

# Web sémantique : Quelles perspectives pour l'information juridique?

Teresa Bascik\*

Le Web sémantique, ayant pour ambition de remplacer le « Web des documents » de plus en plus difficile à exploiter en raison de sa croissance exponentielle, prend finalement son envol après plusieurs années de recherches concertées. La théorie à la base de ce mouvement et les standards proposés deviennent de plus en plus reconnus par l'industrie des nouvelles technologiques et sont désormais utilisés pour le développement d'applications nécessitant un cadre commun d'analyse pour publier, échanger et restituer l'information provenant de sources hétérogènes. Bien que la recherche sur l'automatisation du traitement de l'information juridique ne soit pas nouvelle, la formalisation sémantique de ce corpus documentaire par le biais d'ontologies et sa structuration en langage *Resource Description Framework* (RDF) apportent des résultats de plus en plus prometteurs. Cette nouvelle approche apporte des réponses concrètes pour des grands enjeux de gestion de l'information juridique, en particulier en ce qui concerne la modélisation, la normalisation et l'interopérabilité de diverses bases de données.

Dans cet article, nous introduisons le cadre théorique du Web sémantique et définissons les notions essentielles qui s'y attachent. Nous présentons ensuite les apports potentiels des standards du Web sémantique dans la gestion de l'information juridique avant de discuter plus spécifiquement de l'enjeu de l'ingénierie

The Semantic Web, that aims to replace the «Web of documents» increasingly difficult to operate due to its exponential growth, finally takes off after several years of collaborative research. The theory behind this movement and the standards it proposes become increasingly recognized by the information technology industry and are now being used for the development of the applications that require a common framework to publish, share and restore the information from heterogeneous sources. Although research on automating the processing of legal information is not new, semantics formalization of this documentary corpus through ontologies and by the means of the *Resource Description Framework* (RDF) provide promising results. This new approach proposes concrete answers to the major challenges of managing legal information, especially modeling, standardization and interoperability of different legal information databases.

In this article, we introduce the theoretical framework of the Semantic Web and we define related notions. We then present the potential contribution of the Semantic Web standards in the management of legal information before discussing more specifically the issue of legal engineering ontologies for the explicit representation of the legal information resources. The question that drives our research is the following: can legal ontologies, designed by the knowledge engineers in order to reduce the

(2014) 19:1 [Lex-Electronica.org](http://Lex-Electronica.org) 25

Copyright © 2014 Teresa Bascik.

\* Teresa Bascik est diplômée d'une maîtrise en sciences de l'information (M.S.I.) et possède une vaste expérience en modélisation de systèmes de gestion stratégique de l'information. Elle est actuellement analyste au sein d'une société privée spécialisée dans le domaine la gestion des connaissances.

des ontologies juridiques permettant la représentation explicite des contenus des ressources documentaires. La question qui anime notre recherche est la suivante : les ontologies juridiques produites par des ingénieurs de connaissances en vue de réduire la complexité du langage juridique et de faciliter le stockage et l'accès aux documents juridiques peuvent-elles englober la complexité des concepts juridiques sans imposer un formalisme réducteur?

Un bref survol de nouvelles perspectives méthodologiques de construction des ontologies viendra en appui à l'hypothèse que le cadre théorique du Web sémantique, basé sur une structuration formelle des concepts juridiques, peut mener à un meilleur accès à l'information législative et est susceptible d'accroître l'interopérabilité et de faciliter l'échange d'informations entre les institutions. Les travaux sur le développement du *Legal Knowledge Interchange Format* (LKIF) seront évoqués afin d'illustrer ce propos.

complexity of legal language and facilitate the storage and access to legal documents encompass the complexity of legal concepts without imposing a binding formalism?

A brief overview of new methodological approaches for building legal ontologies will support the hypothesis that the theoretical framework of the Semantic Web, based on a formal structure of legal concepts can lead to a better access to legislative information and is likely to increase interoperability and facilitate the information exchange between institutions. The development of the *Legal Knowledge Interchange Format* (LKIF) will be discussed to illustrate this hypothesis.

---

<b>1. Esquisse de la problématique</b>	<b>27</b>
1.1. Le Web actuel	27
1.2. Web sémantique en quelques mots	28
1.3. La « sémantique » de l'information juridique	30
<b>2. Qu'est-ce qu'une ontologie ?</b>	<b>32</b>
2.1. L'historique du concept	32
2.2. Les limites de faisabilité	34
<b>3. L'ontologie juridique</b>	<b>34</b>
3.1. Les fondements de l'approche	34
3.2. Méthodologies identifiées	36
3.3. Projet en cours	37
<b>Conclusion</b>	<b>37</b>
<b>Bibliographie</b>	<b>39</b>

# Web sémantique : Quelles perspectives pour l'information juridique?

Teresa Bascik

Il est essentiel que les paroles des lois réveillent  
chez tous les hommes les mêmes idées.

Montesquieu<sup>1</sup>.

## 1. Esquisse de la problématique

### 1.1. Le Web actuel

Le Web que nous connaissons actuellement composé d'ensembles documentaires non structurés devient de plus en plus anarchique<sup>2</sup>. Malgré la convivialité que nous confèrent les moteurs de recherche couramment employés, la recherche d'information réellement pertinente, devant l'expansion exponentielle du Web, peut s'avérer une tâche ardue. Même si les résultats de nos recherches sont parfois étonnamment précis, force est de constater que le Web n'est qu'un amas gigantesque de données en tout genre et qu'il n'est pas exploité de manière optimale<sup>3</sup>. Ainsi, le défi d'identifier et d'extraire les données pertinentes dans cet immense univers en constante expansion peut s'avérer difficile.

