

Adquisición inteligente del conocimiento jurídico

Miguel-Ángel López Alonso
Universidad Carlos III de Madrid

0.1. Resumen

La normalización documental del lenguaje jurídico es una condición imprescindible para crear sistemas jurídicos que atiendan las demandas específicas de los juristas. Se propone un enfoque metodológico que integre las teorías jurídicas y de la Inteligencia Artificial utilizando los razonamientos inductivo y analógico, con la aspiración de identificar y establecer relaciones conceptuales no expresadas mediante el análisis de la información subyacente en las relaciones lexicales y terminológicas de los términos de la red semántica conceptual. Se comentan diferentes prototipos de sistemas expertos jurídicos y algunos sistemas híbridos, que incluyen hipertexto, redes neuronales y algoritmos genéticos.

Palabras clave: Sistemas de documentación jurídica. Sistemas expertos. Inteligencia Artificial. Tesoros. Redes semánticas. Redes neuronales. Algoritmos genéticos.

0.2. Abstract

Standardising juridical terminology is a necessity to create juridical information systems capable to fulfil the specific information needs of jurists. A methodological approach is proposed, that integrates the juridical theories and Artificial Intelligence, using both inductive and analogic reasoning to identify and establish conceptual relations by analysing the information that underlies to the lexical and terminological relations of the terms in the conceptual network. Different prototypes of juridical expert systems and some hybrid systems that include hypertext, neural networks and genetic algorithms are summarised.

Keywords: Juridical information systems. Expert systems. Artificial Intelligence. Thesaurus. Semantic networks. Neural networks. Genetic algorithms.

1. Introducción

Tomando el conocimiento jurídico como uno de los dominios científicos en los que existe más amplia interconexión de información, trataremos de ahondar en las peculiaridades de la normalización de su lenguaje, previa diferenciación entre el sublenguaje técnico-jurídico y las palabras polisémicas, y en su estructuración mediante tesauros conceptuales para su aplicación a las reglas de la base de conocimientos de los sistemas expertos.

Teniendo siempre en cuenta que la nueva generación de gestores “inteligentes” de tesauros, apoyados en las técnicas de la Inteligencia Artificial sobre sistemas expertos, aspiran a identificar y establecer nuevas relaciones conceptuales (no declaradas expresamente en el proceso de creación del tesoro) mediante el análisis de la información subyacente en las relaciones lexicales y terminológicas de los términos de una red semántica conceptual.

Analizaremos brevemente las condiciones de un sistema jurídico documental que atienda las demandas específicas de los juristas, mediante un enfoque que integre las teorías de la Inteligencia Artificial y las jurídicas, utilizando los razonamientos inductivo y analógico.

Comentaremos algunos sistemas expertos jurídicos que tratan flexiblemente los *continuos cambios de la normativa jurídica* y su repercusión en la actualización de las reglas de su base de conocimientos.

Dado que para evitar las ambigüedades en el ámbito jurídico es obligado tener siempre en cuenta el contexto concreto de aplicación, serán preferibles los métodos conceptual-lingüísticos que utilizan el razonamiento inductivo-analógico a los métodos deductivos basados en reglas; lo que dará lugar a la potenciación del estudio de los sistemas neuronales artificiales, capaces de generar sus propias reglas para adaptarse a los problemas que no se presentan “a priori” en términos concretos.

Por último, mencionaremos los últimos prototipos de sistemas expertos que tratan de resolver las ambigüedades en el procesamiento del lenguaje natural a través de interfaces interactivos sistema-usuario.

Así mismo, adelantaremos como apuesta de futuro el avance de los sistemas híbridos: que integran las redes neuronales junto con el hipertexto en los sistemas expertos y las nuevas líneas de investigación sobre algoritmos genéticos.

2. Nomenclatura jurídica: su estructuración y normalización

La Ciencia del Derecho se apoya básicamente en el lenguaje natural, salvo determinadas palabras o expresiones que son propias de la terminología jurídica específica.

Creemos con Jorge Páez (1992) que, debido a la amplia interconexión de información existente en este dominio científico, coexisten dentro de este lenguaje tres tipos de términos:

- los de contenido técnico-jurídico o sublenguaje jurídico, a veces muy alejado del sentido que a dichos términos les otorga el lenguaje natural,
- los del lenguaje común, que han servido como elementos de regulación de aspectos y relaciones jurídicas en cierto contexto social, y
- los ambiguos, que permiten interpretar los documentos jurídicos a tenor de situaciones concretas o del momento en que se aplican.

