

V. INFORMÁTICA JURÍDICA DOCUMENTARIA	29
A. Generalidades	29
B. Principales características	30
C. Esquemas de representación documentaria	32
1. Método del texto integral (<i>full text</i>)	33
2. Método de indización (<i>key word</i>)	33
D. Problemas a nivel gramatical	34
1. Sinonimias	34
2. Polisemias u homografias	34
3. Analogías	35
4. Antonimias	35
5. Formas sintácticas	35
6. Orden de términos	36
E. Instrumentos lingüísticos	36
1. El léxico	36
2. El <i>thesaurus</i>	36
F. Principales sistemas en operación	37
1. LEXIS	37
2. WESTLAW	37
3. AUTO-CITE	38
4. ITALGIURE FIND	38
5. IDG	38
6. CELEX	38
7. JURIS	38
8. PRODASEN	38
9. CREDOC	38
10. IRETIJ	38
11. CEDIJ	38
12. CRIDON	38
13. UNAM-JURE	38

V. Informática jurídica documentaria

A. GENERALIDADES

El reciente desarrollo de los sistemas de documentación automatizada corresponde a una realidad sensible en todos los campos del conocimiento: el crecimiento del volumen documentario se ha dado en proporciones tales, que los métodos tradicionales de búsqueda son hoy obsoletos. En el contexto jurídico, el fenómeno de la “inflación de textos” es en parte responsable de este incremento. Los textos de ley han dejado de ser generales para entrar en detalle y su cantidad, por lo tanto, ha ido en aumento, provocando una labor de legislación más pronunciada en los últimos veinte años.

Este tipo de informatización jurídica constituye un fenómeno irreversible que los profesionales del Derecho deberán considerar seriamente. La búsqueda automatizada es hoy en día, y en ciertos casos, más rentable que las “búsquedas intelectuales-manuales”.

La informática jurídica documentaria es el área más antigua de la informática jurídica; sus orígenes suelen asociarse a los trabajos de John Horty en la Universidad de Pittsburgh.²⁷

En los sistemas de informática jurídica documentaria se trata de crear un banco de datos jurídicos (o *corpus* jurídico documentario) relativo a cualquiera de las fuentes del derecho (menos la costumbre) a efecto de interrogarlo con base en criterios propios acordes a esa información y su relevancia jurídica.

Al principio los criterios jurídicos prevalecieron sobre los informáticos. Se tendía a reproducir el criterio de clasificación usado en los repertorios jurídicos como criterio de almacenamiento y recuperación de la información. Esto constituía una clara limitación ya que la máquina ofrecía muchas más posibilidades, y éstas permanecían generalmente subutilizadas. Es obvio que el

²⁷ Ver *supra*, capítulo IV, letra A.

cambio de los sistemas informáticos contribuyó en medida notable al mejoramiento del uso de los bancos de datos jurídicos.

El rendimiento de un banco de datos de esta naturaleza está en función de la exactitud y extensión de los datos contenidos y de los criterios de búsqueda (*retrieval*) de los documentos pertinentes. Para una fuente jurídica en evolución, como son la legislación y la jurisprudencia, la actualización se convierte en un imperioso menester.

La primera parte de la informática jurídica ha estado fundamentalmente ligada a la creación e interrogación de bancos de datos jurídicos a un grado tal, que es una creencia generalizada que la informática jurídica se agota en este campo.

Si la informática en general ha sido objeto de numerosas variaciones, la informática jurídica documentaria no ha sido la excepción. Al principio los criterios generales no eran muy diferentes de los usados en las técnicas de búsqueda bibliotecaria, y las categorías de clasificación jurídica seguían los modelos de las usadas por los juristas en los repertorios de papel.