1. Charles Louis de Secondat, Baron de Montesquieu, *De l'esprit des lois*, Paris, Gallimard, 1995, Livre XXIX, p. 533.
2. Kimberly Claffy *et al.*, "The (un)Economic Internet?", *IEEE Intelligent Systems*, May-June, 2007, p. 53. "Because no systemic measurements activities exist for collecting rigorous empirical Internet data, in many ways, we don't really know what the Internet actually is. Thus, we don't know the total amounts and patterns of data traffic, the Internet's growth rate, the extent and locations of congestion, patterns and distribution of ISP interconnectivity, and many other things that are critical if we're to understand what actually works in the Internet".
3. Selon étude de Michael K. Bergman, le web profond qui n'est pas indexé par les moteurs de recherche conventionnels constitue environ 95 % du web actuel ; Digimind, *Découvrir et exploiter le web invisible pour la veille stratégique*, 2008, p. 5, <http://www.esi.ac.ma/Dossiers/20140126070100.pdf>, consulté le 10 avril 2014.

La raison est bien simple : le langage de balisage HTML indique aux navigateurs comment afficher l'information, mais ne transmet aucunement sa portée sémantique. De plus, les liens hypertextes qui mettent en relation les pages Web entre elles ne permettent pas aux agents logiciels d'extraire et de relier cette information automatiquement. Ainsi, le besoin d'une formalisation logique permettant de représenter l'information actuellement contenue dans les documents balisés en format HTML, de façon non ambiguë et manipulable par des agents logiciels devient de plus en plus évident. Cette idée a été introduite par Tim Berners-Lee, créateur du web et principal théoricien de l'organisme de normalisation World Wide Web Consortium (W3C) :

*I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The intelligent agents people have touted for ages will finally materialize<sup>4</sup>.*

Le « rêve » de Tim Berners-Lee consiste à rendre les métadonnées « sémantiques » et à permettre aux agents logiciels de les manipuler de manière « intelligente »<sup>5</sup>. Ces derniers pourraient alors extraire et analyser les métadonnées contenues sur le Web avec plus de profondeur et de précision en réalisant la plupart des tâches fastidieuses de la recherche et du traitement de l'information qui reviennent actuellement aux humains. Mais de quelle sémantique s'agit-il au juste?

## 1.2. Web sémantique en quelques mots

Le Web sémantique<sup>6</sup> est avant tout une théorie basée sur le postulat que l'information consignée dans les silos isolés du Web actuel peut être exploitée de manière plus efficace. Issue de l'organisme de normalisation W3C, cette théorie vise à optimiser le potentiel des technologies de l'information en ajoutant au Web actuel une couche de métadonnées supplémentaires qui permettrait de transformer la masse ingérable des pages Web en une sorte d'index structuré facilitant la recherche et la restitution de l'information. La définition du Web sémantique communément acceptée est la suivante :

*The Semantic Web is about **common formats** for integration and combination of data drawn from diverse sources, where the original Web mainly*

- 
4. Tim Berners-Lee et Mark Fischetti, *Weaving the Web. The original design and ultimate destiny of the World Wide Web, by its inventor*, San Francisco, Harper, 1999, p. 13.
  5. Barry Smith et Christopher Welty, *Ontology: Towards a New Synthesis*, dans : "Proceedings of the international conference on Formal Ontology in Information Systems", New York, 2001, p. 4.
  6. Il est à noter que pour les besoins de cet article, les termes « web sémantique » et « web des données » sont utilisés comme des synonymes.

*concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects<sup>7</sup>.*

Cette explication peut paraître complexe, mais l'idée qui l'anime demeure, somme toute, très simple : il s'agit d'établir un **format universel** d'échange d'information et un **langage** permettant d'organiser les silos d'information par le biais d'une grammaire (standard RDF) et d'un vocabulaire commun (ontologies en format OWL). Tout comme dans une langue naturelle, ce vocabulaire commun et cette grammaire universelle permettront la création et la transmission du sens des documents sur le Web. En d'autres termes, contrairement au Web actuel, dont le cadre formel ne dépasse pas le niveau purement syntaxique, le Web sémantique vise la construction des relations raisonnées entre les métadonnées provenant de sources diverses.

Il est à noter que cette idée d'un langage universel permettant de rendre nos recherches d'information plus efficace est à l'origine d'une quête très ancienne exprimée notamment par Gottfried Wilhelm Leibnitz :

*I am working on the point where it is sufficient both to provide a grammar of such a wonderful language and a dictionary for most of the frequent items, to the point of having characteristic numbers for all ideas (...) Once the characteristic numbers of most notions are determined, the human race will have a new kind of tool, a tool that will increase the power of the mind much more than optical lenses helped our eyes, a tool that will be as far superior to microscopes or telescopes as reason is to vision. The compass newer provided navigation with anything more sufficient what this North Star would give us for swimming the sea of experiments<sup>8</sup>.*

Toutefois, le contexte technologique n'a jamais été aussi favorable qu'aujourd'hui pour la réalisation de cette « étoile du Nord » orientant nos quêtes d'information. Bien que le « rêve » d'un Web entièrement sémantique soit loin de devenir une réalité, son cadre théorique est désormais largement reconnu par l'industrie des nouvelles technologies<sup>9</sup>. Les standards qui sont à la base de cette vision sont déjà utilisés pour le développement des systèmes de gestion de l'information spécialisés résultant dans des moteurs de recherche plus efficaces, des applications plus intelligentes et bien d'autres améliorations à notre capacité actuelle de récupérer et traiter les flux de données<sup>10</sup>.

