

Tecnologías de la Web Semántica aplicadas al tratamiento de documentos jurídicos electrónicos

Héctor J. Ruidías^{1a}, Karina B. Eckert^{1b}, Juan M. Lezcano^{2c}, Carolina V. Rosas^{2d}

¹Departamento de Ingeniería y Ciencias de la Producción

²Departamento de Ciencias Sociales y Jurídicas

Universidad Gastón Dachary - Posadas, Misiones, Argentina

{^achandra149, ^bkarinaeck, ^cjuanmanuellezcano, ^dcarolinav.rosas}@gmail.com

RESUMEN

En el terreno jurídico, el crecimiento de Internet ha permitido que las normativas legales, digestos, fallos judiciales, sentencias y leyes estén a disposición a través de portales de acceso público o privado, los cuales cuentan con una forma de organización determinada y búsqueda por términos. La documentación jurídica supone aspectos intrínsecos del objeto judicial que posee relaciones de categorías propias del derecho y de las legislación existente que muchas veces no aparecen en forma explícita y que requieren por lo tanto sean identificadas en un proceso de “curación” del documento legal.

Se presenta así la necesidad de ofrecer una forma de tratamiento documental que ponga atención a la semántica de los textos. Una forma de abordar esta cuestión es a través de las tecnologías semánticas, en particular recurriendo al uso de *ontologías* que ofrecen una forma *explícita* de representación que permite aplicar procesos de *descubrimiento* de relaciones a partir de mecanismos de razonamiento automático. En este modelo de relaciones semánticas explícitas, los mecanismos de búsquedas pueden resultar favorecidos al permitir búsquedas sobre un vocabulario menos dependiente

de la simple ocurrencia terminológica, y sí más orientado a las estructuras de las categorías empleadas (o de los metadatos y sus relaciones).

Palabras clave: Gestión de documentos legales, Web Semántica, Ontologías, Razonamiento automático, Gestión del Conocimiento.

CONTEXTO

La presente investigación se encuentra en ejecución, acreditado en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) por resolución N° 07/A/17. El proyecto es del tipo interdisciplinario abarcando dos áreas disciplinarias principales: Informática (Tecnologías de la Información y Comunicación (TIC)) y Ciencias Jurídicas (Derecho).

1. INTRODUCCIÓN

El fenómeno de la inflación legislativa y la producción de documentos jurídicos hacen necesario desarrollar y promover plataformas y medios de comunicación e información que proporcionen oportunidades de acceder, compartir e intercambiar los recursos científicos, culturales, sociales y económicos que

están disponibles en base de datos vinculadas con las ciencias jurídicas [1].

Atento a que dicho fenómeno se da también en un contexto de crecimiento exponencial de la información producida globalmente por organizaciones, empresas e individuos, alentada sin duda por la dinámica propiciada por Internet, es precisamente que también sobre la misma coyuntura científico-tecnológica se están desarrollando el mayor número de avances tecnológicos que abordan dicha problemática desde dos aspectos fundamentales: uno relacionado al tratamiento de volúmenes masivos de datos y el otro relacionado a la reducción de la opacidad semántica de dichos datos.

Es por ello que en forma concomitante al desarrollo de la Web, tal como fuera planteada por Sir Tim Berners Lee, también ha ido tomando vigor la idea de la Web Semántica [2], donde la información no solamente resultara comprensible por seres humanos, sino también por máquinas.

El hecho de que la búsqueda de información jurídica sea deficiente, no resulta únicamente en un inconveniente de índole técnico, sino que atenta contra los mismos cimientos de la democracia a propiciar la falta de transparencia y celeridad necesaria en la resolución de casos judiciales que esconden un entramado de conflictos sociales [3], [4].

Las ontologías han sido propuestas como artefactos de representación que especifican un vocabulario relativo a cierto dominio en el contexto de los sistemas de información [5]. Dicho vocabulario define entidades, clases, propiedades, axiomas y reglas, además de las relaciones entre dichos componentes, con el objetivo de reducir la ambigüedad, los conflictos terminológicos y las discrepancias semánticas que se presentan en un área de conocimiento determinada. Esto conduce a una definición donde una

ontología es entendida como “una especificación formal y explícita de una conceptualización compartida” [6]. Se destaca en la misma dos ideas claves, la de conceptualización compartida y la de especificación formal. La primera refiere a un modelo abstracto, con sus elementos y la vinculación entre estos, explícitamente definido y además cumple con el requisito de ser consensuada y aceptada por una comunidad. El segundo en tanto aborda la cuestión del lenguaje y el vocabulario con que se representa, de tal manera que sea interpretable de igual manera tanto por personas como por computadoras, siendo por lo tanto una cualidad irrenunciable el que sea procesable por máquina.

