

Perspectivas sobre el uso de la Web Semántica en el tratamiento de información y documentación legislativa

Francisco Javier García Marco

Universidad de Zaragoza (España)

Resumen

Se analizan las perspectivas que ofrece la Web Semántica para el tratamiento de la información y documentación legislativa. Se revisa, en primer lugar, la evolución de la Web desde su proyecto inicial hasta los desarrollos recientes de carácter semántico y social. Seguidamente, se analizan las posibilidades que ofrece para abordar la informatización distribuida del ciclo completo de la información legislativa: desde su discusión, consenso y autorización hasta su publicación, aplicación, impacto, análisis y revisión. Se consideran las posibilidades de convergencia de la Web Semántica y la Web Social, y sus implicaciones para la Web legislativa. Se termina con un análisis estratégico de los retos y oportunidades que afecta a su implementación.

Palabras clave: Legislación. Información y documentación. Web Semántica. Web social. Web 2.0. Perspectivas.

Abstract

The perspectives offered by the Semantic Web project for the treatment of legislative information and documentation are analysed. The original concept and current evolutions of the World Wide Web project is reviewed, attending to the Semantic Web and Social Web developments. From this basis, their implications for the progressive and distributed automation of legislation-based activities are analysed, with a special attention to the convergence of the semantic and social technologies. Finally, a strategic analysis of the challenges and opportunities affecting current and future projects is offered.

Keywords: Legislation. Information and documentation. Semantic Web. Social Web. Web 2.0. Perspectives.

1. La relevancia de un ámbito de futuro

La Web Semántica es uno de los dos nuevos horizontes de la World Wide Web, el conjunto de tecnologías que, sobre la Internet, ha revolucionado la forma en que

publicamos, almacenamos, recuperamos e intercambiamos la información. El otro es, sin duda, la Web Social o Web 2.0. Las dos son evoluciones lógicas de la Web concebida en 1989. La primera —la Web Semántica— supone más y mejor computación distribuida entre los ordenadores que forman la Red; la segunda, más y mejor comunicación entre las personas que usan la Red. Detrás de ambas, la satisfacción del usuario, que gana en potencia de actuación, autonomía, seguridad, relación social, eficacia y eficiencia, aunque, ciertamente, no se deben obviar, por otro lado, los muchos peligros a conjurar, relacionados con la calidad y autenticidad de la información, su preservación y sus implicaciones para la privacidad.

Las expectativas que ha despertado la confluencia de la Internet y la información legislativa son enormes también para todos los implicados en ofrecer servicios de información legislativa a una comunidad de usuarios en expansión.

Por un lado, los profesionales de la información y la documentación legislativa (vid. Maciá, Gonzalo y Martínez-Cañavate, 2002, para una revisión) no pueden —ni quieren— permanecer ajenos a esta revolución en la infraestructura informacional de nuestra sociedad. Antes al contrario, son conscientes de las enormes posibilidades que se ofrecen a una actividad —la legislativa— que se basa sobre todo en la información.

Por otra parte, la información legislativa, en cuanto que es el trasunto y la referencia de la vertebración de las sociedades, tiene un potencial enorme para articular también gran cantidad de información administrativa disponible, cada vez más, en la Red, gracias al movimiento de aplicación de las tecnologías de la información y la comunicación en las administraciones y los poderes públicos que conocemos como *gobierno electrónico*, *administración electrónica* o simplemente, *eGov*, su acrónimo en inglés.

Finalmente, como veremos más adelante, las características de la información legislativa —compleja y formalizada, inherentemente social— ofrecen un ámbito de investigación y desarrollo de gran interés para tecnólogos y equipos interdisciplinares.

En este artículo realizaremos un estudio de la evolución de la World Wide Web y de sus perspectivas futuras y procuraremos entretener ese panorama con las características de la información legislativa para así intentar obtener una visión más clara de por dónde están yendo y pueden ir las cosas en el futuro próximo. Se trata, pues, de un trabajo de análisis y prospectiva.

2. La evolución de la World Wide Web

2.1. La revolución telemática

Es un tópico señalar que el momento actual se caracteriza por la revolución de la información. Este uso de la palabra *revolución* tiene que ver con la exportación

realizada de un concepto procedente del campo de la política —donde denota un cambio repentino, frecuentemente violento, de régimen político— al mundo de la economía, en el que se utiliza para representar la mutación radical de las infraestructuras y estructuras de la producción. En la sociología marxista, que postula la íntima conexión entre los planos ideológico, político, social y económico —con una determinación que va de abajo arriba para los ortodoxos y que se retroalimenta hasta cierto punto para los heterodoxos—, la revolución tiene un sentido sistémico.

En cualquier caso, la primera revolución industrial fue la de la máquina, cuando, a través de la aplicación de la energía fósil —móvil y continua— a ingenios mecánicos se posibilitó la mecanización de trabajo físico, con objetos. La revolución actual, sin embargo, se caracteriza por la automatización de la información, entendida como la gestión automatizada de las relaciones que interesan a agentes inteligentes para dirigir la acción en el mundo físico.

Ambas revoluciones tienen que ver con la liberación del trabajo: del físico en la primera; del comunicacional y cognitivo en la segunda. Comunicacional en la medida en que los ordenadores realizan numerosas transacciones y recodificaciones de información de forma automática; cognitivo porque los ordenadores son capaces de realizar automáticamente labores de cálculo, de computación. Todo ello conlleva la liberación de trabajo humano —que se puede aplicar a otros aspectos, cada vez más relacionados con la atención a otros seres humanos, la creación e innovación y el ocio— y afecta profundamente a las relaciones sociales, porque las potencia y amplía, por un lado, y las trastorna, por el otro, haciendo a los agentes sociales menos dependientes de sus relaciones tradicionales, sean de carácter laboral o personal.

Más allá del momento actual, pero presente ya en muchas pequeñas y grandes aplicaciones, la inteligencia artificial —lógica o “genética”— va abriendo la puerta a la automatización de la decisión, a través de modelos que toman decisiones mediante procedimientos tanto deductivos como inductivos. El futuro que se vislumbra a largo plazo es una presencia generalizada de autómatas capaces de decidir y relacionarse a partir de modelos complejos de representación y adquisición del conocimiento, que actuarán siempre que las personas no quieran o no puedan tomar esa responsabilidad directamente en sus propias manos y los sujetos de poder lo permitan.

2.2. La información “semántica”

El problema es que la información humana no es algo simple y plano, por así decirlo, como una memoria de ordenador. No es un espacio homogéneo de índole lógica, compuesto de estados de activación o desactivación.

Por el contrario, la información humana —expresión de nuestra capacidad de adaptación y modelación del medio ambiente físico, social y espiritual— es algo

complejo que se despliega verticalmente, desde la expresión simbólica de los mensajes hasta su intencionalidad y marcos de interpretación —escenarios, relaciones sociales, experiencias personales—, pasando por esos pasos intermedios a los que nos referimos con los conceptos de semántica y gramática, en la medida en que atendemos, respectivamente, bien a los significados de los signos, bien a su capacidad de articularse en nuevos signos más complejos. Aún por encima, surgen procesos idiosincrásicos, como la creación de sentido, que escapan a la automatización en cuanto que son intrínsecos al ser-en-el-mundo del sujeto, pero que, ciertamente, se pueden simular con tecnologías como los algoritmos genéticos y la vida artificial.

Sin ir tan lejos, la Web Semántica es un esfuerzo de crear un marco de representación y organización de la información contenida en mensajes y compartida en la Internet —ese enorme repositorio mundial al que toda la información disponible se ve abocada— a partir de sus relaciones profundas en el plano del significado. De ahí viene el uso del adjetivo *semántico*. Y ese énfasis en lo semántico nos catapulta sin remedio a la comprensión del mundo del conocimiento, esas estructuras que están subyacentes a los mensajes —en nuestro caso a los documentos, y más específicamente en su forma de recursos web— y que han sido lentamente iluminadas por filósofos y lingüistas.

En este marco, la Web Semántica legislativa aparece como algo más que gestionar mejor la información legislativa y hacerla recuperable y navegable de forma más eficaz y eficiente. Más allá de este objetivo pragmático irrenunciable, la Web Semántica en el campo que nos ocupa se puede entender como un esfuerzo sistemático para organizar el conocimiento de la realidad social y de las técnicas jurídicas que se utilizan para estructurarla dentro de marcos políticos, tal y como aparecen de forma subyacente en la propia legislación, de manera que ese enorme conjunto de documentos sean capaces de responder de forma cada vez más sencilla a preguntas cada vez más inteligentes y sofisticadas.