---

7. W3C, *Semantic Web Activity*, <http://www.w3.org/2001/sw/>, consulté le 10 avril 2014.

8. Gottfried Wilhelm Leibniz, *Philosophical essays*, Indianapolis, Hackett Publishing, 1989, p. 8.

9. David Siegel, *The Power of the Semantic Web to Transform Your Business*, New York, Penguin Group, 2009, p. 236.

10. Citons trois exemples : (1) métadonnées de BBC; "Tom Scott, Case study: Use of Semantic Web Technologies on the BBC Web Sites", *BBC Internet Blog*, [http://www.bbc.co.uk/blogs/legacy/bbcinternet/2010/02/case\\_study\\_use\\_of\\_semantic\\_web.html](http://www.bbc.co.uk/blogs/legacy/bbcinternet/2010/02/case_study_use_of_semantic_web.html), consulté le 10 avril 2014; (2) métadonnées bibliographiques dans le cadre du projet Drug Ontology Project for Elsevier (DOPE), Anita de Waard *et al.*, W3C, *Semantic Web Use Cases and Case Studies*, <http://www.w3.org/2001/sw/sweo/public/UseCases/Elsevier/>, consulté le 10 avril 2014 ; (3) système de gestion des dossiers

### 1.3. La « sémantique » de l'information juridique

Le Web est déjà une source importante d'information juridique<sup>11</sup> : on y trouve des textes législatifs, une grande quantité de jurisprudence, ainsi que des vastes répertoires doctrinaux. De plus, non seulement les textes réglementaires finaux, mais aussi les documents préliminaires, tels des propositions, des amendements ainsi que des transcriptions de débats parlementaires, sont pris en charge par des outils informatiques. Le corpus documentaire constitué de l'ensemble de ces activités se compose de documents électroniques qui sont, ou qui peuvent devenir, des composants du Web.

Cette information mise à disposition du public est souvent répertoriée soit comme un simple flux de texte balisé en HTML<sup>12</sup>, ou pire encore, comme des ensembles de documents PDF non pénétrables<sup>13</sup>. Ainsi, le corpus documentaire d'information juridique est désormais un gigantesque amas des documents sans aucune structure logique. La seule façon de l'exploiter est bien souvent d'utiliser des mots-clés, accompagnés parfois d'opérateurs booléens pour extraire des données selon la date, le tribunal ou d'autres critères. Les bases de données juridiques commerciales sont davantage structurées, mais la recherche est encore menée essentiellement par le biais de mots-clés qui se fondent sur l'appariement syntaxique et ne permet de retrouver que des documents qui contiennent des chaînes de caractères contigus exacts.

Or la documentation juridique est très fortement polysémiotique et son abord dépend de multiples contextes sociaux et culturels. De plus, sa prétention à la clarté n'exclut pas l'ambiguïté engendrée, entre autres, par de multiples abréviations, des sigles, des formulations différentes en fonction du contenu. Ce type de corpus textuel est donc très difficilement exploitable par les moteurs de recherche en plein texte. Ainsi, nos recherches d'information dans ce gigantesque corpus non structuré

---

judiciaires NetCase; Jesús Contreras et Marta Poblet, *NetCase: An Intelligent System to Assist Legal Services Providers in Transnational Legal Networks*, dans V. Richard Benjamins et al., *Law and the Semantic Web Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*, Lecture Notes in Artificial Intelligence 3369, Berlin Heidelberg, Springer, 2005, p. 218–232.

11. Nous définissons la notion de l'information juridique selon la *Déclaration de Montréal sur l'accès libre au droit* : « L'information juridique publique regroupe l'information juridique émanant des organismes publics qui ont l'obligation de produire le droit et de le rendre public. Elle inclut les sources primaires du droit, comme la législation, la jurisprudence, les traités ainsi que diverses sources secondaires ou interprétatives publiques comme le compte-rendu des travaux préparatoires, les rapports visant la réforme du droit et ceux résultant des commissions d'enquête. Elle inclut également les documents juridiques qui résultent de financements publics ». <http://www.fatlm.org/declaration>, consulté le 10 avril 2014.
12. John McClure, *Legal-RDF Vocabularies, Requirements & Design Rationale*, Proceedings of the V Legislative XML Workshop, 2006, p. 150.
13. Ce constat a été notamment exprimé par les chercheurs du mouvement *LegalRDF*, <http://www.hypergrove.com/legalrdf.org/>, consulté le 2 avril 2014.

demeurent fortement aléatoires et la tâche la plus fastidieuse de trier et de sélectionner l'information ainsi obtenue revient aux humains et non pas aux agents logiciels.

Devant cette surcharge informationnelle et l'insuffisance de nos méthodes de gestion de l'information actuellement en usage, il convient de constater la nécessité de développer et de mettre en œuvre de meilleurs outils permettant de rendre cette masse documentaire exploitable de manière adéquate<sup>14</sup>. Toutefois, les initiatives européennes visant, entre autres, à centraliser le corpus de jurisprudence dans une base de données unique démontrent que ce type d'approche est un chemin périlleux à entreprendre<sup>15</sup>. Les recherches actuelles visent plutôt à enrichir les corpus d'information juridique d'une couche de métadonnées supplémentaire permettant à des systèmes informatiques de « comprendre » leur sens<sup>16</sup>. En d'autres termes, les efforts se concentrent davantage à identifier les concepts contenus dans les documents juridiques et à les exposer aux nombreux outils émergents du Web sémantique permettant non seulement un meilleur repérage, mais également l'automatisation d'un grand nombre de processus de traitement de l'information juridique.