El énfasis en el aspecto “consensuado” que propician las tecnologías de la Web Semántica va desde la estandarización y especificaciones, la arquitectura de capas de la W3C, hasta las definiciones de ontologías de alto nivel (o fundacionales) o específicas para distintos dominios, ya sean médicos, de organizaciones y personas o de índole legal. A modo de resumen se puede mencionar el conjunto de tecnologías de base sobre las que se construye la Web actual, que incluye una forma unívoca de identificar recursos (URI), una codificación estándar (UNICODE), y un lenguaje básico para estructurar documentos (XML+XMLSchema) y sobre el cuál se respaldan los distintos lenguajes de marcado, tales como XHTML y HTML. Sobre tales tecnologías se cimientan las tecnologías específicas de la web semántica, que incluyen descriptores de recursos y relaciones entre estos (RDF y RDFSchemas) y un lenguaje específico de ontologías (OWL), que aporta mayor expresividad al permitir describir propiedades y clases con restricciones de cardinalidad, exclusión (relaciones disjoint), y una mayor riqueza a la hora de

definir propiedades (relaciones). Naturalmente tales características imponen algunas restricciones en función de la expresividad deseada, lo cual permite diferenciar entre tres tipos básicos de familias de ontologías, OWL Lite, OWL DL y OWL FULL [7].

Las ontologías han demostrado su utilidad desde sus orígenes al ofrecer un marco conceptual útil para reducir la ambigüedad inherente al uso del lenguaje natural. Puesto que en muchos dominios del conocimiento humano se establecen definiciones en términos de construcciones verbales susceptibles a varias interpretaciones, principalmente aquellas que se vinculan a las ciencias sociales.

En el derecho, la interpretación es un eje fundamental sobre el que se articula el accionar aplicativo de la justicia, y el deterioro legislativo debido a la sobresaturación de leyes y normas, puede suponer un reto a la hora de conocer la vigencia real sobre cierta normativa. Todo esto pone sobre relieve la necesidad de establecer un proceso de ordenamiento jurídico [3] que culmina en la creación de un Digesto Jurídico. Tal proceso supone básicamente el escrutinio permanente sobre el crecimiento legislativo y el consecuente ordenamiento del mismo. Existen tres categorías fundamentales de intervención que se consideran y que definen la complejidad del proceso (consolidación), que incluye 1) un aspecto sistemático, es decir la forma en que se organizará la normativa, 2) otro lingüístico (morfológico, sintáctico y semántico) y 3) uno de interrelación normativa.

Estas tres categorías resultan de vital importancia, no únicamente para el ordenamiento jurídico, sino para la digitalización y su tratamiento a través de sistemas de gestión documental, que

permitan almacenar y acceder a dicha normativa.

La creación de un Digesto Jurídico por otra parte supone la conformación a una doctrina, un marco jurídico y normativo y también el entrecruzamiento entre las ramas clásicas del derecho y otras transversales tales como “equidad de género”, “biodiversidad”, “derecho a la salud” entre otras [4].

En el marco de un sistema de gestión documental enriquecido con tecnologías semánticas, un Digesto Jurídico se define a partir de vinculaciones semánticas (anotado semántico) entre el texto de la norma y una ontología de referencia. Para ello lo apropiado supone la reutilización de ontologías de alto nivel [8], [9] para objetos que no son propios del derecho, tales como aquellos que refieren a eventos temporales y ubicaciones espaciales, y para aquellos que pertenecen a categorías propias del Derecho, emplear ontologías de dominio tales como FOLAW o LRI-CORE [10], [11].

Un antecedente destacado que se puede mencionar es el e-COURT de la Unión Europea, un sistema para manejo de información legal, textual y multimedia, que permite el almacenamiento y recuperación de la misma, y que es tolerante a ambigüedades de índole lingüística gracias a las tecnologías semánticas que emplea [12]. Otro antecedente es el Digesto Jurídico basado en ontologías creado para la Universidad Nacional del Litoral [13].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La presente línea de investigación comprende aspectos propios del derecho, como el tratamiento y ordenamiento jurídico de documentos legales, y otros propios de las TIC, como ser todo lo

relativo a la Web Semántica y tecnologías relacionadas, así también como sus aspectos metodológicos (Ej. Metodologías para desarrollar ontologías: methontology). Por otra parte se emplea documentos legales reales para tener una mejor dimensión de la problemática, en base a los cuales se diseña y pone a prueba una infraestructura de software desarrollada a partir de distintas tecnologías de la web semántica y sistemas de base de datos no relacionales y/o documentales. Cabe mencionarse que la propuesta no se centra únicamente en el hipotético sistema de una respuesta respecto a qué documentos recuperar a partir de un término de búsqueda, sino también a un dato puntual solicitado (Ej. todos los fallos vinculados a “violencia de género” que fueron “apelados”), eventualmente en lenguaje natural, aunque podría formularse a través de una interfaz ad-hoc similar al lenguaje de consulta SPARQL.