Los primeros sistemas de interrogación de bancos de datos jurídicos fueron los sistemas *batch*, es decir, aquellos que permitían la búsqueda en los archivos (de texto y de palabras ordenadas alfabéticamente), que indicaban, una vez señalada una palabra, la "dirección" donde estaba almacenada en todos los documentos del archivo, comparando simplemente los archivos. La utilización de los operadores proposicionales permitía (y permite aún por ser una fase en uso) la combinación de palabras, a fin de tornar más específico el campo de los documentos buscados. Así por ejemplo, consultar las palabras "contrato o compra-venta", permiten recuperar todos los documentos que contengan las dos palabras, y si agregamos "hipoteca (no)" excluimos todos los documentos del primer conjunto que contienen la palabra "hipoteca".

De los sistemas *batch* se pasó a los sistemas *on line* que permiten la posibilidad de establecer interrelación a distancia a fin de precisar la pregunta mediante un diálogo entre el usuario y la máquina. Hoy en día los sistemas de interrogación son lo suficientemente sofisticados como para ayudar en forma considerable al interrogador, perfilándose un periodo de transición en la utilización de sistemas expertos a efecto de mejorar la calidad de la búsqueda y para hacer "aprender" a la máquina.

B. PRINCIPALES CARACTERÍSTICAS

La finalidad de la informatización en un sistema documentario consiste en encontrar lo más rápida y pertinentemente posible la información que ha sido almacenada. El conjunto de esas informaciones constituye el banco de datos o corpus (la expresión "base de datos" es por momentos reservada a la designación de subconjuntos del corpus total).

La interfase almacenamiento-recuperación de información jurídica provoca el siguiente proceso:

a) La entrada de documentos (leyes, reglamentos, jurisprudencia, doctrina, acuerdos, etcétera), se efectúa bajo una forma codificada comprensible para la máquina. Dependiendo de la capacidad del sistema la codificación de textos será más o menos voluminosa. Esta codificación será elemental si la entrada de información es hecha carácter por carácter (letra, cifra, puntuación, etcétera). La codificación será más compleja si el texto ha sido condensado previamente de tal forma que sólo contenga sus elementos característicos; cada elemento del texto (palabra o expresión, identificadores, etcétera) corresponde a un número de código.

b) La búsqueda de documentos significativos se efectúa de la manera siguiente: la codificación de palabras deseadas, según la combinación escogida, será introducida en la computadora que comparará toda la base documentaria y señalará una concurrencia importante cada vez que la comparación de los códigos puedan ser combinados gracias a los operadores *booleanos* (provenientes de la lógica de Boole) “y”, “o” y “salvo” de tal modo que la búsqueda gane en eficacia.

Todos los sistemas prevén la facultad para los usuarios de conocer el número de documentos ligados a cada pregunta a fin de restringir el campo de éste y obtener una información más precisa.

El principio general de un sistema jurídico automatizado es simple, pero su puesta en funcionamiento presenta varias dificultades, de entre las que se cuentan:

a) La reticencia de los profesionales del Derecho a este uso de la informática en función de múltiples razones sociológicamente fáciles de determinar. Los argumentos habituales externados por los juristas como el empobrecimiento del trabajo de análisis, sustitución de lo cuantitativo por lo cualitativo, riesgos de inutilidad o sustitución, etcétera son respetables, pero rebasan el cuadro técnico de este trabajo.

b) Los argumentos tendientes a denunciar los resultados relativos de estos sistemas son, por el contrario, más arbitrarios; sin embargo a este respecto cabe mencionar que toda búsqueda documentaria deja escapar documentos que pueden ser interesantes o, inversamente, conduce a documentos que no conciernen de manera directa a la cuestión estudiada y que va en función de la misma documentación informatizada. En este caso las informaciones inútiles constituyen un ruido²⁸ y las informaciones útiles que no han sido comunicadas constituyen un silencio.²⁹

²⁸ El ruido informático constituye la recuperación de gran cantidad de información distinta de la que es requerida. Es consecuencia directa del grado de análisis tan general propio de una indización.

²⁹ El silencio informático consiste en la imposibilidad de recuperar toda la información requerida sobre un punto específico. Es la falta de información pertinente en una consulta dada.