Afin d'explicitier cet enjeu, prenons comme exemple un paragraphe de la *Loi sur la protection du consommateur*<sup>17</sup> :

```
<p>La présente loi s'applique à tout contrat conclu entre un consommateur
et un commerçant</p>
```

En examinant le code source de la page Web qui héberge ce paragraphe, nous nous apercevons que le langage HTML n'est qu'une sorte de structure syntaxique qui indique au navigateur que le texte entre les balises doit être affiché comme un seul paragraphe (balises <p>), mais que celui-ci reste muet quant à sa portée sémantique. L'idée qui anime le mouvement du Web sémantique est d'enrichir cette structure syntaxique et d'y ajouter des métadonnées indiquant la portée sémantique donc, en occurrence, d'identifier la signification que pourraient prendre dans ce contexte les concepts « consommateur » et « commerçant ». Ce même contenu pourrait donc être enrichi par le balisage « sémantique » :

```
<p>La présente loi s'applique à tout contrat conclu entre un consommateur
<!DOCTYPE rdf:RDF [<!ENTITY law "http://domain.tld/otherpath/law#" >]>
<owl:Class rdf:ID="Consumer">
  <rdfs:subClassOf rdf:resource="#Person"/>
</owl:Class>
```

14. Comme en témoigne la *Déclaration de Montréal sur l'accès libre au droit*, citée dans la note 11.

15. Council of the European Union, "Council Conclusions Inviting the Introduction of the European Case Law Identifier (ECLI) and a Minimum Set of Uniform Metadata for Case Law", *Official journal of the European union* (C/127) 1, Apr. 29, 2011, p. 2, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:127:0001:0007:EN:PDF>, consulté le 10 avril 2014.

16. Marc van Opijnen, *European Case Law Identifier: Indispensable Asset for Legal Information Retrieval*, dans : Maria Angela Biasiotti et Sebastiano Faro, *From Information to Knowledge – Online access to legal information: methodologies, trends and perspectives*, IOS Press, 2011, p. 4.

17. *Loi sur la protection du consommateur*, RLRQ c. P-40.1.

```

<owl:inverseOf>
  <owl:ObjectProperty rdf:ID="Merchant"/>
</owl:inverseOf>
et un commerçant
<!DOCTYPE rdf:RDF [<!ENTITY law «http://domain.tld/otherpath/law#» >]>
<owl:Class rdf:ID="Merchant">
  <rdfs:subClassOf rdf:resource="#Person" />
</owl:Class>
<owl:inverseOf>
  <owl:ObjectProperty rdf:ID="Consumer"/>
</owl:inverseOf>
</p>

```

À la lumière de cet exemple, nous pouvons constater que la couche de métadonnées sémantiques référant à une ontologie juridique en format *Web Ontology Language* (OWL) peut améliorer le travail des agents logiciels et leur donner la capacité d'accéder plus facilement aux contenus des ressources textuelles et à effectuer des raisonnements sur ceux-ci. Au croisement de la théorie du droit et du domaine de l'intelligence artificielle, la modélisation des ontologies juridiques soulève toutefois les questions suivantes : (1) Pouvons-nous saisir la complexité des concepts juridiques et en construire des ontologie? (2) Le cadre théorique du Web sémantique, incarnant l'espoir de mieux organiser des ensembles documentaires non structurés par le langage de représentation des connaissances formellement défini, permettrait-il d'améliorer la gestion de l'information juridique ou, au contraire, il conduira vers un formalisme réducteur? Telles sont les questions auxquelles je tâcherai à répondre dans les lignes qui suivent.

## 2. Qu'est-ce qu'une ontologie ?

### 2.1. L'historique du concept

Le terme « ontologie » (du grec ὄν désignant le participe présent du verbe « être » et λόγος désignant « discours »<sup>18</sup>) réfère d'abord vers la discipline philosophique dont l'origine semble provenir du désir humain de caractériser « l'être » et tout ce qui existe en général, c'est-à-dire de parvenir à une conceptualisation raisonnée du monde. Cette idée a été notamment exprimée par Aristote qui proposait de décrire « ce qui est » dans ses propriétés générales et dans ce qu'il peut y avoir d'absolu, par opposition à la seule considération de ses apparences ou de ses attributs séparés<sup>19</sup>.

18. Alain Rey (dir.), *Dictionnaire historique de la langue française*, Paris, Le Robert, 2012, p. 1369.

19. Aristote, *La Métaphysique. Livre Gamma*, 1003, 21-32, traduction de Gaston Colle, Louvain, Édition de l'institut supérieur de philosophie, 1931, p. 3.

La pensée ontologique a jalonné l'histoire de la philosophie occidentale en prenant diverses formes, mais « procédant toujours par des concepts *a priori* »<sup>20</sup> afin de démontrer « pourquoi et comment ils sont possibles (...) »<sup>21</sup>.