En un estudio reciente de Haas (1993), basado en resúmenes para determinar el porcentaje de palabras que se pueden considerar términos del sublenguaje científico, sólo un 25% de ellos lo son plenamente, pero, variando ligeramente de unas a otras áreas del conocimiento.

Algunas de las características de este sublenguaje jurídico son:

- La abundancia de multitérminos para definir los conceptos o entidades jurídicas.
- La diferenciación con el lenguaje natural general se establece a través de su variada “riqueza terminológica”, lo que obliga a estrechar el cuidado en la elección de definiciones jurídicas correctas.
- Los términos no tienen cabida en los diccionarios generales y cuando la tienen no incluyen la acepción deseada para el dominio jurídico.
- La utilización por los especialistas jurídicos de una gramática propia de este sublenguaje, no claramente diferenciada de la del lenguaje natural.
- La existencia de un estilo diferenciado en los documentos jurídicos de uso profesional.

La adecuada representación de los conceptos o situaciones jurídicas exige gran precisión en la selección de los términos, tanto en la indización como en las ecuaciones de búsqueda de documentos.

La respuesta tradicional pasa por la compilación de una lista autorizada de términos preferidos (vocabulario controlado o tesauro) que proporciona una normalización terminológica y un registro de sus relaciones.

Estos tesauros tradicionales son en realidad una superserie de sublenguaje controlado en un dominio científico específico, que se usa tanto durante:

- El proceso de indización, como ayuda en la identificación de los conceptos representados por los términos desconocidos.

- El proceso de recuperación, como fuente de identificación de nuevos términos controlados que permitan la formulación de nuevos conceptos y aumenten la precisión de las búsquedas booleanas.

El avance en el procesamiento del lenguaje natural tiende cada día más a que el usuario sea una “máquina” o sistema automático de tratamiento de la información, precisándose tesauros específicamente desarrollados para cada aplicación que representen los conceptos de forma mucho más formal, detallada y explícita de lo que sería necesario para la mera utilización humana.

La normalización de terminologías no sólo simplifica notablemente la comunicación entre los especialistas, sino que ayuda al establecimiento de bancos de datos automatizados y su recuperación mediante sistemas expertos; previo establecimiento de las jerarquías de sus conceptos terminológicos en los tesauros conceptuales.

La normalización de terminos técnico-jurídicos es un proceso complejo que comprende varias operaciones:

1. La definición de términos, tendente a la utilización de aquellos términos cuyos conceptos han sido evidenciados por la práctica como suficientemente precisos para la delimitación de determinados aspectos (eludiendo su sustitución por sinónimos o cuasinónimos más coloquiales).
2. La unificación de clasificaciones y jerarquías de conceptos, dada la estructuración del Derecho y los sucesivos intentos de codificación del mismo, existiendo unas específicas relaciones jerárquicas entre los diferentes conceptos jurídicos, en función de su ubicación sistemática en los textos legales, que facilitan su interpretación y delimitación.
3. La eliminación de sinónimos, como una de las características más sobresalientes del lenguaje jurídico, que obliga a recurrir a la creación de multitérminos para evitar la ambigüedad y diferenciar las múltiples formas con que puede definirse un mismo concepto (acto, hecho, negocio jurídico, etc.); similar problema tendremos con la reducción de los homónimos, debiendo adoptarse los términos que eliminan los equívocos en nuestro contexto jurídico y sean de más fácil traducción a otras lenguas del entorno socio-lingüístico (isomerismo).
4. La fijación de abreviaturas y símbolos, en la que al igual que en otras disciplinas se deberá establecer una sinonimia absoluta entre el conjunto de términos y su representación abreviada en siglas.

5. La eliminación de términos vacíos, no sólo artículos, preposiciones, adjetivos o nombres de aplicación generalizada, pronombres o adverbios de muy alta frecuencia de uso, sino incluso los términos de aplicación demasiado generalizada dentro del campo jurídico.
6. La eliminación de frases semánticamente equivalentes, debido a que el distinto orden de los términos puede confundir al usuario en la representación de diferentes conceptos (isomorfismo).

Deberemos destacar que, en el momento actual de desarrollo de las comunicaciones a nivel mundial, la necesidad de normalización de un *formato* intermedio para los intercambios electrónicos de ficheros terminológicos (TEI), permitiendo la conversión adecuada entre los distintos formatos de los bancos de datos nacionales e internacionales, es cada día más indispensable (Melby, 1995).

3. Teoría de los sistemas jurídicos

El enfoque está basado en los métodos conceptuales que parten de la observación del proceso analítico-sintético, dado el consenso sobre la obtención de un mayor número de documentos y de mayor relevancia mediante procedimientos de búsqueda basados en la *interpretación jurídica* (Blair y Maron, 1985), con preferencia a los procedimientos lingüísticos de comparación entre términos y sus interrelaciones.