Una primera causa de silencio es propiamente contingente; los bancos de datos son raramente exhaustivos respecto a un sector jurídico: los documentos son objeto constante de una selección dictada para las facilidades de búsqueda. Son entonces extraños los sistemas que buscan sobre todo el material "bruto" y no se remiten en lo esencial a una primera elección efectuada por la edición jurídica (lo que de por sí podría presentar problemas en cuanto a la propiedad literaria). El corpus almacenado es inferior entonces al corpus real en una proporción más o menos importante.

Una causa de silencio más delicada de controlar es aquélla resultante de una función de análisis; el análisis de un texto trae aparejado un desglosamiento de palabras clave que permitirá dirigir la búsqueda y constituir un índice, introduciendo un factor importante de silencio: aquél de nociones no consideradas en el momento del análisis; así como un primer factor de ruido, una noción puesta en evidencia que no corresponde más que directamente al contenido del documento.

Una vez incorporada la documentación analizada, la concepción del sistema puede hacer aparecer nuevas fuentes de silencio o ruido de origen técnico (programación deficiente, etcétera).

Una última fuente de ruido o silencio surge de la inadecuada formulación de la pregunta por parte del usuario, pues no lo hace de conformidad exacta con las características del sistema.

C. ESQUEMAS DE REPRESENTACIÓN DOCUMENTARIA

El problema principal que se enfrenta en la constitución de un sistema de documentación jurídica automatizado está en función de la naturaleza híbrida del lenguaje jurídico. A diferencia de la mayoría de las otras disciplinas, el Derecho no utiliza un lenguaje "científico" al no ser puramente descriptivo o preciso.

De esta forma, el lenguaje jurídico no describe al Derecho como objeto, sino en gran medida al Derecho mismo; por otra parte, cada uno de los términos empleados no corresponde biunívocamente a una realidad o a un objeto: la ambivalencia estriba en la misma naturaleza profunda del Derecho, a lo cual difícilmente se presenta objeción.

Las palabras que integran los documentos jurídicos no pueden ser consideradas como unidades fundamentales con un sentido cada una. De esta forma, cada documento está caracterizado de manera única por las palabras utilizadas (esto a causa del estilo, de la sintaxis y del vocabulario jurídico), empleando esos términos en su sentido más general.

El estilo jurídico recurre con frecuencia a comparaciones o efectos que pueden hacer intervenir una noción análoga en el curso de un texto; pudiendo

también sugerir una noción o situación sin que el término que menciona con mayor frecuencia no aparezca realmente.

Las dificultades inherentes al estilo jurídico presentan dos métodos de almacenamiento de datos documentarios que analizaremos en forma específica.

1. Método del texto integral (*full text*)

Consiste en introducir los textos en memoria tal cuales, siendo integralmente almacenados en la computadora. Este método es invariablemente oneroso por el espacio de memoria; por otra parte, disminuye notablemente los riesgos de silencio, siendo por el contrario altamente generador de ruido.

2. Método de indización (*key word*)

Este método implica el análisis previo del documento para extraer y concretar las características fundamentales (conceptos, circunstancias, elementos de decisión, etcétera). Es costoso por la necesidad de personal calificado, pero disminuye teóricamente los riesgos de ruido. Cada documento es objeto de un resumen más o menos sintético que recibe el nombre de *abstract*, siendo éste un proceso informático-jurídico de tratamiento de información jurídica que tiene por objeto la recuperación y presentación de universos de información de manera automática, a partir de la elaboración de un soporte derivado en el que se plasman los sintagmas descriptores del soporte documentario de origen, relacionables sintagmáticamente dentro de unidades de ideas. Éstas son estructuradas en forma lógico-deductiva a través del empleo de restrictores de distancia que las hacen concordantes con módulos de interrogación. En función de la lógica de la computadora, actualmente se distinguen dos tipos de *abstracts* jurídicos fundamentales: el legislativo y el jurisprudencial.