L'emploi du terme « ontologie » en informatique puise son origine dans les recherches en logique formelle, en intelligence artificielle et en sciences cognitives. Il a été introduit en 1980 par John McCarthy affirmant que les systèmes informatiques basés sur l'intelligence artificielle doivent préalablement « *list everything that exists, building an ontology of our world* »<sup>22</sup>. Le choix de ce terme n'a rien d'aléatoire. Dans ce contexte, le terme ontologie désigne un ensemble de concepts, censé être indépendant d'une langue naturelle donnée, qui permettrait d'automatiser certains processus de traitement d'information en imitant le raisonnement humaine. L'ontologie dans le domaine de l'informatique est également définie comme « une spécification formelle d'une conceptualisation d'un domaine, partagée par un groupe de personnes, qui est établie selon un certain point de vue imposé par l'application construite »<sup>23</sup>. Les concepts faisant partie d'une ontologie sont organisés de manière hiérarchique et entretiennent des relations de l'hyperonymie et de la subsomption qui, tout comme des langages de programmation orientés objet, se basent sur les mécanismes d'héritage indiquant aux agents logiciels que, sauf exception explicitement déclarées, les catégories inférieures héritent des propriétés et fonctions des entités « parents ». Une fois construite et validée par une communauté d'experts donnée, l'ontologie est détachée du langage naturel et traduite vers le langage de représentation RDF (*Ressource Description Framework*). Le fondement de ce dernier repose sur l'idée que toute information peut être structurée sous forme d'un triplet : sujet => prédicat => objet. Ses trois composants sont principalement des identificateurs de ressources uniques sous formes URI (*Universal Resource Identifier*) liant ainsi différents composants du Web de manière qui reflète leur portée sémantique. Ils permettent ainsi d'élargir les critères de recherche d'information sur la base de relations sémantiques contrôlées offrant ainsi des réponses concrètes aux problèmes de diffusion et d'exploitation des corpus textuels massifs et polysémiotiques.

20. Roger Verneaux, *Les sources cartésiennes et kantienne de l'idéalisme français*, Paris, Beauchesne, 1936, p. 461.

21. *Ibidem*, p. 461.

22. John McCarthy, "Circumscription – A Form of Non-Monotonic Reasoning", *Artificial Intelligence*, 1980, 5: 13, p. 27.

23. Patrick Séguéla, Nathalie Aussenac-Gilles, *Extraction de relations sémantiques entre termes et enrichissement de modèles conceptuel*, dans : *Ingénierie des Connaissances*, Paris, École Polytechnique, 1999, p. 10.

## 2.2. Les limites de faisabilité

Toutefois, la construction des ontologies, qui sont alors centrales pour le Web sémantique, exige une représentation formelle pourvue d'une sémantique dénotative qui repose sur les deux postulats dérivés du cognitivisme classique : (1) les connaissances sont discrètes et représentables<sup>24</sup>; (2) les connaissances sont indépendantes de leurs substrats sémiotiques et se manifestent dans les concepts abstraits et indépendants des langues naturelles<sup>25</sup>.

L'approche ontologique peut donc être perçue comme l'aboutissement du formalisme logique soutenant qu'il soit possible de représenter un monde objectif, indépendant de toute langue naturelle et de tout autre système de signes<sup>26</sup>. Ces enjeux ont été débattus dans le domaine de l'intelligence artificielle et de la représentation des connaissances depuis trois décennies. Toutefois, il est à noter que la modélisation ontologique, telle que préconisée actuellement par W3C et pratiquée dans l'industrie des nouvelles technologies, n'a pas pour ambition de spécifier le sens global des concepts qui y figurent, mais d'explicitier les relations qu'ils peuvent entretenir et de démontrer « pourquoi et comment ils sont possibles ».

En somme, il est possible de conclure que l'approche ontologique se contente actuellement de proposer une modélisation abstraite et simplifiée d'un domaine de connaissance qu'on souhaite représenter sur le Web ou, de manière générale, partout où il va être nécessaire d'appuyer des agents logiciels sur les métadonnées sémantiques extraites pour les besoins d'une application donnée.

## 3. L'ontologie juridique

### 3.1. Les fondements de l'approche

Le besoin d'une représentation formelle de l'information juridique permettant de rendre sa gestion plus optimale n'est pas récente, tout comme les tentatives de représenter les concepts juridiques afin qu'ils soient interprétables par l'ordinateur.

---

24. Il s'agit d'un postulat hérité du positivisme logique introduit par le Cercle de Vienne et décrit notamment par Rudolf Carnap, *The Logical Structure of the World*, Londres, Routledge & Kegan Paul, 1967. Ludwig Wittgenstein résume ce postulat dans sa célèbre phrase : « Le monde est tout ce qui a lieu », *Tractatus logico-philosophicus*, Gallimard, Paris, 1993, p.33.

25. Ces préceptes méthodologiques se basent sur la théorie générale de la terminologie élaborée par Eugen Wüster dans son ouvrage *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Vienne, Springer, 1979.

26. Ce type d'approche a été vigoureusement dénoncé par François Rastier, « Sémantique du web vs sémantique web ? », *Texto! Textes & Cultures*, Volume XIII - n°3, 2008, p. 4, <http://www.revue-texto.net/index.php?id=1729>, consulté le 12 avril 2014.

L'idée de reformuler le droit sous forme d'un système, basant sa démarche sur les conjonctions syllogiques <SI> => <ALORS> avec les précisions des entités et des règles qui régissent leurs interactions, a été vivement débattue au cours des années 1980, sans pourtant porter les fruits espérés<sup>27</sup>. De même, les recherches sur les techniques permettant de représenter et imiter le raisonnement juridique ont pris une place importante dans le domaine de l'ingénierie des connaissances et ont donné lieu à un vigoureux débat sur nos capacités d'extraire et représenter des concepts juridiques<sup>28</sup>. Les résultats de ces études démontrent clairement que le raisonnement juridique ne se résout pas à une logique formelle. Toutefois, il est possible de constater que :

*The Law is applied in legal sources, therefore it relies on documents. Documents are the basic infrastructure for legal operations: they have a physical perspective (the support), a representational one (the language), and a cognitive one (the intended content)*<sup>29</sup>.