Partiendo de la generalizada opinión sobre la necesidad de fusión de los conocimientos tradicionales de la Teoría de la Recuperación de la Información (IR) y los de la Inteligencia Artificial e Ingeniería del Conocimiento (AI/KO), trataremos de articular un marco global teórico que nos sirva de punto de partida para el diseño de más perfeccionados sistemas expertos jurídicos.

Los documentalistas jurídicos debemos fijarnos como meta primordial el estudio pormenorizado de la *base de conocimientos* del sistema, dejando el diseño del núcleo del sistema o del interface de usuario a los ingenieros del conocimiento.

3.1. Requerimientos previos para la integración de la recuperación de la información

3.1.1. Integración de la recuperación con la indización previa.

Para ello, destacaremos que la integración de dichos conocimientos requiere, en primer lugar, que la recuperación de la información no sea una actividad aislada, debiendo *integrarse* con la indización previa y mantener un alto nivel de interrelación entre sus distintos sistemas informáticos.

Aunque en el establecimiento de las reglas y procesos para construir el núcleo de los sistemas expertos parece existir un amplio consenso (Gibb y Smart, 1990), no ocurre lo mismo con las *metodologías de indización* de los documentos a clasificar en la base de conocimientos.

Además de las diferentes preferencias del indizador, las distintas formas de indización pueden variar según las situaciones a cubrir: algunos tipos de indización son apropiados cuando se desea generar índices perfectos y otros tipos son adecuados para que el usuario pueda ejercitar sus conocimientos con todo cuidado.

Dado que la única indización “perfecta” es el texto completo del documento, la tarea de asignar términos del lenguaje controlado a su representación provoca distintos *problemas de indeterminación*, siendo los más conflictivos los derivados de los diversos esquemas de clasificación del área jurídica, y aquellos relacionados con el segundo.

En cuanto al primer problema, recordemos que la efectividad en la recuperación de un documento viene unida a la forma en que el documento haya sido clasificado en la base de datos del sistema informativo; de acuerdo con los términos y frases usadas para su indización, deberá permitir la descripción de su contexto (*autor, título, fecha, longitud, tipo, editor, etc.*) y/o de sus materias tratadas.

La inconsistencia en la selección de términos por el indizador varía en función del área temática, el conocimiento del dominio temático, el entrenamiento y/o experiencia del indizador, el uso de vocabulario controlado o no, la carga emocional que el indizador pone en la argumentación, etc.

Parece generalmente reconocido que la indización automática, basada en técnicas estadísticas o probabilísticas, ha alcanzado su máximo techo de efectividad y que nuevas aproximaciones deben investigarse para mejorar sustancialmente la efectividad de la recuperación (Croft y Thompson, 1987).

De la reunión de las teorías de la Ciencia de la Información y de la Inteligencia Artificial han surgido *poderosas técnicas de indización automática*, integradas en sistemas comerciales (ej.: TOPIC o *Construe*) o en prototipos que tratan de capturar el alcance cognoscitivo de los documentos mediante la identificación de los descriptores adecuados (por medio de tesauros, esquemas clasificatorios o estructuras específicas).

Actualmente, destacan como nuevos procedimientos de indización (distintos de los silogismos u otros procedimientos deductivos) el *razonamiento inductivo* de las situaciones prácticas en cada dominio temático; definidas como por Sillence (1992) como:

complejas unidades de conocimiento que codifican atributos y tipifican ejemplos de categorías generales (formadas por entidades o categorías funcionales y reglas de interrelación), dejando a un lado las propiedades atípicas.

Para solucionar el segundo, aquel referente al problema de la organización del conocimiento jurídico debemos seguir los siguientes pasos:

- Primeramente se debe describir el ámbito conceptual abarcado por cada entidad del dominio (estructuras, situaciones, redes semánticas, etc.), en tres fases: nombre del concepto, diversas formas en que el concepto puede ser identificado dentro de un pasaje del texto y descripción de como se relacionan entre sí dichos conceptos (sinonimia, jerarquía superior, jerarquía inferior, referencias cruzadas, ejemplos, etc.).
- A continuación se definirá la estructura conceptual completa del dominio para ser almacenada en la base de conocimientos del sistema.