Precisamente gracias a las técnicas de representación que ofrece la Web Semántica es posible intentar organizar y representar el conocimiento legislativo como una superestructura intermedia que ponga en contacto las necesidades de los agentes que componen el cuerpo social —con sus diferentes perspectivas y roles— y el mundo en permanente expansión de la documentación legislativa. Es desde esta perspectiva como cobra sentido hablar de representación y organización del conocimiento legislativo y no solo de información legislativa.

En fin, el concepto de sociedad del conocimiento se ha convertido en un eslogan para denominar a nuestra época, en particular dentro el contexto europeo. Desde el punto de vista que nos ocupa, eso significa colaborar en la tarea de conectar el conocimiento almacenado en los documentos legislativos con las necesidades de conocimiento de todos los agentes que requieren de esas fuentes, y hacerlo en el contexto de la red Internet, con las posibilidades que ofrece de

colaboración e integración en un espacio público, permanente y global de comunicación y conocimiento.

2.3. La Web 1.0: una tecnología de comunicación documental

A principios de la década del 2000, Tim Berners-Lee y su equipo propusieron su nuevo horizonte para la World Wide Web: la Web Semántica (Berners-Lee, Hendler y Lassila, 2001). En resumidas cuentas, se trataba de convertir la Web en un entorno auténticamente interoperable de computación distribuida, y, en definitiva, de ir más allá de lo que ha sido el uso principal de la Web y la razón fundamental de su éxito: el intercambio de información entre personas a través de páginas web, esto es, de documentos. Sin embargo, la Web fue concebida en un entorno de alta computación y contenía la semilla de objetivos más ambiciosos. En realidad, el problema de la Web Semántica y los conceptos que luego darán lugar a RDF están ya presentes en el proyecto inicial Tim Berners-Lee (1989-1990):

An intriguing possibility, given a large hypertext database with typed links, is that it allows some degree of automatic analysis. It is possible to search, for example, for anomalies such as undocumented software or divisions which contain no people. It is possible to generate lists of people or devices for other purposes, such as mailing lists of people to be informed of changes. It is also possible to look at the topology of an organisation or a project, and draw conclusions about how it should be managed, and how it could evolve. This is particularly useful when the database becomes very large, and groups of projects, for example, so interwoven as to make it difficult to see the wood for the trees.

In a complex place like CERN, it's not always obvious how to divide people into groups. Imagine making a large three-dimensional model, with people represented by little spheres, and strings between people who have something in common at work.

Now imagine picking up the structure and shaking it, until you make some sense of the tangle: perhaps, you see tightly knit groups in some places, and in some places weak areas of communication spanned by only a few people. Perhaps a linked information system will allow us to see the real structure of the organisation in which we work.

No obstante, el éxito de la World Wide Web se debió más bien no tanto a sus promesas como al hecho de que ofrecía una solución sencilla y abierta al problema de la publicación electrónica. Efectivamente, la Web ha constituido desde sus primeros momentos, básicamente, una tecnología de comunicación documental. Está orientada fundamentalmente a transmitir documentos de forma no solo eficaz, sino, sobre todo, muy eficiente, gracias a la codificación digital electrónica y óptica. Frente a la documentación en papel —eficaz y relativamente barata—, la comunicación mediante documentación digital en red aporta cinco ventajas que la superan de forma decisiva: velocidad —inmediatez, de hecho—, economía de transporte y almacenamiento, una disponibilidad universal y un medio compatible con cualquier tipo de documento —sonido, texto, imágenes fijas, vídeo, interactivos y sus combinaciones— y de carácter convergente —proporciona una plataforma para

medios tanto privados como públicos, personales o institucionales, y tiene un carácter multicanal—. De hecho, su carácter multicanal y universal permite reunir en un espacio virtual medios de información hasta ahora dispersos en el espacio como editoriales, medios de comunicación, instituciones culturales y científicas, etcétera. Como el papel, y a diferencia de otras soluciones informáticas, la World Wide Web constituye una tecnología asequible, abierta, distribuida y libre que todo el mundo puede utilizar en cuanto dispone de un ordenador de cualquier tipo conectado a la Red.

El inmediato éxito de la tecnología se enfrentó al principio a la ausencia de una amplia estructura de servicio como la que existe en torno al papel. Casi desde un primer momento se detectaron los problemas que planteaba el nuevo medio frente a servicios de información más exigentes. Por un lado, faltaban catálogos que permitieran una búsqueda adecuada en ese nuevo espacio público de información. Por el otro, la información podía desaparecer —y aparecer— sin ningún control. Finalmente, la autenticidad de la información en estas redes abiertas quedaba gravemente en entredicho, pues la información digital en ficheros no protegidos es fácilmente manipulable.

Pero, dadas las enormes ventajas de la tecnología, las soluciones vinieron también muy rápidamente. En primer lugar, la Red empezó a dotarse de catálogos y buscadores en un entorno competitivo y abierto que ha llevado a la aparición de gigantes de los servicios de intermediación de la información como Google o Yahoo! En segundo lugar, comenzó la regulación del depósito legal de los materiales digitales y aparecieron instituciones privadas dedicadas al almacenamiento de la información digital. En tercer lugar, para las necesidades de recuperación y control más exigentes, se importó el modelo de representación de la información de las bibliotecas y bases de datos basados en estándares de metadatos. Finalmente, a través de la firma electrónica y las huellas digitales se ha restaurado en gran medida la confianza en la autenticidad de los documentos electrónicos, posibilitando que sirvan para muchas funciones que todavía les estaban vedadas, en particular en el ámbito que nos interesa, el legal. Falta, eso sí, todavía por resolver el problema de la disociación del estrecho nexo que une la firma al individuo y que se está intentando suplir mediante sistemas de captura de biodatos en el momento de la firma digital.

Gracias a todos estos avances —y con alguna salvedad, como se ha visto—, hoy por hoy nada impide alcanzar en la World Wide Web los estándares de autenticidad, conservación y recuperación que han sido tradicionales en los archivos y bibliotecas, siempre que, lógicamente, se dediquen el trabajo y los recursos necesarios.

Como consecuencia, la World Wide Web se ha convertido en una plataforma integral de publicación, almacenamiento y recuperación de la información, y, progresivamente, de autenticación, organización y conservación de la información (García Marco, 1997). Se trata, sin duda, de uno de los grandes hitos de nuestra cul-

tura, que está dando lugar a la aparición de una inmensa biblioteca digital universal multimedia distribuida. Y una parte fundamental de esa biblioteca debe ser lógicamente la información jurídica, que constituye una de las columnas vertebrales de las organizaciones sociales.

2.4. Los límites de la Web 1.0

Ahora bien, ese éxito, como vislumbraba Tim Berners-Lee, puede ser mayor. Para comprender por qué, es necesario analizar los límites de la representación del conocimiento en la web tradicional, cuya gran aportación es una implementación computacional del concepto de hipertexto.

2.4.1. El centro conceptual de la Web 1.0: el hipertexto

Efectivamente, la World Wide Web es el resultado de la confluencia de la Internet —red telemática de carácter público y universal y arquitectura distribuida— con el proyecto de automatización de la recuperación y acceso a la información que cristalizó a partir de la II Guerra Mundial en el concepto de hipertexto. Fue la inmediatez de la transferencia documental en las redes telemáticas —fruto de un factor cuantitativo, la velocidad de transmisión de la información— la que hizo posible el salto cualitativo que suponía el hipertexto. Siguiendo la visión de Vannevar Bush (1945) y de Theodor Nelson —simplificadamente, y con su crítica (1)—, los documentos digitales en la World Wide Web se ligan mediante enlaces que permiten recuperar un nuevo documento pulsando sobre una fracción del documento de partida que ha sido convenientemente marcado —el enlace o *link*—.

De esta manera, el hipertexto permite generar un espacio “virtual” de información —y de experiencia— en la medida en que documentos, medios, emisores y receptores anteriormente dispersos y alejados en el espacio quedan ahora reunidos en un isoespacio en el que todos se encuentran en teoría a la misma distancia ideal, en concreto a un clic del ratón o de otro dispositivo homólogo. En ese espacio, si todo va bien, desaparece el factor tiempo, y los documentos quedan contiguos, al menos desde el punto de vista del trabajo necesario para consultarlos. Lo mismo ocurre con los procesos de comunicación mediados por documentos en la Internet, como el correo electrónico, el chat o la videoconferencia (2).