Cabe expresar que tanto el método del texto integral como el de indización no son antitéticos más que en apariencia: la indización puede en ciertos casos derivar en un resumen que será introducido y explotado según los métodos del texto integral; es extraño que un texto sea introducido integralmente sin ser acompañado de descriptores que realizan una indización en el cuerpo del texto. La oposición entre los dos métodos está más marcada en las primeras experiencias, cuando los sistemas eran aún elementales. Más recientemente, es la naturaleza constitutiva del banco de datos la que determina la elección de una mayor o menor indización (ésta conviene más en la jurisprudencia o doctrina a diferencia del texto integral considerado como más adaptable a los textos legislativos).

La tendencia generalizada para los sistemas importantes es la de combinar las ventajas del texto integral (trabajo sobre los documentos jurídicos en sí mismos sin pérdida de información ni interpretación errónea, resultante de la vía de análisis) y aquellas derivadas de la indización, aumentando la eficacia de la búsqueda, de aquí que se mencione la existencia de un tercer método: el combinado, a fin de preservar el “espíritu” del legislador, juzgador o autor.

D. PROBLEMAS A NIVEL GRAMATICAL

Por otra parte, la sintaxis jurídica presenta dificultades a nivel gramatical, que constituyen serios problemas lingüísticos para la recuperación de información jurídica y que podemos agrupar en las siguientes manifestaciones:

1. Sinonimias

Se da cuando una idea se puede expresar con dos o más palabras diferentes, por ejemplo:

CEMENTERIO PANTEÓN CAMPOSANTO

Este fenómeno constituye un problema respecto de la recuperación, debido a que si se solicita información por la primera palabra y la computadora no nos provee los documentos que contienen también la segunda o la tercera, entonces la información es parcial, en este caso, el problema que se ha de resolver es el del silencio informático.

2. Polisemias u homografías

Este fenómeno se presenta cuando una palabra (sintagma autónomo) tiene dos o más sentidos. Por ejemplo, el término *CONSTITUCIÓN* puede aludir a una acepción jurídica equivalente a *Carta Magna* o bien referirse a la *formación de una sociedad*, dependiendo del contexto en que se use.

Las polisemias constituyen un problema para efectos de recuperación, ya que si se interroga a la computadora por la palabra de nuestro ejemplo, en su primer sentido (y éste no se puede distinguir del segundo), nos dará todos los soportes derivados en que se encuentre tal término, independientemente de la acepción que en él se emplee. Aquí el problema que debe resolverse es el del ruido informático.

3. Analogías

Encontramos este problema cuando dos o más palabras, provenientes de diferente raíz, se refieren a ideas parecidas, por ejemplo:

MORA RETARDO o TÉRMINO PLAZO

Este fenómeno constituye un problema, ya que es posible que la información requerida quede incompleta si no se relaciona la palabra por la que se interroga con sus análogos. En este caso debe resolverse el silencio informático.

4. Antonimias

Se da entre una palabra y otra con sentido opuesto, por ejemplo:

CONSTITUCIONAL INCONSTITUCIONAL

Es relevante para efectos de recuperación debido a que una palabra precedida o seguida de una negación es sinónimo del sintagma autónomo que sin dicha negación sería su antónimo, por ejemplo:

NO CONSTITUCIONAL = INCONSTITUCIONAL

De esta manera, a alguien que solicita información por la palabra *inconstitucional* también le interesa recuperar aquella en la que se alude a *no constitucional*.

El problema a que da lugar la antonimia es el silencio informático.