3.1.2. La recuperación de la información en el marco del trabajo habitual

La recuperación documental debe producirse en el marco del *trabajo habitual* de los usuarios con sus habituales sistemas informáticos de tratamiento documental (ej.: procesadores de texto, correo electrónico, archivos, etc.). Para desarrollar todo ello debemos realizar una serie de propuestas con respecto a la anterior premisa:

3.1.2.1. *Premisas para un buen diseño de la base de conocimientos del sistema experto*. Entre las premisas a destacar para desarrollar un buen diseño nos hallamos con:

- a) Que el trabajo de tratamiento de la documentación, en un entorno jurídico integrado, es una parte esencial del trabajo de los especialistas jurídicos, dado que dichos sistemas contienen documentos que deben soportar continuos cambios (inserciones, anulaciones o actualizaciones).
- b) Que la flexibilidad necesaria para soportar fácilmente los sucesivos cambios legislativos debe apoyarse en un conocimiento previo de cuales serán los grupos de usuarios y que cambios exigirán de éste.

Entre las posibilidades de ajustar el sistema (por determinados grupos de usuarios especializados en derecho) deben tenerse principalmente en cuenta que:

- 1) Para controlar los cambios en la estructura de la base de conocimientos y en el acceso a ella, debe conocerse cuando ha sido introducido cada documento y mantener un tesoro actualizado automáticamente que permita la homogenización semántica.

- 2) La evolución de los datos contenidos en la base de conocimientos, debe permitir la adición de nuevas leyes y la modificación de otras sin necesidad de conocimientos informáticos avanzados.
- 3) La posibilidad de añadir nuevos descriptores a los documentos previamente indizados, alterando ciertas subestructuras de clasificación del dominio jurídico o incluso añadir notas personales, debe realizarse bajo la supervisión periódica del administrador del sistema.

3.2.2. *Problemas para la formulación de una teoría integrada multidisciplinar*

La formulación de una teoría integrada interdisciplinar implica tratar dos tipos de problemas: los procesos relacionados con la Ingeniería del Software y/o del Conocimiento y los problemas teóricos relacionados con la representación del conocimiento y/o del razonamiento jurídicos.

3.2.2.1. *Los procesos relacionados con la Ingeniería del Software y/o del Conocimiento.* Estos no son específicos de los Sistemas Expertos Jurídicos sino de las distintas fases de diseño e implementación general de sistemas cognitivos (por ejemplo, interfaces de cooperación sistema-usuario, paradigmas de representación del conocimiento: estructuras, situaciones, redes semánticas, etc., o de recuperación del mismo), cuyos desarrollos más consolidados comentaremos en el epígrafe cuarto de este trabajo.

3.2.2.2. *Los problemas teóricos relacionados con la representación del conocimiento y/o del razonamiento jurídicos.* Estos problemas incluirán a su vez:

- a) La comprensión de la *naturaleza teórica* del conocimiento y razonamiento jurídico, derivada de:
 1. La naturaleza compleja del derecho, pues no se puede subestimar la complejidad del campo jurídico, debiendo alcanzar un grado de precisión en su representación que permita capturar las sutilezas del sublenguaje jurídico.
 2. La categorización del conocimiento jurídico y la estructura normativa del derecho, partiendo de la hipótesis de que detrás de cada fragmento jurídico existe una norma completa, se puede establecer una clasificación de categorías limitada que permita intercalar cualquier fragmento jurídico arbitrario dentro de esas normas generales, sin distorsionar o cambiar la estructura general del derecho (Brouwer, 1994).
 3. La integración de los componentes de la estructura interna de cada una de estas categorías. La teoría de la representación del conocimiento jurídico debe permitirnos alcanzar una mayor integración entre los distintos componentes de la norma general, fijando patro-

nes que deben ser tenidos en cuenta en la plasmación escrita de las normas obligatorias y sus consecuencias para el razonamiento jurídico (por ejemplo, tanto el lugar de un fragmento jurídico dentro de una norma general como el momento de uso pueden afectar el significado de ésta).

4. La relación entre las diferentes categorías jurídicas. La tarea de relacionar estas categorías es muy ardua, exigiendo estudios teóricos de semántica jurídica, de teoría de los actos jurídicos y especialmente una cuidadosa investigación de numerosos ejemplos de práctica jurídica.

b) *La aplicación práctica* de las teorías de la Inteligencia Artificial al campo jurídico (por ejemplo, los distintos tipos de razonamiento: deductivo, inductivo o analógico, procesamiento del lenguaje natural, etc.).

Tratando de integrar las teorías de la Inteligencia Artificial y las Jurídicas, seguiremos a Wahlgren (1994) en que la obtención de una verdadera teoría interdisciplinar no puede obtenerse:

- ni importando más teoría de la AI en el campo jurídico (aunque sean nuevas investigaciones),
- ni articulando aspectos jurídicos conocidos en el campo de la AI (a pesar de que sean reexaminados e interpretados desde una perspectiva informática).