Desde un punto de vista lógico y comunicacional —si se quiere, filosófico—, el concepto subyacente al hipertexto no es innovador; es una función de la comunicación oral —la referencia a otros mensajes—, y aflora desde siempre en ámbitos como la academia, la literatura, el arte, los textos sagrados o la ciencia. Se trata de la referencia intertextual o cita “bibliográfica”.

Lo que hace el hipertexto es la traslación de la metáfora del botón —máximo impacto con mínimo trabajo— al mundo de la referencia intertextual: si se quiere, la mecanización de la referencia intertextual.

2.4.2. *La referencia intertextual, centro de la World Wide Web*

La referencia intertextual es una necesidad ineludible porque el documento es un mensaje y no está diseñado para transmitir el conocimiento de forma integral, sino para efectuar un proceso de comunicación, un proceso de armonizar conocimientos en un contexto de saberes compartidos entre emisor y receptor. Cuando el conocimiento previo y los fines son muy semejantes en el emisor y el receptor, es fácil realizar este proceso. Sin embargo, cuando o bien los fines, o bien el conocimiento previo son diferentes —como ocurre frecuentemente en la lectura con fines de creación, estudio e investigación—, el texto interesa solo de forma muchas veces fraccionaria, y, sobre todo, necesita ser completado por otros textos relacionados. Desde el punto de vista del lector de un documento —que puede estar vivencialmente alejado del emisor y ligado a él por una relación muy limitada—, este se convierte más bien en un recurso para sus necesidades, en función de las cuales procesará solo partes limitadas de él, aquellas que le permiten alcanzar sus fines; aunque, claro está, puede ser “infectado” por las ideas que contiene, con lo que se generaría un proceso de comunicación bidireccional.

Así, pues, el universo de los mensajes reconstruye, de alguna manera, el universo del conocimiento, o, más precisamente, lo “representa”. Sin embargo, lo hace de una manera, por una parte, redundante, y por otra, incompleta. Esto es así porque los mensajes representan solo una fracción del conocimiento disponible, concretamente aquella que un emisor consideró interesante transmitir a un receptor en su debido momento. Por ello, solo pueden ser interpretados desde ese conocimiento socialmente compartido. En definitiva, los mensajes transfieren estados de conocimiento en función de una situación de comunicación específica y están sujetos a distorsiones en los planos pragmático, semántico y simbólico. Por otro lado, el mundo del conocimiento se extiende a lo largo de la línea del tiempo, generando la interesantísima dinámica que el filósofo Karl Popper (1979) discute en su teoría de los tres mundos.

Como bien señalaron estos expertos, el saber humano es multidimensional y posee una integración que supera las posibilidades de la comunicación, que se realiza mediante secuencias de símbolos. La comunicación, de hecho, presupone ese mundo de saber que permite su decodificación, y que constituye, por así decirlo, la cuarta dimensión de la documentación. De alguna manera, el hipertexto permite reconstruir esas relaciones de sentido de forma transversal a las expresadas en los discursos específicos de los documentos, que, por otra parte, están llenos de referencias implícitas y explícitas a otros documentos.

Obviamente, la World Wide Web realiza una reconstrucción muy limitada de las relaciones entre las fracciones de saber —por así decirlo— que contienen los documentos. La acerba crítica del creador del concepto de hipertexto (Nelson; vid. nota 1) va precisamente en esa dirección. Los enlaces de la Web aportan muy poca

información y solo permiten establecer nexos 1 a 1. Además requieren mucho trabajo, razón por la que la mayor parte de los autores se resiste a poner más enlaces de los que son estrictamente imprescindibles para que el documento sea localizado o sirva de puerta a otros documentos absolutamente necesarios desde el punto de vista de su creador o su editor. La representación del conocimiento por encima de los mensajes —las páginas web— queda limitada a una nube difusa y muy incompleta de enlaces entre páginas, que resulta útil sobre todo para estimar la popularidad de las páginas, más que para relacionarlas rigurosamente mediante vínculos conceptuales. En lo referente a los mensajes propiamente dichos, la representación se realiza en el nivel de la codificación simbólica del plano del significativo y de las relaciones entre textos.

Así pues, el hipertexto es revolucionario no tanto por sus bases —lógicas, epistemológicas o tecnológicas— cuanto por sus efectos: la generación de un espacio virtual en el que todos los documentos se convierten en contiguos unos de otros, situándose a la misma distancia “psicológica” y con un mismo esfuerzo de obtención. Todos los documentos pueden ser, recíprocamente, enlazados con el mismo trabajo. Esa simplicidad y esa potencia ofrecen, por sí solas, unas posibilidades emergentes cuyos resultados están a la vista de todos.

2.4.3. El procesamiento de datos

La otra gran limitación de la World Wide Web como sistema informático es que no permite distinguir los datos que están contenidos en los documentos, que aparecen, desde el punto de vista del programador —no, obviamente, desde el del lector— como información no estructurada. Eso impide su procesamiento automático, y convierte a toda esa información en una especie de cultivo extensivo por contraposición al cultivo intensivo y preciso que se realiza por medio de tecnologías como los gestores de bases de datos. El marcado de datos que se realiza mediante HTML en la World Wide Web se orienta sobre todo a la transmisión de los documentos y su correcta presentación, y queda muy por detrás de las posibilidades ofrecidas por los lenguajes de marcado, diseñados como medio de interconexión entre la gestión de datos y la ofimática documental.

Las limitaciones de la World Wide Web tanto en su implementación del hipertexto como en su capacidad de gestión de datos son precisamente el trampolín del que parte el proyecto de la Web Semántica.

2.5. El futuro de la World Wide Web: Web Semántica y Web Social

Sin embargo, la evolución actual de la World Wide Web solo parcialmente sigue esta lógica; el otro gran ramal de desarrollo de la Web tiene que ver más bien con su potencial como medio de comunicación. Como consecuencia, en la actualidad parece que el futuro de la World Wide Web se va dibujando en dos capas, que no son excluyentes, sino más bien complementarias.

La primera tiene que ver con la computación propiamente dicha y con su horizonte mítico: la inteligencia artificial. El movimiento que desarrolla este enfoque es precisamente el de la Web Semántica. Su centro es la automatización de los procesos de información en sus diferentes niveles: simbólico, semántico y pragmático. Su objetivo —inherente a la informática en cuanto que disciplina de la “información automática” (3)— es aislar fracciones de información —datos— que sean procesables dentro de un algoritmo para obtener un nuevo dato.

Es importante señalar que, aunque el discurso teórico de la informática se elabora en torno al concepto de información, lo cierto es que esta disciplina se ocupa del procesamiento de datos. En realidad nadie sabe muy bien qué es la información, más allá de algo que reduce la incertidumbre en el marco de un esquema de acción que conduce a determinados resultados —una definición heredada de la teoría de la probabilidad, devenida en teoría de la decisión—. Apenas se consideran otros aspectos de la información “humana”, particularmente la creación de marcos de acción o la definición de objetivos. En consecuencia, la informática solo hace, lógicamente, abordables los aspectos computables de la información, aunque interesen, por supuesto, los demás. Al final, la información se reduce a lo manejable: la relación entre datos, los procesos que conducen de un conjunto de datos a otro y el esfuerzo de representar aspectos relevantes o interesantes mediante conjuntos de datos y algoritmos de transformación.

No es extraño que, por tanto, la Web Semántica se plantee sobre todo como un esquema para identificar y procesar automáticamente los datos que contiene la información. Sin embargo, aporta la construcción de niveles de sucesiva abstracción y control: el documento estructurado (XML), la definición del esquema documental (DTD, XML Schema), el nivel de relaciones entre metadatos (RDF), la definición de esquemas de metadatos (RDF Schema) y, finalmente, la modelización de dominios de conocimientos (OWL). Esa representación en niveles de sucesiva integración y control se semeja a la estructura real de los sistemas cognitivos humanos y permite una representación del conocimiento que va más allá del simple procesamiento de la información.

La segunda gran tendencia de evolución de la World Wide Web tiene que ver con las posibilidades que ofrece la Web como medio de comunicación, y en particular con sus usos emergentes tanto para el trabajo como para el ocio colaborativo. Se trata de la llamada *Web Social* o *Web 2.0* —en el popular acrónimo creado por O’Reilly (2005).