L'approche ontologique se situe au niveau cognitif de cette typologique et vise à affranchir l'information juridique de la complexité et la diversité sémiotique des documents qui les manifestent. En référant aux concepts, et non pas au texte, et faisant des liens entre ces concepts, le cadre formel du web sémantique propose ainsi une stratégie de désambiguïsation et une règle d'explicitation contextuelle de l'information implicite. Grâce aux agents logiciels intelligents étant capables d'analyser et d'extraire le sens des concepts en prenant en compte le contexte de leur apparition, leur portée sémantique se forge dans un réseau multiple<sup>30</sup> et se renouvelle en faisant apparaître des significations temporaires et des polysémies exclues du langage formel lors de la modélisation. Il est donc possible de conclure que l'approche ontologique, visant à extraire les concepts juridiques, mais aussi à les contextualiser grâce au travail des agents logiciels intelligents, constitue une méthode grandement plus efficace que les systèmes de l'intelligence artificielle qu'on a tenté de construire dans les années 1980. Avec un degré de formalisation proposé par l'approche ontologique, les agents peuvent véritablement repérer l'information contenue dans les documents textuels au lieu de simplement récupérer des chaînes de caractère utilisées dans nos requêtes.

27. Alexander Boer *et al.*, *Specification of the Legal Knowledge Interchange Format*, Deliverable 1.1, 2007, p.6, <http://www.estrellaproject.org/doc/D1.1-LKIF-Specification.pdf>, consulté le 12 avril 2014.

28. Ce débat est notamment illustré par Joost Breuker, André Valente, et Radboud Winkels, *Use and Reuse of Legal Ontologies in Knowledge Engineering and Information Management*, dans : V. Richard Benjamins *et al.*, préc., note 10.

29. *Ibidem*, p.11

30. Nous empruntons cette idée de Gérard Genette soutient que un sens s'établie dans les relations dites transtextuelles, manifestes ou cachées, entre plusieurs textes reliées entre elles par des liens de parenté ou de filiation et reliées à un modèle commun, dont elles dérivent toutes; *Palimpsestes. La Littérature au second degré*, Paris, Seuil, coll. « Essais », 1982, p.11.

## 3.2. Méthodologies identifiées

La conception d'une ontologie est une tâche difficile nécessitant la mise en place de procédés élaborés. Étant donné la masse croissante de documents sur le Web d'une part et les avancées technologiques dans le domaine du traitement automatique des langues naturelles de l'autre, les recherches actuelles s'orientent davantage vers les procédés automatiques de génération d'ontologies à partir des textes<sup>31</sup>. Toutes ces méthodes se composent de plusieurs étapes, mais elles consistent en premier lieu à construire un corpus représentatif pour le domaine modélisé. Des outils de traitement linguistique sont ensuite mis en œuvre et les concepts identifiés ainsi que les relations entre eux sont soumis à l'évaluation par des experts du domaine concerné. La dernière étape est celle de la formalisation des concepts et du réseau sémantique précédemment obtenu et sa traduction vers le langage OWL.

L'enjeu de la construction d'une ontologie à partir des textes consiste à exploiter au mieux les éléments sémantiques fournis par ceux-ci, tout en s'affranchissant de l'empreinte linguistique et discursive d'une situation de communication donnée<sup>32</sup>. Pour ce faire, on emploie notamment les techniques élaborées par la linguistique du corpus visant à extraire à partir des documents textuels des termes représentant des concepts ainsi que les relations de l'hyponymie et de la subsumption qui les unissent. L'approche contemporaine ne se contente pas de la recherche des cooccurrences de certaines lexies, mais prend un compte la complexité d'une langue naturelle qui ne se résume pas à une collection des termes agencés par des règles syntaxiques. Grâce aux logiciels d'ingénierie linguistique de plus en plus sophistiqués, il est désormais possible d'analyser les marqueurs du discours et l'occurrence de divers signes sémiotiques, tels que les caractères gras ou les lettres majuscules.

Le survol des méthodologies actuellement employées a également permis d'identifier la méthode d'extraction des concepts à partir des schémas de documents XML<sup>33</sup>. Ce langage de balisage intègre les notions de description sémantique et, par le phénomène d'imbrication des balises, des liens existant entre les unités textuelles<sup>34</sup>. En prenant en compte la structure explicite des textes balisés en XML, cette méthode permet alors d'identifier les concepts ainsi que les relations sémantiques

---

31. Thibault Mondary, *Construction d'ontologies à partir de textes. Rapport de l'analyse de concepts formels*, thèse de doctorat, Université Paris 13, 2011, p. 3.

32. *Ibidem*, p. 2.

33. Citons comme exemple l'application OntoBuilder permettant l'extraction des ontologies à partir de fichiers XML. Cette application est disponible à l'adresse suivant : <http://ontobuilder.bitbucket.org/>, consulté le avril 2014.

34. Mouna Kamel et Nathalie Aussenac-Gilles, *Construction automatique d'ontologies à partir de spécifications de bases de données*, Publication du Laboratoire IRIT, Université Paul Sabatier de Toulouse, [http://geonto.lri.fr/Publications\\_fichiers/Papiers/KamelAssenacGilles\\_IC2009\\_10.pdf](http://geonto.lri.fr/Publications_fichiers/Papiers/KamelAssenacGilles_IC2009_10.pdf), consulté le 10 avril 2014.

entre ces concepts<sup>35</sup>. Cette approche nous semble d'autant plus légitime que l'élaboration des normes XML pour la documentation juridique prend de plus en plus d'envergure en fournissant ainsi un riche corpus documentaire<sup>36</sup>. Il est donc à noter qu'en exploitant la sémantique véhiculée par les documents textuels et la structure hiérarchique des balises XML, on s'éloigne considérablement d'une modalisation s'appuyant sur une conception syllogistique du droit.