3.2.3. Necesidad de una teoría de naturaleza disciplinar

El campo unificado de la Inteligencia Artificial y el Derecho requiere una *teoría de naturaleza interdisciplinar* que refleje requerimientos y características unívocas de la materia combinada y no sea meramente el resultado de *investigaciones teóricas* entre filósofos del derecho e investigadores de la AI (McCarthy, 1990)), sino realizada a través de la integración de variados *esfuerzos prácticos* a menor escala que fijen los elementos básicos, relaciones y problemas a investigar (siguiendo criterios predefinidos de estructura, generalidad, transparencia, formalización, particularización, flexibilidad, etc.).

a) Algunos como Mital y Johnson (1992), afirman que la ley es reducible a un conjunto de reglas que pueden ser extraídas de los textos jurídicos y que la forma más sencilla de representar una situación jurídica concreta pasa por entresacar todos los rasgos que la definen, aplicándoles valores o pesos a cada uno de ellos según su mayor o menor importancia en la toma de una decisión específica. El problema básico de estos sistemas de *razonamiento deductivo* surge cuando se necesita la interpretación de códigos o leyes ambiguas.

- b) Otros como Wildemast y De Mulder (1992), sostienen que dado que los conceptos pueden variar de un usuario a otro, de una época a otra o incluso dentro de la misma sesión de búsqueda (interactuando diferentes alternativas según la actuación de la parte contraria); para representar los documentos referentes a una situación jurídica concreta se les debe agrupar en índices (teniendo en cuenta su similitud con respecto de unos llamados “superdocumentos”), cuyo perfil (Gelbart y Smith, 1990) se genere automáticamente a partir del texto original, por medio de un método mixto conceptual-lingüístico que permita una interpretación contextual. Estos sistemas utilizan el *razonamiento inductivo* (especie de análisis estadístico de las interpretaciones dadas por distintos juristas) junto con el *razonamiento analógico* (especie de comparación de los precedentes existentes y la situación en estudio) para resolver la ambigüedad legal.

4. Sistemas de información jurídica

Hasta los años 90, las investigaciones en derecho, Inteligencia Artificial e informática para la construcción de sistemas expertos legales, han sido mayoritariamente de tipo metodológico, limitándose los sistemas prácticos a prototipos para investigación o demostración.

En estos últimos años, los intentos por capturar el conocimiento y el razonamiento de los expertos jurídicos, han tenido un limitado progreso y escasos beneficios directos para la profesión jurídica, dado que:

- 1) A pesar de que han surgido variados productos comerciales para ayudar a los profesionales jurídicos en el acceso a bases jurídicas, elaboración de borradores o consulta de información; su apoyatura teórica no rebasa la de las bases documentales relacionales u orientadas a objetos (por ejemplo, *21st Century Lawyer* o *Gryphon*).
- 2) Ha habido pocos intentos para producir sistemas inteligentes de ayuda al razonamiento jurídico de los profesionales (toma de decisiones prejudiciales o emisión de informes jurídicos), utilizando esterotipos y estudiando cuestiones conceptuales generales, y los pocos consolidados proporcionan limitadas soluciones que no abarcan a todo el dominio jurídico (por ejemplo, LEDA en Holanda).

Para continuar avanzando en estas tareas jurídico-cognitivas deberá recurrirse al campo de la ingeniería del conocimiento como:

- Ayuda para la creación de un entorno de información que proporcione soporte a la formulación de borradores legislativos (Allen, 1980) y, también, como ...

- Apoyo de los consejeros jurídicos de los políticos para la toma de decisiones sobre el tipo de leyes a desarrollar (Bench-Capon, 1994).

Los *sistemas aplicados* de mayor utilidad, funcionando e intensivamente probados, provienen de investigadores australianos que han implementado las diversas teorías difundidas en la década anterior (por ejemplo, La “Privacy Workstation” de los investigadores MowBray y GreenLeaf, en un sistema experto hipertextual desarrollado para diversos departamentos del gobierno australiano).

Debido a las peculiaridades de la información jurídica: excesivo volumen de textos y almacenamiento a texto completo, cuando se recupera este tipo de información en los sistemas tradicionales (basados en operadores booleanos), suele surgir *ruido* en las referencias, debido a las interrelaciones entre: leyes, casos jurisprudenciales y leyes o doctrina legal, casos y leyes.