Una cuestión importante es que la automatización del procesamiento de información está ligada fundamentalmente a procesos de eficiencia y economías de escala, mientras que el movimiento de colaboración en la Web es, ante todo, un proceso de generación de oportunidades de interacción y de mercados y espacios sociales. No es extraño, por tanto, que el interés comercial y político esté ahora si-

tuado sobre todo en la Web 2.0. Es posible pronosticar que, conforme el potencial de la World Wide Web como espacio de interacción social virtual se vaya alcanzando, el énfasis volverá a los procesos de automatización para los que la Web Semántica ha sido diseñada.

La Web Semántica ofrece un camino flexible sin sacrificar el objetivo último de cualquier método de representación del conocimiento: la sistematicidad —tanto en el sentido de totalidad como en los de exhaustividad, granularidad, coherencia y consistencia—. Uno de los objetivos de la Web Semántica es interrelacionar profundamente la información de la World Wide Web en un todo cada vez más coherente. En realidad, una mayor integración y aprovechamiento de las sinergias de la información disponible en la Web es un objetivo compartido de las dos evoluciones de la World Wide Web, esto es, tanto de la Web Semántica como de la Web Social.

3. La World Wide Web legislativa

Una vez analizados la filosofía de la World Wide Web y sus desarrollos actuales es relativamente sencillo introducirnos en el campo específico de su aplicación a la información legislativa. En principio, se trata de un problema de aplicación de unas tecnologías de propósito general a las peculiaridades y objetivos de un campo específico de la actividad humana. Concretamente, la Web Semántica legislativa no sería otra cosa que la aplicación de las posibilidades de la Web Semántica a las características específicas de la información legislativa, y, de manera muy especial, a las necesidades que subyacen tras ellas. Sin embargo, no hay que descartar que los desarrollos con tecnologías semánticas en el campo de la información legislativa no puedan aportar perspectivas nuevas al propio proyecto de la Web Semántica.

Desde este punto de vista, se puede definir la Web Semántica legislativa como un sistema de información, comunicación y documentación de carácter abierto, distribuido e interoperable en la Internet, dotado de las características avanzadas que proporciona la Web Semántica, que evoluciona y se desarrolla con objeto de resolver de forma integrada y sinérgica las necesidades del ciclo completo de la actividad legislativa y de todos sus agentes. A partir de una definición como esta conviene precisar tanto los aspectos tecnológicos como los propiamente legislativos.

Desde el punto de vista tecnológico cabe precisar algunas de las principales características que aporta la Web Semántica: fundamentalmente, la anotación del contenido mediante lenguajes de marcado (XML) (Arellano, Nogales y Martín, 2003), la definición de estándares de descripción de documentos (XML Schema), el establecimiento de interconexiones entre etiquetas y datos etiquetados (RDF), la definición de esquemas de descripción de recursos y de relaciones entre datos (RDF Schema) y la construcción de ontologías legales distribuidas (OWL) para facilitar la navegación, recuperación y explotación de la información legal (figura 1).

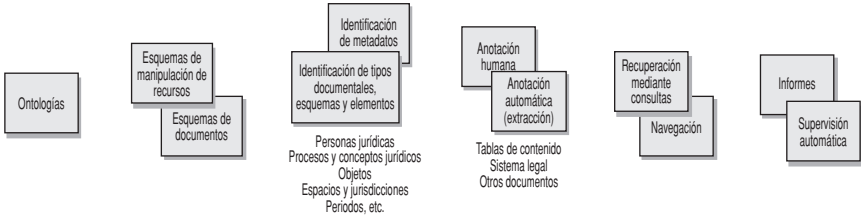


Figura 1. Las tecnologías semánticas y su aplicación a la información legislativa.

Desde el punto de vista propiamente legislativo, las características y necesidades que plantea la información legislativa se pueden visualizar con más precisión atendiendo tanto al ciclo de su producción y uso (figura 2) como a los agentes que intervienen en él (figura 3). Paralelamente, es necesario tener en cuenta también la naturaleza evolutiva y diacrónica de la actividad legislativa y su íntima conexión con la geografía política y con los otros factores que intervienen en esta, notablemente los aspectos lingüísticos.

Desde una óptica global, parece claro que una Web Semántica legislativa debe atender a todos los actores y procesos, facilitando su interacción e integración en el entorno distribuido de Internet para facilitar la obtención de sinergias de muy diverso tipo. Al reutilizar, enlazar, representar y retroalimentar la información generada en cada fase por cada agente surgirán procesos nuevos, más eficaces y eficientes.

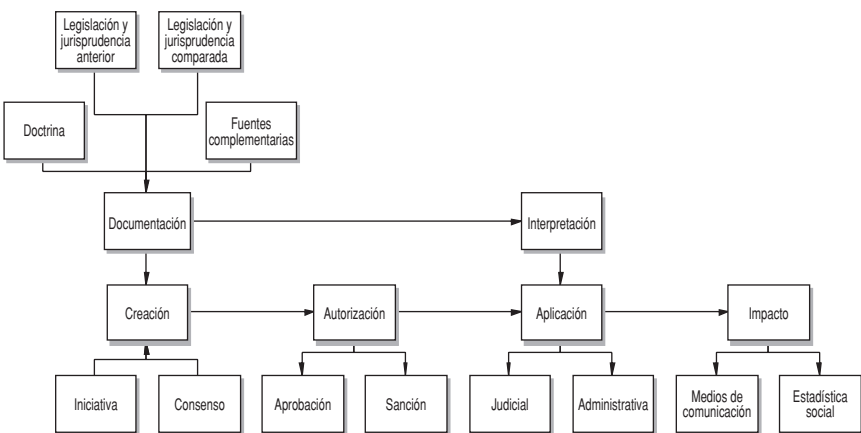


Figura 2. El ciclo legislativo.

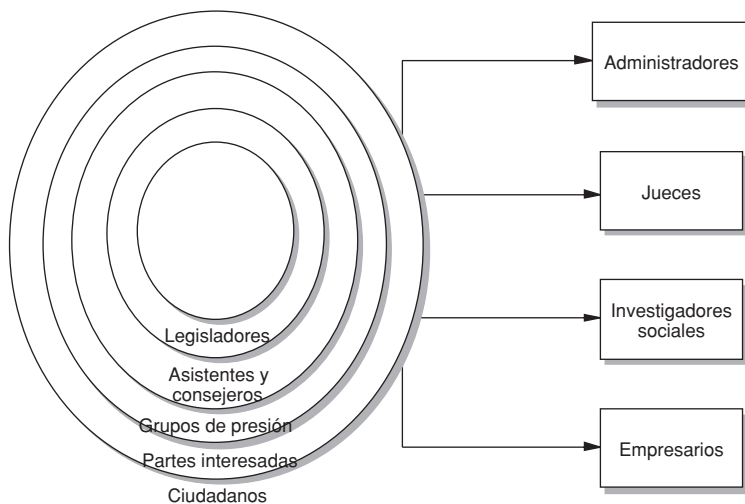


Figura 3. Los agentes del proceso legislativo.

3.1. La Web 1.0 como tecnología de publicación jurídica

La creación, el debate, la aprobación y la publicación de legislación son procesos que realizan un uso intensivo de documentación: sus productos finales, sub-productos y productos intermedios son documentos. Recíprocamente, los procesos políticos y administrativos necesarios para su producción se controlan igualmente mediante documentos y quedan reflejados en documentos.

Ciertamente, legislación y documento son dos realidades íntimamente unidas desde el comienzo de la civilización. El derecho legislativo se ha beneficiado históricamente de los desarrollos en el mundo de la documentación, permitiendo su codificación, conservación y publicación de forma más adecuada. Recíprocamente, las actividades documentales —especialmente las relacionadas con la preservación y el acceso, que fundamentan las profesiones de bibliotecario, archivero y documentalista— han encontrado su justificación en el valor que mantienen en la larga duración los documentos legales, científicos o literarios. La documentación legislativa ha constituido uno de los pilares de la documentación como actividad profesional.