5. Formas sintácticas

Consisten en la manera en que una misma idea se puede expresar a través de diversas relaciones sintagmáticas, por ejemplo:

CREACIÓN DE UNA EMPRESA
 UNA EMPRESA SE CREÓ
 UNA EMPRESA SERÁ CREADA
 UNA EMPRESA CREADA
 UNA EMPRESA SE CREARÁ

Los citados son sintagmas equivalentes en caso de una interrogación, por lo que, a efecto de no perder información, es preciso relacionar todas las variaciones de la palabra "creación" que pueden aparecer en diferentes redacciones y bajo un tipo específico de interrogación, y que representan consecuencias jurídicas distintas.

6. Orden de términos

En el caso de modificación del orden de los términos, podemos encontrar un cambio en el sentido, por ejemplo:

ACCIÓN DE ENRIQUECIMIENTO ILEGÍTIMO
ILEGÍTIMA ACCIÓN DE ENRIQUECIMIENTO
ILEGÍTIMO ENRIQUECIMIENTO EN LA ACCIÓN

Esto sin duda se desprende de un problema de ruido informático.

E. INSTRUMENTOS LINGÜÍSTICOS

Para resolver estos y otros problemas lingüísticos existen dos instrumentos fundamentales con el objetivo de una apropiada recuperación de información por ideas, independientemente de la forma en que se expresen. Éstos son: el léxico y el *thesaurus*.

1. El léxico

Consiste en la organización de todas las palabras con contenido informativo,³⁰ almacenadas en computadora por nociones y subnociones, bajo un criterio morfológico semántico con la finalidad de resolver los problemas de la sinonimia y ayudarnos a resolver los de homografía o polisemia.

2. El *thesaurus*

Éste por su parte pretende servir de medio de apoyo al usuario para resolver los problemas de analogía, antonimia y aislamiento semántico de las palabras polisémicas al momento de la interrogación.

³⁰ A las palabras carentes de contenido informativo se les denomina "palabras nulas", como lo son, por ejemplo: para, de, a, etcétera.

Se construye a partir de las nociones del léxico y su connotación es distinta a la empleada en los términos de biblioteconomía (lista de descriptores propia de los sistemas de lenguaje cerrado).

Por lo expuesto anteriormente podemos decir que la especificidad marcada del lenguaje jurídico convertiría en inoperante todo sistema documentario si no estuviera integrado a éste y en combinación con los problemas de búsqueda y conversación con los usuarios.

De esta forma, el *thesaurus* se convierte en un "léxico jerarquizado que comprende una red de interconexiones, exclusiones, discriminaciones y proximidades semánticas bajo la forma de listas de sustitutivos de contrarios, términos vecinos o genéricos, etcétera". A cada interrogación del *corpus*, el *thesaurus* orienta la explotación a fin de disminuir automáticamente en la conversación con el usuario los problemas de ruido y silencio.

Sus funciones principales son las siguientes:

a) Como diccionario analógico en su función de conexión permitiendo reconocer situaciones y expresiones jurídicas, tomando en cuenta la sintaxis gramatical, buscando antónimos y reagrupando sinónimos.

b) Como diccionario analítico en su función de discriminación excluyendo nociones afines no necesarias o incidentales, así como en las formas no deseadas, disminuyendo las figuras polisémicas y sinónimas.

c) Como índice en su función de adición de términos que conforman la base de datos o *corpus*.

Los *thesaurus* pueden ser abiertos o cerrados, dependiendo de que se pueda o no agregar a ellos nuevos elementos.

F. PRINCIPALES SISTEMAS EN OPERACIÓN

Existen en el mundo varios sistemas en operación a nivel de informática jurídica documentaria. Sin pretender ser exhaustivos, enumeraremos algunos de ellos resaltando en su caso sus características primordiales.

1) LEXIS de la compañía Mead Data Central (Estados Unidos). Entró en funciones desde 1973 y se caracteriza por tener el banco de datos jurídicos más grande en el mundo con casi treinta billones de caracteres cargados y con una entrada anual en memoria de alrededor de 35 000 sentencias estructuradas en texto integral. Actualmente el sistema LEXIS no sólo opera en los Estados Unidos, sino también en Inglaterra y Francia, donde se han almacenado las resoluciones del Consejo de Estado.