Cuando no existen estructuras conceptuales de los términos del dominio jurídico (tesauros, tablas de clasificación, etc.) que orienten las preguntas hacia el dominio concreto de la aplicación, se suele perder gran cantidad de información debido a la *ambigüedad* en el uso de los términos o a la no utilización de los sinónimos.

Para evitar largos tiempos de acceso a los terminales, se deben plantear las consultas de la manera más exacta posible, para lo que se necesitan gestores de tesauros “inteligentes” basados en estructuras semánticas conceptuales que puedan inducir relaciones asociativas capaces de analizar las preguntas, compararlas con los documentos almacenados de la base de conocimientos e identificar aquellos o partes de ellos que sean relevantes.

Con la finalidad de que los modernos *tesauros* sirvan más para la representación de relaciones entre conceptos que entre términos, distintas adiciones se han tratado de efectuar partiendo de un nuevo enfoque dentro de la perspectiva de la Inteligencia Artificial, considerándolos como *una red semántica de conceptos* que cubren un dominio particular del conocimiento y cuyas relaciones semánticas puedan considerarse como etiquetas “oficiales” (reglas de representación) de dichos conceptos.

Seguidamente revisamos brevemente sucesivas aportaciones al enfoque anterior (Paice, 1991):

- 1º) Reconocimiento de que un tesoro puede ser explorado como un documento hipertextual cualquiera dentro de una red semántica. A partir de uno o más nodos desde los que se formule la pregunta y recopilando términos para aumentar o reemplazar a sus predecesores en la búsqueda (según la mayor o menor prioridad asignada, en función de la distancia simbólica entre el nuevo documento y el núcleo más próximo).

- 2^a) Mejora de la red semántica anterior mediante la ponderación de los vínculos entre nodos, dando preferencia a los términos alcanzados desde mayor número de caminos. Otros criterios de ponderación pueden ser: el tipo de relación (*use* o *use for*, o bien BT, NT o RT), el número de ocurrencias de los descriptores en documentos (establecidos como modelos de la base de datos), la longitud desde el nodo inicial, etc.
- 3^a) Expansión de los términos iniciales del texto para la comparación de situaciones ideales “no preexistentes” en un determinado texto sino “creadas” (comprendidas, descritas y almacenadas) para representar la “idea implicada” en dicho texto. Estos “resúmenes” de pasajes del texto contendrían los términos del tesoro incluidos en el texto en análisis, los términos adyacentes a los incluidos que no aparecen en el texto y los términos interrelacionados entre los dos anteriores.
- 4^a) Investigaciones sobre sistemas dotados de interfaces interactivos hombre-máquina o “inteligentes”, capaces de permitir a los usuarios encontrar los mejores caminos a seguir mediante la selección de los mejores términos; utilizando el tesoro como herramienta de precisión en la recuperación de documentos. El conocimiento práctico de los expertos en un dominio científico (heurística) es almacenado en la base de un sistema experto para ser utilizado en la ejecución automática de procedimientos analíticos que permitan lograr una acertada representación del contenido de los documentos.

La habilidad para la aplicación de la experiencia durante el proceso de recuperación puede contestar a preguntas de los usuarios tales como (Jones, 1995) ¿cuándo usar el tesoro?, ¿dónde empezar a usarlo?, ¿qué relaciones seguir?, ¿qué nuevos términos seleccionar?, ¿cómo incluir los nuevos términos en la pregunta?, etc o permitir la ampliación o restricción de la búsqueda, mediante interacciones al sistema del tipo de:

- use términos más generales y/o términos relacionados para un más amplio tratamiento de la situación buscada,
- muévase más arriba o más abajo en la jerarquía del tesoro para modificar su especialización, o
- use sólo vocabulario controlado para lograr una mayor especificidad en la búsqueda.

Partiendo de la experiencia acumulada, por los ingenieros del conocimiento holandeses, en la implementación de sistemas inteligentes que sirvan no solo para responder preguntas jurídicas sino también para la preparación de proyectos de ley; parece que los modelos directos (deductivos o inductivos) están siendo abandonados en beneficio de los *modelos indirectos* (conocimiento conceptual)

que utilizan un lenguaje intermedio para la representación de las situaciones jurídicas y permiten una adecuada indización-recuperación integrada de los documentos jurídicos.

Trabajando en la línea de anteriores teóricos jurídicos como Brouwer, Hart, Kelsen, Ross, Ruiter y Von Wright o Kralingen (1993) propone el uso de un lenguaje intermedio de “estructuras normativas” como medio de formalización y conceptualización del conocimiento jurídico que sea capaz de expresar todo tipo de relaciones de lógica-normativa entre la legislación específica en estudio.