Por todo ello, la información y la documentación legislativa se han beneficiado enormemente de la World Wide Web, de manera que tanto los organismos productores —parlamentos y cámaras legislativas— como las partes intervinientes —partidos, sindicatos y otras organizaciones no gubernamentales— usan de forma generalizada servicios de información World Wide Web públicos y privados (intranets y extranets) para apoyar su reproducción, producción y difusión. Se

publican en ellos un conjunto de documentos relacionados con la producción y difusión de la legislación —propuestas, críticas, valoraciones, borradores, leyes, sentencias, presupuestos, reglamentos, noticias, discusiones, etcétera— que deben ser ligados en una estructura interrelacionada para sacar adelante las acciones relacionadas con esta misma legislación: documentación, discusión, consensos, aprobación, publicación, crítica, aplicación y clarificación judicial y administrativa, estudio del impacto, modificación y derogación.

En el ámbito de la difusión documental, la World Wide Web ha facilitado también de forma decisiva el acceso a las bases de datos jurídicas (Alvite, 2003, 2004a y 2004b) y a los servicios de alerta de legislación que permiten trabajar de forma más eficaz y eficiente a los profesionales del derecho en sus diferentes actividades. La World Wide Web se ha convertido en el soporte básico de la intermediación en el campo de la información legislativa, y se han abierto muchas posibilidades de mejora.

Como resultado, la situación actual consiste en numerosos repositorios de información legislativa y paralegislativa ligada a los gobiernos, asambleas y partes interesadas, así como a los medios de comunicación y a los editores y distribuidores comerciales de información jurídica.

El gran reto es poner en valor y relación todos esos repositorios de manera que se integren en un escenario de acelerada globalización en los diferentes niveles en que esa integración es necesaria: dentro de los propios ámbitos jurisdiccionales, entre las diferentes coadministraciones y gobiernos, con los profesionales jurídicos, con las empresas y con los ciudadanos, individuales u organizados en grupos, asociaciones y fundaciones. De esta manera, la legislación cumplirá los fines previstos y los requisitos necesarios tanto en el marco estricto de los sistemas jurídicos como en el de la propia vida social en toda su complejidad y niveles. Estas mejoras en el plano de la eficacia irán de la mano de enormes ahorros de costos en el procesamiento y difusión de la información legislativa.

3.2. La Web Semántica y el universo de la información legislativa

El mundo del derecho, el de la Administración pública y, en particular, el de la legislación constituyen campos de investigación y aplicación apasionantes para la Web Semántica.

Por una parte, componen un campo de información altamente formalizado en sus procesos y formatos documentales y de datos, susceptibles de ser trasladados a los protocolos y formatos de representación de la Web Semántica. Es necesario, claro está, un esfuerzo añadido de normalización para asegurar la interoperabilidad, pero existe un sustrato de formalización muy avanzado. Por otra, se trata de información muy compleja, con muchas relaciones, formatos, usos y actores; justo el tipo de problemática para la que se pensó la Web Semántica: un conjunto de tec-

nologías orientadas precisamente a la resolución de problemas complejos en los que intervienen muchas partes, y que requieren una computación a la vez compleja y potente, por un lado, y abierta y distribuida por otro.

Además, la Web Semántica permite bajar un escalón en el nivel de automatización, desde el de la publicación, distribución y acceso de la información legislativa —el nivel de la comunicación— al del control de los datos que contiene y de su consistencia interna. Un problema que plantea la legislación y que no hace sino crecer con el aumento de las regulaciones y de la globalización es su coherencia y consistencia interna y externa. Al atender al nivel de la información etiquetada, esto es, al nivel del dato, la Web Semántica permite enlazar la legislación relevante, en un primer paso, y establecer sobre esa información etiquetada mecanismos de análisis y navegación mediante ontologías en un segundo paso.

Sin ir a cuestiones tan complejas, la codificación a nivel de datos que permite la Web Semántica ofrece ventajas prácticas inmediatas, ya que se trata de una codificación estructurada, que facilita extraordinariamente todo un conjunto de aplicaciones.

A continuación vamos a examinar ese contínuum de tecnologías que van desde la anotación estructurada al desarrollo de ontologías, y que constituyen las puntas de lanza de los desarrollos semánticos en el campo de la información legislativa. Las cuatro primeras tienen que ver sobre todo con la familia XML —anotación colaborativa, preservación, reutilización, distribución multicanal y navegación mediante tablas de contenido y enlaces—, y las demás, con RDF y OWL —navegación y recuperación conceptual, abordaje de la barrera lingüística, minería de datos y generación de reportes, y análisis de coherencia y consistencia—. Dado el enorme trabajo que supone, el desarrollo de procedimientos de detección y marcado automáticos es uno de los aspectos críticos implicados en casi todos los esfuerzos.

3.2.1. La anotación colaborativa como característica básica

Sin duda, una de las posibilidades naturales de la familia de estándares XML es la anotación de textos legales, que constituye la base de todas las tecnologías semánticas, como anotación de elementos o de contenido añadido.

Nada impide, lógicamente, dar un paso más y convertir esa anotación en un proceso colaborativo con todas las garantías de control necesarias, identificando autores y datando los cambios adecuadamente. La anotación interesa —con fines en algunas ocasiones diversos y en muchas coincidentes— a agentes muy diferentes: legisladores, jueces, académicos, documentalistas, periodistas y ciudadanos interesados, entre otros sectores.

Un objetivo deseable sería ir clasificando los actores y los tipos de anotaciones de cara a la construcción de estándares en RDF Schemas que, posteriormente, pudieran ligarse a ontologías.

3.2.2. Almacenamiento permanente y preservación a largo plazo

Una de las implicaciones más importantes de los estándares de la Web Semántica resulta de uno de sus objetivos manifiestos: la estabilidad. Al definir formatos estables para codificar caracteres, imágenes, estructuras documentales, metadatos y relaciones conceptuales, la Web Semántica sienta las bases para la preservación a largo plazo de la información.

Hay mucho trabajo, sin embargo, por hacer. Especialmente, es necesario desarrollar estándares para almacenar los documentos “máster” de forma independiente de las anotaciones, sin perjuicio de su accesibilidad. Con el tiempo, uno de los efectos más importantes de la Web Semántica debería ser la radical reducción en la redundancia de la información, de manera que se dirija inequívocamente a los documentos originales y sus copias autorizadas y que el estatus de cada documento legal de la Internet quede claro en cuanto a su autenticidad y autoridad.

3.2.3. Reutilización de contenidos y distribución multicanal

También las empresas que ofrecen servicios añadidos sobre la información y documentación legislativa utilizan la World Wide Web de forma generalizada para la labor de difusión de la información. Para estas empresas, la utilización del enfoque de Web Semántica permite, sobre todo, tener la información estructurada y organizada en un repositorio de cara a su reutilización en subsiguientes productos de información —manuales, compilaciones, boletines de difusión selectiva, bases de datos en Internet o en diferentes soportes, etcétera— que se ajusten a las necesidades y hábitos de conjuntos específicos de usuarios (clientes objetivo), y cuyas características pueden ser expresadas en lenguajes de formateo (4).

Una vez formalizados los sistemas de representación para el almacenamiento y la difusión, el proceso global puede beneficiarse enormemente de la gestión automatizada de flujos de trabajo y, al final de cada flujo, de la gestión automatizada de productos editoriales.

3.2.4. Navegación intradocumental e interdocumental

Puesto que la referencia intradocumental e interdocumental es un aspecto fundamental de la documentación legislativa —y en general de la jurídica—, el hipertexto se ha convertido en una herramienta imprescindible de los servicios de información legislativa, al dotar de agilidad al manejo de las referencias y ahorrar tiempo a los usuarios.

Las referencias documentales en los textos legislativos se pueden clasificar en varios tipos. En primer lugar, las que vertebran jerárquica y verticalmente el texto en función de las características del tipo documental concreto, y que permiten navegarlo mediante tablas de contenido y sumarios, fáciles de extraer mediante XML-XLS. En segundo lugar, las que conectan los diferentes documentos legislativos

entre sí, reconstruyendo el enorme entramado del sistema jurídico en la sincronía y la diacronía. En tercer lugar, la conexión de los textos legislativos con todos los documentos que los utilizan, los desarrollan, los interpretan y los valoran. La extracción automática de estos enlaces es una de las líneas de investigación más prometedoras para alcanzar fácilmente una masa crítica de enlaces (Martínez, Fuente y Derniame, 2003; Martínez y Fuente, 2007). La precisión y categorización de los enlaces gracias a estándares más avanzados como XPath o XLink es, como muestran estos autores, un aspecto clave de los desarrollos en este campo.