El trabajo desarrollado sentará las bases para la transferencia de soluciones tecnológicas al sector judicial, partiendo desde un prototipo funcional que empleará datos aportados desde el Poder Judicial de la Provincia de Misiones mediante la firma de un convenio de transferencia de conocimientos entre la UGD y dicho organismo estatal. Asimismo desde el punto de vista académico, desde el trabajo llevado a cabo con tecnologías semánticas, se puede ver ya no solo un campo fértil para la aplicaciones de tales tecnologías, sino también la posibilidad de acceder a una fuente de escenarios de prueba para la validación de las propuestas desarrolladas tanto en el marco del presente proyecto como de futuros proyectos.

3. RESULTADOS Y OBJETIVOS

Objetivo General

Proponer una infraestructura orientada a la búsqueda de información jurídica en documentos electrónicos legales a partir de la integración de tecnologías semánticas.

Objetivos Específicos

- Establecer un marco teórico-conceptual sobre los antecedentes de la problemática en torno al ordenamiento jurídico de leyes, fallos, sentencias y otros tipos de documentos jurídicos, como así también de los medios manuales o mecanizados para acceder a los mismos, principalmente enfocado en los sistemas informatizados y de búsqueda por ocurrencia de términos ya sea de coincidencia plena o aproximada (ya sea por criterios estadísticos o semánticos).
- Comprender la relación entre términos de búsqueda, tesauros, digestos jurídicos y ontologías aplicadas al ámbito jurídico.
- Determinar la pertinencia de distintas soluciones tecnológicas de base de datos no relacionales para el almacenamiento y la recuperación de documentos judiciales con tecnologías semánticas.
- Especificar una ontología integrada con ontologías de alto nivel y específicas del dominio para la gestión de documentos jurídicos.
- Proponer una solución unificada de organización de la información jurídica en relación a los las estrategias de búsqueda semántica.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación está conformado de la siguiente manera:

Director del Proyecto:
 Ing. Héctor Javier Ruidías
 Co-Director:
 Dr. Juan Manuel Lezcano
 Docentes-Investigadores:
 Ing. Karina Beatriz Eckert
 Abogada Carolina Vanesa Rosas
 Tutor/Asesor
 Dra. María L. Caliusco

El Ing. Héctor J. Ruidías se encuentra desarrollando su tesis doctoral sobre tecnologías semánticas en UTN Facultad Regional Santa Fe bajo la dirección de la Dra María L. Caliusco.

Además el equipo incluye a estudiantes de grado de las carreras de Ingeniería en Informática y Derecho quienes colaboran en diversas tareas.

5. REFERENCIAS

- [1] G. L. Paul y J. R. Baron, "Information Inflation: Can the Legal System Adapt?", *Rich JL Tech*, vol. 13, pp. 10–17, 2007.
- [2] T. Berners-Lee, J. Hendler, O. Lassila, y others, "The semantic web", *Sci. Am.*, vol. 284, no 5, pp. 28–37, 2001.
- [3] R. G. Brenna, "El Digesto Jurídico. Un camino hacia la democratización de la información jurídica". [En línea]. Disponible en: <http://www.dab.com.ar/articles/53/eldigesto-jur%C3%ADdico-un-camino-hacia-la-democratiza.aspx>. [Accedido: 14-nov-2017].
- [4] M. A. Ciuro Caldani, "EL COMPLEJO DEL FUNCIONAMIENTO DE LAS NORMAS", *INVESTIGACIÓN Y DOCENCIA*, vol. 40, 2007.
- [5] N. Guarino, *Formal Ontology in Information Systems: Proceedings of the First International Conference (FOIS'98)*, June 6-8, Trento, Italy IOS Press, 1998.
- [6] T. R. Gruber, "A translation approach to portable ontology specifications", *Knowl Acquis*, vol. 5, pp. 199–220, jun. 1993.
- [7] T. Berners-Lee, "Semantic Web on XML". [En línea]. Disponible en: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html> [Accedido: 26-2-2018].
- [7] K. Kaneiwa, M. Iwazume, y K. Fukuda, "An Upper Ontology for Event Classifications and Relations", en *AI 2007: Advances in Artificial Intelligence*, M. A. Orgun y J. Thornton, Eds. Springer Berlin Heidelberg, 2007, pp. 394–403.
- [8] V. Mascardi, A. Locoro, y P. Rosso, "Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation", *IEEE Trans. Knowl. Data Eng.*, vol. 22, no 5, pp. 609–623, may 2010.
- [9] J. a. P. J. Breukers y R. J. Hoekstra, "Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law", *CEUR Workshop Proc.*, 2004.
- [10] R. Benjamins, *Law and the Semantic Web - Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*. Springer, 2005.
- [11] J. Breuker, A. Elhag, E. Petkov, y R. Winkels, "Ontologies for legal information serving and knowledge management", en *Legal Knowledge and Information Systems, Jurix 2002: The Fifteenth Annual Conference*, 2002, pp. 1–10.
- [12] F. Enrique y M. Lopez, "Desarrollo de algoritmos de procesamiento para la indexación y la búsqueda en lenguaje natural del contenido de un digesto basado en tecnologías semánticas", *Universidad Nacional del Litoral*, 2016.