En la misma línea Voermans y Verhareem (1993), de la Facultad de Derecho/Instituto de Lenguaje Tecnológico e Inteligencia Artificial de la Universidad de Tilburg en Holanda, han utilizado la tecnología del hipertexto dentro del sistema semi-inteligente LEDA (usado por expertos jurídicos del Ministerio Holandés de Justicia como ayuda en la elaboración de proyectos de ley), para proporcionar a los expertos jurídicos:

- documentos estructurados en una red semántica de referencias cruzadas que les permita conservar su significado propio (leyes, casos jurisprudenciales, doctrina legal, etc.), junto con ...
- una representación visual de la colección de documentos a través de la definición de varios tipos de conexiones dinámicas (relaciones), utilizando el formalismo de las estructuras normativas de Kralingen.

Apoyando este desarrollo de LEDA, Svensson (1993) aplica el método de la micro-simulación al estudio de los efectos de la aplicación estricta de diversos apartados de una ley a casos específicos, midiendo sus desviaciones prácticas y la frecuencia de uso de las distintas subsecciones de ella, para tratar de ayudar a la toma de decisiones sobre futuros cambios legislativos.

5. Conclusiones

Existe una tendencia creciente a proporcionar al usuario soluciones de acceso a la información que, mediante el análisis de los *corpora* existentes y la utilización de los conocimientos del usuario, permitan sesiones interactivas de búsqueda y mejoren las preguntas iniciales al sistema.

En un estudio jurídico sobre “máquinas inteligentes” tenemos a favor de la buena construcción de sistemas expertos el que las leyes se pueden descomponer en conjuntos de reglas, y tenemos en contra que no todo el derecho está reflejado en forma de códigos sino también mediante decisiones jurisprudenciales y doctrina jurídica.

Principalmente, se trata no solo de identificar los documentos relevantes, sino de localizar previamente los pasajes y las frases significativas, para ser utilizadas

en las búsquedas que toman como modelo un documento tipo en la base de conocimientos y lo utilizan como perfil para la búsqueda de sus similares.

Estos “perfiles de información” relevante se utilizan para construir *árboles de decisiones* que representen la pregunta inicial mediante reglas y, usando la analogía y la creación de lazos hipertextuales entre los pasajes de texto más representativos, permitan al usuario la navegación entre textos legales con propiedades similares; tanto para su recuperación inmediata como para servir como fuente para el diseño de nuevos documentos legislativos.

Aunque la mayoría de los prototipos jurídicos (PROLEX5, GREBE, CABARET o IKBALS) (Hunter y Zelevnikow, 1995) no incluyen técnicas de Inteligencia Artificial y los estudios sobre recuperación de la información de bases de datos documentales siguen siendo importantes para el soporte de los sistemas jurídicos, la mayor parte de los investigadores tratan de desarrollar dichas técnicas sobre redes neuronales artificiales, dotándolas de *instrumentos de información* que permitan utilizar la inducción y la inferencia en la toma de decisiones de los juristas .

Estos instrumentos de información (denominados *interfaces* “inteligentes”) deben ser capaces de dar información de las relaciones entre conceptos jurídicos, mediante su interconexión en base a relaciones conceptuales preestablecidas; de esta forma el usuario tiene acceso a ellos mediante una representación gráfica de la red, por la que puede rastrear los documentos siguiendo su línea de interés, con sus conexiones visibles a través de vínculos hipertextuales.

Las teoría de las *redes neuronales* artificiales (experimentadas con éxito en medicina y finanzas) proporcionan un mapa de las variables introducidas en la búsqueda y las obtenidas en la recuperación, siendo capaces de almacenar información de forma distribuida (interconexionando pesos entre las unidades documentales mínimas: términos, frases o párrafos), de aprender tras un previo período de adiestramiento (actualizando pesos) y de generalizar cuando reciben más variables de nuevas búsquedas (Martín y Serrano Cinca, 1995).

Dado que las propuestas anteriores no bastan, por sí solas, para alcanzar una visión suficientemente general de la compleja naturaleza de la inteligencia humana y de sus ciencias cognoscitivas; otras líneas de investigación tratan de profundizar en la utilización de técnicas de tipo ascendente, como por ejemplo, los *algoritmos genéticos*, debido a sus enormes propiedades de auto-evolución; habiendo dado buenos resultados en áreas tan diversas como: informática, ingeniería, finanzas, investigación penal, etc. (Goldberg, 1994).

Por último apuntar, como apuesta de futuro, que con estos modelos conceptuales será posible la creación de una macroestructura de conocimientos que abarque la terminología usada por distintos sistemas, sus sistemas clasificatorios

y sus bases de conocimientos; como vía de unión para la comunicación entre diferentes agentes jurídicos inteligentes del tipo usado en las búsquedas con agentes virtuales de Internet.