### 3.3. Projet en cours

Face à l'importante inflation législative européenne, les travaux sur le web sémantique prennent de plus en plus d'envergure<sup>37</sup>. Comme preuve, les travaux menés dans le cadre du projet *Estrella*<sup>38</sup> ayant pour objectif de développer le format *Legal Knowledge Interchange Format* (LKIF). Ce dernier vise à adopter un profil de métadonnées commun pour la description des actifs sémantiques élaborés et utilisés par les gouvernements de l'Union européenne. Cette norme commune pour la représentation de l'information législative fournira une méthode pour la structuration, la publication et le référencement des documents législatifs. Le LKIF promet de rendre les requêtes véritablement sémantiques, à la différence d'une approche basée sur la cooccurrence des mots dans les textes. De plus, l'utilisation d'un cadre commun facilitera également l'analyse et l'interprétation des données législatives par les agents logiciels favorisant ainsi l'interopérabilité entre les différents systèmes d'information, particulièrement importante pour les organismes législatifs souhaitant communiquer en réseau sans ambiguïté ni perte.

## CONCLUSION

Le Web d'aujourd'hui est le fruit des efforts constants de la communauté participant à la définition de ses standards. Né de ces travaux concertés, le mouvement du Web sémantique a désormais atteint le niveau de maturité scientifique et les formats de représentation de l'information qu'il propose (RDF et OWL) sont suffisamment stables pour soutenir le développement de plusieurs applications. Il est donc légitime d'espérer que l'utilisation de ces standards améliorera les méthodes de stockage, de recherche et de restitution de l'information juridique. L'adoption généralisée des

---

35. Mouna Kamel et Nathalie Aussenac-Gilles, *Ontology Learning by Analyzing XML Document Structure and Content*. "Proceedings of the International Conference on Knowledge Engineering and Ontology Development", Funchal - Madeira, Portugal, October 6-8, 2009.

36. Citons comme exemple le développement du model XML pour les sources de l'information juridique européenne, CEN Metalex, <http://www.metalex.eu/>, consulté le 13 avril 2014.

37. Marc van Opijnen, *The European Legal Semantic Web: Completed Building Blocks and Future Work*, dans: European Legal Access Conference, Paris, 21-23 November 2012, p. 2.

38. Disponible à l'adresse suivant : [http://www.estrellaproject.org/?page\\_id=3](http://www.estrellaproject.org/?page_id=3), consulté le avril 2014.

normes et des standards du Web sémantique favorisera l'interopérabilité entre les systèmes de gestion documentaire et contribuera à ce que l'information juridique soit plus structurée et plus facilement accessible aux juristes professionnels et aux citoyens.

Les travaux sur la conceptualisation des ontologies juridiques entraînent également, par le regroupement sémantique et la clarification qu'ils opèrent, une réflexion sur l'amélioration possible des corpus législatifs et réglementaires pour que « les paroles des lois réveillent chez tous les hommes les mêmes idées »<sup>39</sup>. L'analyse linguistique de ce corpus documentaire peut mettre en évidence les incohérences ainsi que les modifications à apporter<sup>40</sup> afin de favoriser la démocratisation des règles juridiques en rendant l'information juridique plus simple et cohérente<sup>41</sup>.

---

39. Charles Louis de Secondat baron de Montesquieu, *De l'esprit des lois*, Paris, Gallimard, 1995, Livre XXIX, p. 533.

40. André Valenté, *Types and roles of legal ontologies*, dans : V. R. Benjamins, cité dans la note no 10.

41. Cette réflexion est d'ailleurs à la base de l'initiative du gouvernement britannique *Good Law* <https://www.gov.uk/good-law#relationship-with-other-initiatives>, consulté le 14 avril 2014.

## BIBLIOGRAPHIE

Monographies, thèses et ouvrages collectifs

V. Richard Benjamins *et al.*, *Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*, Lecture Notes in Artificial Intelligence 3369, Berlin Heidelberg, Springer, 2005.

Tim Berners-Lee, Mark Fischetti, *Weaving the Web. The original design and ultimate destiny of the World Wide Web, by its inventor*, San Francisco, Harper, 1999.

Alexander Boer *et al.*, *Using Ontologies for Comparing and Harmonizing Legislation*, dans : “Proceedings of the 9th International Conference on Artificial Intelligence and Law”, ACM Press, New York, 2003.

Mouna Kamel et Nathalie Aussenac-Gilles, *Construction automatique d'ontologies à partir de spécifications de bases de données*, Publication du Laboratoire IRIT, Université Paul Sabatier de Toulouse, [http://geonto.lri.fr/Publications\\_fichiers/Papiers/KamelAssenacGilles\\_IC2009\\_10.pdf](http://geonto.lri.fr/Publications_fichiers/Papiers/KamelAssenacGilles_IC2009_10.pdf), consulté le 10 avril 2014.

Mouna Kamel et Nathalie Aussenac-Gilles, *Ontology Learning by Analyzing XML Document Structure and Content*, dans : “Proceedings of the International Conference on Knowledge Engineering and Ontology Development”, Funchal - Madeira, Portugal, October 6-8, 2009.

Laurens Mommers, *Application of a knowledge-based ontology of the legal domain in collaborative workspaces*, dans: “Proceedings of the 9th International Conference on Artificial Intelligence and Law”, ACM Press, New York, 2003.