3.2.5. Descripción de contenidos, recuperación y navegación conceptual

La utilización de RDF y de esquemas de descripción de recursos como Dublin Core ofrece la capa intermedia para conectar los documentos estructurados con las ontologías que permiten organizar la información en respuesta a preguntas —recuperación de información— mediante menús navegables y mapas conceptuales, que es la forma en que la Internet —más allá del hipertexto plano— se hace inteligible para los usuarios.

Como se ha señalado, una de las aplicaciones más útiles de las ontologías es precisamente la de facilitar la anotación semántica: identificación de lugares, organizaciones, personas, documentos, procesos y objetos mencionados en los documentos legales y en todo el cuerpo de textos relacionados con ellos. Así, por ejemplo, un sistema basado en ontologías puede asociar fácilmente varias denominaciones de la misma entidad —El Govern, Gobierno de Cataluña, Govern of Catalunya, etcétera—. O puede ampliar la búsqueda a partir de las relaciones contenidas en la ontología: por ejemplo, desglosar los países europeos en una búsqueda sobre legislación europea no comunitaria. Pero su mayor potencial reside en que conectar ontologías entre sí es relativamente fácil, pues es posible la recuperación en repositorios de información diferentes. Como señala Davies (Benjamins et ál., 2008, p. 78), “Ontologies can help to address this issue by providing a uniform access layer to heterogeneous data sources. Linking multiple structured, semistructured, and unstructured information sources using a consistent vocabulary makes it easier to build applications that pull data together from across the enterprise. It also facilitates the introduction of new systems and databases”.

Otra gran potencialidad de las ontologías es, precisamente, ofrecer vistas personalizadas y agregadas de esos recursos a partir de la descripción sistemática de los conceptos de recuperación y de la interconexión entre ontologías diversas —pues toda ontología se configura en respuesta a necesidades específicas, que varían según grupos de usuarios y contextos—. Se trata de agrupar diferentes clasificaciones en una misma ontología de manera que la información se busque y se muestre mediante las categorías y la terminología más naturales para el usuario.

Es importante señalar, sin embargo, que las ontologías requieren, recíprocamente, la descripción de los recursos y servicios. El problema es la cantidad ingente de trabajo humano, tanto en análisis como en coordinación y supervisión, que resulta necesaria para llevarla a cabo. Por ello, al igual que en otros campos de aplicación, una de las líneas en desarrollo más atendidas es la asignación automática o asistida de metadatos.

3.2.6. *Puentes multilingües*

La legislación está ligada al ámbito jurisdiccional en el que se produce y debe estar expresada en las lenguas oficiales legalmente previstas. Sin embargo, la creciente globalización, el respeto a las lenguas minoritarias dentro de los Estados y la consolidación de bloques regionales con idiomas distintos —Unión Europea, Mercosur, etcétera— exigen la consideración transversal de esos entornos jurídicos formulados en diferentes lenguas.

Por ello, el mundo de la información legislativa es uno de los que se puede beneficiar en mayor medida de la interconexión de terminologías en diferentes idiomas en el ámbito de la Internet, por medio de tecnologías como RDF u OWL.

3.2.7. *Obtención de datos y reportes*

Una aplicación fundamental de la Web Semántica será la obtención de datos sobre los procesos legislativos, su valoración, sus efectos y su impacto. Las ontologías legales servirían en este campo como instrumentos para recuperar información etiquetada que, después, sería resumida mediante técnicas estadísticas por los denominados *robots semánticos* para facilitar los procesos de opinión y decisión política.

Uno de los campos de aplicación que se vislumbra más interesante es la conexión con los sistemas de estadística social. Otro, que obviamente interesa enormemente a los políticos, es el análisis del impacto mediático de la actividad legislativa. Por una parte, los políticos son especialmente sensibles a los problemas que aparecen en los medios, los cuales, en muchas ocasiones, condicionan el noventa por ciento de sus iniciativas políticas. Por otra, los medios dedican una cobertura importante en noticias y opinión a la actividad legislativa.

Otro campo de aplicación de gran importancia es, ciertamente, el mapeo de legislaciones relacionadas. Se trata de un aspecto que está avanzando en el contexto europeo de la mano de grupos de investigación como el Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) de Florencia, conocido también por sus trabajos de marcado automático en XML.

3.2.8. *Vigilancia de la coherencia del sistema legal*

Algunos expertos están incluso especulando con la posibilidad de utilizar un marcado muy fino de la legislación con el objeto de procesar la coherencia de los

textos legales y su conformidad con otras normas de carácter superior o promulgadas anteriormente. Se trata de desarrollar sistemas formales para asistir a los legisladores y a las agencias que están obligadas a informar o supervisar las nuevas leyes y normas en un entorno de crecimiento continuo de las regulaciones, exacerbado por los acuerdos internacionales que conlleva la globalización. Este tipo de aplicaciones enlaza directamente con los sistemas expertos y con el programa de la inteligencia artificial.

3.2.6. Las ontologías legislativas

Las ontologías legislativas ocupan un lugar central en el desarrollo de todos los campos de aplicación que se han analizado someramente en los apartados anteriores, y merecen, posiblemente, una atención algo más especial.

El campo de las ontologías nació y se desarrolló dentro de la inteligencia artificial. Su objetivo era desarrollar formalismos para describir mediante el lenguaje de la lógica el sistema conceptual de un determinado dominio de la actividad humana de tal manera que dicha descripción pueda soportar mecanismos de inferencia automatizados (Gruber, 1993). Este esfuerzo se incorporó al proyecto de Web Semántica y se ha concretado en estándares sólidos como OWL (World Wide Web Consortium, 1994-2004 y 2004), que han despertado grandes expectativas, tal como demuestra la literatura creciente sobre el tema (García Marco, 2007). Los campos de experimentación se multiplican y el campo de la información jurídica es uno de los que ha recibido un interés especial, tanto por su carácter transversal a las diferentes realidades sociales, políticas y económicas como por las propias características de ese material (García Marco, 2006).

En línea con lo sucedido en otros campos, las ontologías legales se desarrollaron inicialmente dentro del programa de la inteligencia artificial como una tecnología de representación del contenido legal —Frame Based Ontology (FBO), LRI Core, Legal Knowledge Interchange Format (LKIF)—. Sin embargo, en el actual panorama de su incorporación a Internet, ha crecido el interés en su aplicación a dos aspectos más específicos íntimamente relacionados: la recuperación de la información basada en vocabularios controlados y la organización de sitios web mediante taxonomías. En este sentido, muchos de los proyectos de los últimos años se han centrado en la expresión en OWL de tesauros o taxonomías legales de alcance general o especializado —Lexical Ontologies for legal Information Serving (LOIS), Legal Taxonomy Syllabus, Metalex, Ontoweb— (Schweighofer, 2008).

La mayor parte de los grupos de investigación interactúan en torno a las International Conferences on Artificial Intelligence and Law (ICAAIL) y a diversos seminarios, entre los que se cuentan el International Seminar on Law and the Semantic Web (Barcelona, 2003), que fue el punto de partida del conocido libro de Benjamins et ál. (2005), el International Workshop on Legal Ontologies and Web

Based Legal Information Management (Edimburgo, 2003), el Workshop on Regulatory Ontologies (Chipre, 2004) y el II Workshop on Legal Ontologies and Artificial Intelligence Techniques (Universidad de Stanford, 2007). Entre los centros de investigación internacionales destaca en Europa el Laboratory for Applied Ontology del CNR, ahora adscrito al Institute of Cognitive Sciences and Technology (ISTC). En España existen proyectos en curso, entre otros, en el Departamento de Inteligencia Artificial de la Universidad Politécnica de Madrid (Gómez-Pérez, 2006; Costilla et ál., 2005; Corcho et ál., 2005) y en el Instituto de Derecho y Tecnología de la Universidad Autónoma de Barcelona, cuyos investigadores se encuentran activos en los foros mencionados.

3.3. La convergencia de la Web Semántica y la Web Social en el ámbito de la información legislativa

3.3.1. La Web Social y la información legislativa

En la línea del concepto de “redes vivas” acuñado por Aires Rover (2007), la construcción social de la Internet no se debe limitar tan solo al desarrollo tecnológico, sino que por el contrario las redes telemáticas han de servir para alimentar el desarrollo positivo de las redes sociales que subyacen tras ellas y les dan sentido.

La Web 2.0 significa, sobre todo, que la Web como espacio virtual de comunicación no es solo el resultado de arquitecturas de servicios y contenidos inventados por individuos y asumidos o avalados por grupos o instituciones destacados, sino que, además, es el resultado de la contribución de muchísimas personas, grupos e instituciones dotados de la posibilidad y las herramientas para aportar su colaboración. Es también un proceso de “apropiación” de la tecnología por los ciudadanos, incluyendo a los desfavorecidos.