6. Bibliografía

- Allen, L.E. (1980). *Language, law and logic; plain legal drafting for the electronic age.* // *Computer Science and Law*, Cambridge : Cambridge U. Press, 1980. p. 75-100.
- Bench-Capon, Trevor (1994). *An Incremental Approach to Legal Drafting Support.* // *Proceedings of JURIX'94.* Lelystad : Kroninklijke Vermande (Netherlands), 1994.
- Blair, D.C. ; Maron, M.E. *An evaluation of retrieve effectiveness for a full-text document-retrieval system.* // *Communications of the ACM.* 28 : 3 (1985) 289-299.
- Brouwer, F. W.(1994). *Legal Knowledge Representation in the Perspective of Legal Theory.* // *Proceedings of JURIX'94.* Lelystad : Koninklijke Vermande, Netherlands, 1994.
- Croft, W.B.;Thompson, R. *I3R: A new approach to the design of document retrieval systems.* // *Journal of the ASIS.* 38 : 6 (1987) 389-404.
- Geifbart, D. ; Smith, J.C.(1990). *Toward a Comprehensive Legal Information Retrieval System.* // *Proceedings of DEXA'90.* Vienna,; 1990.
- Gibb, F. y Smart, G. *Structured information management: using new techniques for processing text.* // *Online review.* 14 : 3 (1990) 159-171.
- Golberd, D.E. (1994). *Genetic and evolutionary algorithms come of age.* // *Communications of the ACM.* 37 (1994) 113-119.
- Hunter, Dan ; Zeleznikow, John (1995). *An Overview of some Reasoning Formalisms as Applied to Law.* // *KUB home page - Arthur van Horck - .* Institute for Language Technology and Artificial Intelligence. Tilburg University : Netherlands, 1995
- Jones, Susan et al.(1995). *Interactive Thesaurus Navigation: Intelligence Rules Ok?* // *Journal of the American Society for Information Science.* 46 : 1 (1995) 52-59.
- Martin, Bonifacio ; Serrano Cinca, Carlos (1995). *Fundamentos de redes neuronales artificiales: hardware y software.* // *Scire : Representación y Organización del Conocimiento.* 1 : 1(1995) 103-125.
- Mc Carthy, J. (1990). *Some Expert Systems need Common Sense.* // *Proceedings of the Fourth International Symposium of Knowledge Engineering.* Universidad Politécnica de Madrid, 7-11 may 1990, Barcelona, Spain.
- Melby, Alan; E-Tif (1995). *An Electronic Terminology Interchange Format.* // *Computers and the Humanities,* 29 (1995) 159-165.
- Mital, V. Y ; Johnson, L.(1992). *Advanced information system for lawyers.* London: Chapman & Hall, 1992. 306 p.
- Paez, Jorge (1992). *Lenguajes Jurídicos Estructurados.* Madrid : ICADE, 1992.
- Paice, Chris D. A. (1991) *Thesaural Model of Information Retrieval.* // *Information Processing Management.* 27 : 5 (1991) 433-447.
- Sillince, J.A.A (1992). *Argumentation-based indexing for information retrieval from learned articles.* // *Journal of Documentation.* 48 : 4 (1992) 387-405.

- Svensson, Jorgen et al. (1993). Using Knowledge Based Micro Simulation in Analysing the Application of Legislation. // Proceedings of JURIX'93. Lelystad : Kroninklijke Vermande (Netherlands), 1993.
- Van Kralingen, Robert et al. (1993). Norm Frames in the Representation of Laws. // Proceedings of JURIX'93. Lelystad : Kroninklijke Vermande (Netherlands), 1993.
- Voermans, W.; Verharem, E.M.(1993). LEDA: a semi-intelligent Legislative Drafting-Support System. // Proceedings of JURIX'93. Lelystad : Kroninklijke Vermande, Netherlands, 1993.
- W. Haas, Stephanie ; HE, Shaoyi (1993). Toward the automatic identification of sublanguage vocabulary. // Information Processing Management. 29 : 6 (1993) 721-732.
- Wahlgren, P. A.(1994). General Theory of Artificial Intelligence and Law. // Proceedings of JURIX'94. Lelystad: Koninklijke Vermande, Netherlands, 1994.
- Wildemast, C.A.M. ; De Mulder, R.V.(1992). Same design considerations for a conceptual legal information Retrieval System. // Proceedings of JURIX'92. Lelystad : Koninklijke Vermande, Netherlands, 1992.