John McClure, *Legal-RDF Vocabularies, Requirements & Design Rationale*, dans : “Proceedings of the V Legislative XML Workshop”, European Press Academic Publishing (EPAP), 2007.

Thibault Mondary, *Construction d'ontologies à partir de textes. L'apport de l'analyse de concepts formels*, thèse de doctorat, Université Paris 13, 2011.

Marc van Opijnen, *European Case Law Identifier: Indispensable Asset for Legal Information Retrieval*, dans : Maria Angela Biasiotti et Sebastiano Faro, *From Information to Knowledge – Online access to legal information: methodologies, trends and perspectives*, IOS Press, 2011.

Marc van Opijnen, *The European Legal Semantic Web: Completed Building Blocks and Future Work*, dans: European Legal Access Conference, Paris, 21-23 November 2012.

Barry Smith et Christopher Welty, *Ontology: Towards a New Synthesis* dans : “Proceedings of the international conference on Formal Ontology in Information Systems”, New York, 2001.

Patrick Séguéla et Nathalie Aussenac-Gilles, *Extraction de relations sémantiques entre termes et enrichissement de modèles conceptuel*, dans : *Ingénierie des Connaissances*, Paris, École Polytechnique, 1999.

David Siegel, *The Power of the Semantic Web to Transform Your Business*, New York, Penguin Group, 2009.

#### Articles de revue

Kimberly Claffy *et al.*, “The (un)Economic Internet?”, *IEEE Intelligent Systems*, May-June, 2007.

Council of the European Union, “Council Conclusions Inviting the Introduction of the European Case Law Identifier (ECLI) and a Minimum Set of Uniform Metadata for Case Law”, *Official journal of the European union* (C/127) 1, Apr. 29, 2011, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:127:0001:0007:EN:PDF>, consulté le 10 avril 2014.

Thomas R. Gruber, “Towards Principles for the Design of Ontologies Used for Knowledge Sharing”, *International Journal Human-Computer Studies*, Vol. 43, Issues 5-6, November 1995.

John McCarthy, “Circumscription – A Form of Non-Monotonic Reasoning”, *Artificial Intelligence*, 1980, 5: 13.

François Rastier, « Sémantique du web vs semantic web ? », *Texte!Textes & Cultures*, Volume XIII - n°3, 2008, <http://www.revue-texto.net/index.php?id=1729>, consulté le 12 avril 2014.

#### Ouvrages généraux

Aristote, *La Métaphysique. Livre Gamma*, 1003, 21-32, traduction de Gaston Colle, Louvain, Édition de l'institut supérieur de philosophie, 1931, p. 3.

Rudolf Carnap, *The Logical Structure of the World*, Londres, Routledge & Kegan Paul, 1967.

Gérard Genette, *Palimpsestes. La Littérature au second degré*, Paris, Seuil, coll. « Essais », 1982.

Gottfried Wilhelm Leibniz, *Philosophical essays*, Indianapolis, Hackett Publishing, 1989.

Roger Verneaux, *Les sources cartésiennes et kantienne de l'idéalisme français*, Paris, Beauchesne, 1936.

Charles Louis de Secondat, Baron de Montesquieu, *De l'esprit des lois*, Paris, Gallimard, 1995.

Ludwig Wittgenstein, *Tractatus logico-philosophicus*, Gallimard, Paris, 1993.

Eugen Wüster, *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Vienne, Springer, 1979.

## Législation

*Loi sur la protection du consommateur*, RLRQ c. P-40.1.

## Ouvrage de référence

Alain Rey (dir.), *Dictionnaire historique de la langue française*, Paris : Le Robert, 2012.

## Autres publications

Alexander Boer *et al.*, *Specification of the Legal Knowledge Interchange Format*, Deliverable 1.1, 2007 <http://www.estrellaproject.org/doc/D1.1-LKIF-Specification.pdf>, consulté le 12 avril 2014.

*Déclaration de Montréal sur l'accès libre au droit*, <http://www.fatlm.org/declaration>, consulté le 20 avril 2014.

## Sites Internet

Cabinet Office of the Government of the United Kingdom, *Good law principles, current challenges and what the Office of the Parliamentary Counsel are doing to promote good law*, <https://www.gov.uk/good-law#relationship-with-other-initiatives>, consulté le 14 avril 2014.

Digimind, *Découvrir et exploiter le web invisible pour la veille stratégique*, 2008, <http://www.esi.ac.ma/Dossiers/20140126070100.pdf>, consulté le 2014-04-10.

Estrella consortium, *The LKIF Core Ontology* [http://www.estrellaproject.org/?page\\_id=3](http://www.estrellaproject.org/?page_id=3), consulté le avril 2014.

*LegalRDF*, <http://www.hypergrove.com/legalrdf.org/>, consulté le 2 avril 2014.

Tom Scott, "Case study: Use of Semantic Web Technologies on the BBC Web Sites", *BBC Internet Blog*, [http://www.bbc.co.uk/blogs/legacy/bbcinternet/2010/02/case\\_study\\_use\\_of\\_semantic\\_web.html](http://www.bbc.co.uk/blogs/legacy/bbcinternet/2010/02/case_study_use_of_semantic_web.html), consulté le 10 avril 2014.

Word Wide Web Consortium (W3C), *Semantic Web Activity*, <http://www.w3.org/2001/sw/>, consulté le 10 avril 2014.

Word Wide Web Consortium (W3C), *Semantic Web Use Cases and Case Studies*, <http://www.w3.org/2001/sw/sweo/public/UseCases/Elsevier/>, consulté le 8 avril 2014.