¿Qué significa esto para la información legislativa? Pues, fundamentalmente, las enormes posibilidades que ofrecería un anillo de servicios de colaboración en torno a los servicios legislativos, dedicados tanto a los agentes cualificados como a los ciudadanos en su conjunto. Las implicaciones para la mejora de la participación ciudadana en la construcción y reforma del ordenamiento jurídico son obvias, y se ayudaría a resolver problemas de etiquetado y mapeo de relaciones que la automatización no está todavía en condiciones de resolver plenamente.

3.3.2. Los dos escenarios futuros de la WWW y las dos visiones del derecho

Es interesante constatar cómo la perspectiva sistemática del conocimiento jurídico que pretendería un enfoque desde la Web Semántica y la perspectiva emergente y colaborativa que ofrece el enfoque de la Web Social se relacionan directamente con las dos dinámicas de construcción del derecho que han sido expresadas por diferentes filósofos del derecho (Aguiló, 2005). La primera de ellas, a veces denominada *legalista*, es la centrada en las normas y el rigor del método y el ra-

zonamiento deductivo. La segunda, o *constitucionalista*, es de carácter más dinámico, orientada a los resultados sociales y centrada en los procesos de conflicto y consenso, y, en general, en las relaciones de poder que subyacen al establecimiento de normas jurídicas.

Parece lógico pensar que, en principio, la perspectiva formalista ha de ser mejor servida por las tecnologías de representación del conocimiento y, por ende, por las de la Web Semántica, más orientadas al procesamiento automático e inferencial, mientras que la perspectiva *constitucionalista* lo será por las tecnologías de comunicación social (Web 2.0), orientadas precisamente a la potenciación de las redes sociales humanas y a su articulación.

Igualmente, el sentido común sugiere que ambas perspectivas —y por ende ambos grupos de tecnologías— son complementarias antes que contrapuestas, aunque en ocasiones puedan entrar en conflicto.

Así, por un lado, la Web Semántica legal, a pesar de su énfasis en los procesos automatizados y automatizables, debe estar al servicio de redes sociales humanas que constituyen el sentido último del derecho. Recíprocamente, la Web 2.0 requiere la racionalización de los procesos de información y asistencia automatizada en la recuperación, procesamiento e integración de la información generada, o, de otra manera, corre el peligro de caer en el continuo descubrimiento de mediterráneos, las oleadas de cambio basadas en la moda, la desestructuración y el presentismo.

Sin embargo, a pesar de su indudable atractivo intelectual, hay que moderar la validez de esta dicotomía, puesto que la Web Semántica también es *social*, si bien requiere una capacidad técnica mayor. Su objetivo es también la integración de servicios generados por diferentes agentes, para construir realidades emergentes. Desde esta perspectiva, ambas son redes *sociales*, si bien la Web Semántica es la Red de los técnicos, y la Web 2.0, la de los usuarios. Seguramente, este enfoque es más ajustado a la realidad.

3.3.3. Los problemas: la convergencia como un reto

Aunque la integración entre ambos mundos ha sido defendida recientemente por diversos autores, tampoco es posible desde una perspectiva rigurosa obviar los problemas que plantea.

Ciertamente, la Web 2.0 reproduce el proceso de información no estructural característico de la Web 1.0. En este sentido, la Web 2.0 está enormemente alejada del proyecto de la Web Semántica, y, en la práctica, ambos constituyen dos proyectos independientes, aunque puedan ser complementarios.

Un ejemplo claro de las diferencias de enfoque entre ambas concepciones de la web son las folcsonomías, las etiquetas descriptivas del contenido que los usuarios

asignan libremente a los documentos. Es cierto que existe la posibilidad de conectarlas con los esquemas más estructurados, y que hasta cierto punto se pueden utilizar tecnologías de detección automática para ello. Sin embargo, ello requeriría, durante mucho tiempo, una enorme cantidad de trabajo humano dedicado no solo al análisis, sino a la supervisión y al mantenimiento de los sistemas de interconexión. Y no hay que olvidar que los sistemas de Web Social basan su éxito precisamente en el aporte limitado por parte de los administradores y masivo por parte de los usuarios, lo cual es el enfoque contrario al de la Web Semántica.

En conclusión, la convergencia de los proyectos de la Web Semántica y de la Social ofrece una de las líneas de investigación más interesantes y provocadoras. En este sentido, existen ya proyectos tan interesantes como los *semantic wikis*, en los que los usuarios colaboran en anotar y enlazar páginas partiendo de una ontología estática —aportada por los administradores— o abierta —construida dinámicamente por los usuarios—. Ciertamente, la colaboración en la creación de enlaces semánticos y en la construcción de ontologías es una de las grandes promesas para intentar reducir los costes de unas actividades que consumen recursos humanos abundantes, especializados y, por consiguiente, caros.

4. Conclusión

En las páginas anteriores hemos intentado ofrecer un panorama de la evolución de la World Wide Web y las oportunidades que ofrece para mejorar la creación, acceso y evaluación de la legislación, una actividad central de las sociedades humanas. Se ha visto cómo el germen de la Web Semántica latía ya en el proyecto inicial planteado por Tim Berners-Lee a los responsables del CERN. Después, se ha examinado cómo la explosión de la información y la comunicación que ha conllevado la Web ha generado procesos emergentes de carácter social que están colmando las expectativas iniciales de la Red como entorno de colaboración. Más computación gracias a la Web Semántica y más interacción humana gracias a la Web Social son las dos grandes tendencias de la World Wide Web actual. Ambas ofrecen posibilidades interesantísimas para la actividad legislativa y, en definitiva, para la difusión, el acceso, el uso y la evaluación de la legislación. Las sinergias entre ambos enfoques prometen dar alas a ambos proyectos y ayudarles a superar los escollos con los que se enfrentan. En este sentido, son complementarias.

Como colofón, vamos a intentar completar este análisis prospectivo de la aplicación de las tecnologías semánticas a la información legislativa con un análisis de carácter más operativo en el que se examinan los aspectos que favorecen y perjudican a su adopción y desarrollo práctico. Al final, a pesar de todas sus ventajas teóricas, la Web Semántica legislativa debe ser implementada en la práctica, y un posicionamiento estratégico sobre su viabilidad y las fuerzas positivas y negativas que entran en juego es una parte importante del análisis de sus perspectivas. Con

este fin, se presenta en la tabla I un análisis DAFO que intenta cubrir este objetivo, sintetizando las debilidades, amenazas, fortalezas y oportunidades que se han detectado a lo largo de este estudio.

Debilidades	Fortalezas
<p>Falta de conocimiento de la tecnología por parte de numerosos proveedores de servicios informáticos para las administraciones públicas.</p> <p>Falta de una comunidad suficientemente grande para alcanzar masa crítica.</p> <p>Complejidad de la información jurídica.</p> <p>Ausencia de servicios reales en explotación que muestren las ventajas a los agentes y al gran público.</p>	<p>Existencia de una comunidad científica interesada de carácter interdisciplinar.</p> <p>Existencia de organizaciones clientes con el volumen de datos, la visión estratégica y los recursos necesarios.</p> <p>Solidez epistemológica y tecnológica del proyecto de la Web Semántica y de sus estándares, que abarcan todas las áreas críticas de desarrollo.</p> <p>Carácter altamente formalizado de la documentación jurídica y de los procesos legislativos.</p> <p>Disponibilidad de una enorme cantidad de información estructurada y semiestructurada de carácter legislativo, y relacionada con esta.</p> <p>Capacidad de pacto y racionalidad de los agentes legales.</p>
<p>Escasa visualización de las tecnologías semánticas por parte del gran público y de la mayor parte de los agentes implicados.</p> <p>Mayor impacto de la computación social y mejor modelo de negocio.</p> <p>Resistencia a perder monopolios económicos en el uso de información debido a formatos que facilitan la reutilización de información.</p> <p>Miedo a perder el gradiente de poder que conlleva el control sobre el acceso a la información y su procesamiento.</p> <p>Coste alto de desarrollo y mantenimiento.</p> <p>Desconfianza ante los problemas de autenticidad y preservación de la información.</p>	<p>Carácter directamente motivador de la información legal, que genera derechos y obligaciones sujetos a sanciones y penas.</p> <p>Mejorar el acceso a la información legislativa incrementa la eficacia del principio de publicidad de la norma.</p> <p>Necesidad de mejorar la eficiencia en la utilización de la información legislativa y de reducir los costos de todos los agentes implicados.</p> <p>Experiencias constatadas de ahorro de costes en la reutilización de información.</p> <p>Urgencia de mejorar la conexión de sistemas legislativos distintos en un entorno de creciente globalización y regionalización.</p>
Amenazas	Oportunidades

Tabla I. Análisis DAFO de la aplicación de las tecnologías semánticas a la información legislativa.

