en est notamment des différents conférenciers de la journée !

¹¹ R. SUSSKIND, *préc.*, note 2, p.282

Enfin, l'ultime objectif du développement d'une telle technologie est la mise en place de systèmes automatisés intelligents qui pourraient résoudre des problèmes juridiques posés en langage naturel, à l'image de Watson. Nous sommes évidemment aux prémises d'un tel rêve.

Pour finir le développement de systèmes sémantiques, et en l'espèce plus particulièrement du web va rendre le droit et les données publiques plus accessibles aux populations, à l'image des différents organismes qui prônent l'ouverture des données, comme la coopérative Accessibilité Web ou le collectif Montréal Ouvert. Dès lors la présomption " Nul n'est censé ignorer la loi" s'approchera peut-être plus de la réalité ! Et si il n'est pas rare de voir dans la rue des graffitis réclamant la libéralisation des drogues, je m'en tiendrais pour ma part à réclamer celle de droit !

Conclusion

Le web sémantique nous paraît être une problématique d'avenir inévitable et favorable aux différents acteurs du droit qui auraient tout à gagner à se pencher dès à présent sur la problématique au lieu d'attendre passivement que d'autres le fassent, car comme le dit Michel Serres¹², *les nouvelles technologies nous ont condamné à devenir intelligent*, maintenant que nous disposons d'outils au potentiel inégalé pour gérer nos savoirs !

¹² Michel Serres, « Les nouvelles technologies : révolution culturelle et cognitive », Conférence INRIA 2007, en ligne: http://interstices.info/jcms/c_33030/les-nouvelles-technologies-revolution-culturelle-et-cognitive