2) WESTLAW de la West Publishing Company (Estados Unidos). Casa editorial especializada en textos jurídicos que ha tomado de la sociedad canadiense QL/Systems, el programa necesario para hacer funcionar su banco de datos desde 1975.

3) AUTO-CITE de la Lawyer's Cooperative Publishing Company (Estados Unidos). Desde 1970 brinda a los abogados norteamericanos la posibilidad de obtener en forma automática las citas de jurisprudencia, reconstruidas históricamente de un archivo automatizado que contiene más de cuatro millones de referencias.

4) ITALGIURE FIND de la Suprema Corte de Casación de Italia. Este sistema cuenta con casi cuatro billones de caracteres almacenados en 24 archivos en materia legislativa, jurisprudencial y doctrinaria.

5) IDG del Instituto para la Documentación Jurídica de Florencia, Italia. Dispone de un extenso banco de datos jurídicos (fundamentalmente bibliográficos), constituyendo un apoyo importante para los sistemas italianos de la Corte y la Cámara de Diputados.

6) CELEX de la Comunidad Económica Europea con sede en Bruselas, Bélgica. Contiene el Derecho comunitario, tratados, acuerdos, Jurisprudencia de la Corte de Justicia y decisiones del Parlamento Europeo, en una serie de archivos consultables en los idiomas oficiales de dicha comunidad.

7) JURIS del Ministerio de Justicia (Alemania). Con un banco de datos jurídicos en materia de Derecho fiscal y protección social, con su programa Golem-Passam y agregados interesantes en el tratamiento lingüístico de los vocablos contenidos en el banco denominado Parat.

8) PRODASEN de la Cámara de Senadores (Brasil). Con sistema stairs en uso y una serie de archivos en materia legislativa, doctrinal y jurisprudencial, así como una base de datos sobre proyectos de ley y otros más.

9) CREDOC de la Federación Real de Abogados y Notarios (Bélgica). Que funciona desde 1969 (el más antiguo en Europa) con una vasta red de datos de interés jurídico y un *thesaurus* bilingüe (francés y flamenco).

10) IRETIJ del Instituto de Investigación y Estudio para el Tratamiento de la Información Jurídica (Francia). Con un banco de datos jurisprudencial y doctrinario apoyado en un programa especialmente concebido para tal efecto y con interesantes resultados a pesar de sus limitaciones presupuestarias.

11) CEDIJ del Consejo de Estado y el Tribunal Supremo (Francia). Con un sistema propio de interrogación denominado Docilis y con un banco de datos en *full text* con las normas legislativas y de jurisprudencia, en particular de las decisiones del Consejo Constitucional, el Tribunal Supremo y el Consejo de Estado.

12) CRIDON del Centro de Investigación en Materia de Información y Documentación Notarial (Francia). Con una extensa base de datos en materia notarial con un sistema propio de investigación denominada Sydoni.

13) UNAM-JURE del Instituto de Investigaciones Jurídicas de la UNAM (México). Con un banco de datos en materia legislativa a nivel federal, estatal y municipal, así como el sistema VALLARTA en materia jurisprudencial.

Hay otros sistemas importantes como el Juris-Data (Francia), Status (Inglaterra), Noris (Noruega), Find (Argentina), Sinade (España), Datum (Ca-

nadá) y otros más de carácter operativo además de otros tantos en fase experimental que pronto estarán en posibilidad de dar servicio al público.

Como podemos percatarnos, la informática jurídica documentaria ha alcanzado niveles muy significativos que seguramente se verán incrementados por la difusión de las microcomputadoras y computadora personales que permitirán que prácticamente cada operador jurídico (funcionario, legislador, juez, abogado, notario, estudiante, etcétera), tenga su propio y pequeño banco de datos adaptado a sus exigencias particulares, por lo que difícilmente podemos prever un límite dentro del desarrollo de esta nueva herramienta para el Derecho.