Como se puede ver, se trata de un conjunto de fuerzas favorables y opositoras de muy diverso cariz e importancia, la mayor parte abordables de una manera u otra, pero que deben tenerse presentes de forma global.

En el platillo del haber, las aplicaciones semánticas para el procesamiento de la información legislativa ofrecen para todas las personas informadas posibilidades

reales de abordar problemas que hasta ahora han sido en gran medida intratables por la cantidad masiva de documentación que está en juego. Por otra parte, a nadie se le escapa que los corpus legales ocupan un lugar central en los sistemas sociales que regulan, por lo que constituyen un excelente punto de amplificación en el que demostrar y generalizar el uso de estas tecnologías. Se puede afirmar con casi total seguridad que las tecnologías semánticas diseñadas para este tipo de documentos resultarían automáticamente útiles en otros ámbitos como el político, el de la Administración pública, el judicial y el comercial.

Sin embargo, a pesar de las grandes promesas que ofrece, el desarrollo de tecnologías semánticas legislativas constituye una actividad cara, tanto en su inicio como en su mantenimiento constante, especialmente si se la compara con la otra gran corriente de la Web: la Web Social. Requieren además información muy estructurada, en un volumen suficiente, que pueda ser explotada por los sistemas de razonamiento lógicos inventados por los expertos en inteligencia artificial. Son precisamente los entes gubernamentales y las grandes empresas que generan y explotan la información legislativa los que poseen el volumen de datos estructurados y semiestructurados, la visión estratégica, el grado suficiente de formalización de sus procedimientos y los recursos disponibles necesarios para acometer la tarea.

El esfuerzo debe dirigirse a trabajar con estos clientes y desarrollar servicios específicos que les permitan desarrollar sus actividades con costes más reducidos. Al final, estos ahorros fruto de la automatización de tareas y de la reingeniería de procesos, junto con las subvenciones y los beneficios fiscales de la actividad I + D + I son los factores que van a comprometer a los clientes. Por su parte, los grupos de investigación tendrán que cubrir las debilidades de un enfoque excesivamente pragmático. Deberán, por un lado, intentar mantener una visión estratégica de la integración de servicios que supere una focalización estrecha en aplicaciones aisladas; por otro, preservar un enfoque humanístico —inherente a las tecnologías de las que estamos hablando— de la necesidad de compartir información, redistribuir responsabilidades y cargas, y generar redes y sinergias con una amplia perspectiva social.

Es cierto que, en la medida en que la actividad legislativa está connaturalmente ligada a los procesos de control social y ejercicio del poder, las iniciativas para compartir información mediante procedimientos automáticos —en definitiva, delegados— se producirán previsiblemente con grandes fricciones y resistencias. Es un factor clave que debe ser gestionado de forma atenta y adecuada. Los beneficios que puede aportar una Web Semántica legislativa merecen el esfuerzo.

Notas

- (1) “The Xanadu® project did not ‘fail to invent HTML’. HTML is precisely what we were trying to PREVENT —ever-breaking links, links going outward only, quotes you can’t follow to their origins, no version management, no rights management” (Nelson, 1999).

- (2) Es cierto que, en estos dos últimos casos, el documento es tan efímero que pasa desapercibido, pero está presente tanto en la comunicación asíncrona como en la síncrona.
- (3) Véase una interesante presentación del término en el artículo de Michael Fourman (2003) en la *International Encyclopedia of Information and Library Science*.
- (4) Véase la experiencia en la reutilización de contenidos con un enfoque multicanal —difusión selectiva, bases de datos, libros, formación— de la empresa Lexnova (Escudero, 2007).

Referencias

- Aguiló Regla, Josep (2005). Introduction: legal informatics and the conceptions of the law. // Benjamins, V. Richard, et ál. (eds.) (2005). *Law and the Semantic Web: legal ontologies, methodologies, legal information retrieval, and applications*. LNAI. Londres: Springer, 2005. 18-24.
- Alvite Díez, M.^a Luisa (2003). Tendencias a la investigación sobre la recuperación de Información Jurídica. *Revista Española de Documentación Científica*. 26:2 (2003). <http://redc.revistas.csic.es/index.php/redc/article/view/137/191> (2008-04-12).
- Alvite Díez, M.^a Luisa (2004a). Evolución de las bases de datos jurídicas en España. // *Anales de Documentación*. 7 (2004) 7-27. <http://www.um.es/fccd/anales/ad07/ad0701.pdf> (2008-04-12).
- Alvite Díez, M.^a Luisa (2004b). Las herramientas terminológicas en los sistemas de información jurídica. // *Scire: Representación y Organización del Conocimiento*. 10:1 (en-jun. 2004) 77-90.
- Arellano Pardo, María del Carmen; Nogales Flores, J. Tomás; Martín Galán, Bonifacio (2003). La organización hipertextual del ordenamiento jurídico: posibilidades de XML y estándares relacionados. // *Revista general de información y documentación*. 13:2 (2003) 181-191.
- Benjamins, V. Richard; Casanovas, Pompeu; Breuker, Joost; Gangemi, Aldo (eds.) (2005). *Law and the Semantic Web: legal ontologies, methodologies, legal information retrieval, and applications*. LNAI. Berlín, Londres: Springer, 2005.
- Benjamins, Richard; Davies, John; Baeza-Yates, Ricardo; Mika, Meter; Zaragoza, Hugo; Greaves, Mark; Gómez-Pérez, José; Manuel Contreras, Jesús; Domingue, John; Fensel, Dieter (2008). Near-Term Prospects for Semantic Technologies. // *IEEE Intelligent Systems*. 23:1 (Jan.-Feb. 2008) 76-88.
- Berners-Lee, T. (1989-1990). *Information Management: a proposal*. Ginebra: CERN, March 1989 – May 1990. <http://www.w3.org/History/1989/proposal.html> (2008-04-12).
- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). The Semantic Web. // *Scientific American*. 284:5 (2001) 34-43.
- Bush, V. (1945). As we may think. // *Atlantic Monthly*. 176 (July 1945) 101-108.
- Corcho, O.; Fernández-López, M.; Gómez-Pérez, A.; et ál. (2005). Building legal ontologies with METHONTOLOGY and WebODE. // *Lecture Notes in Computer Science*. 3369 (2005) 142-157.
- Costilla, C.; Palacios, J. P.; Cremades, J.; et ál. (2005). e-government: a legislative ontology for the 'SIAP' parliamentary management system. // *Lecture Notes in Computer Science*. 3416 (2005) 134-146.
- Scire*. 15 : 1 (en-jun. 2009) 59-85. ISSN 1135-3716.

- Rover, Aires (2007). Intervención en el Seminario Hispano-Brasileño de la Red AECI “Gobierno Electrónico”. Florianópolis, 27 de noviembre de 2007.
- Schweighofer, Erich (2008). Legal Ontologies and XML. // XII Congreso Iberoamericano de Derecho e Informática (12-17 de mayo de 2008). Centro de Congresos y Exposiciones de IberCaja, Zaragoza. Presentación en Power Point.
- Schweighofer, E.; Liebwald, D. (2007). Advanced lexical ontologies and hybrid knowledge based systems: first steps to a dynamic legal electronic commentary. // *Artificial Intelligence and Law*. 15:2 (Jun. 2007) 103-115. <http://dx.doi.org/10.1007/s10506-007-9029-1> (2008-04-12).
- World Wide Web Consortium (1994-2004). Semantic Web Activity. Cambridge, Keio, París: W3C, 1994-2004. <http://www.w3.org/2001/sw/> (2008-04-12).
- World Wide Web Consortium (2004). OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004. Cambridge, Keio, París: World Wide Web Consortium, 2004. <http://www.w3.org/TR/2004/REC-owl-features-20040210/> (2008-04-12).

Recibido: 2008-05-08. Aceptado: 2008-09-14