

CAPITULO TERCERO

LA DOCUMENTACION JURIDICA AUTOMATIZADA

A. Aspectos generales

1. ALGUNAS NOCIONES PREVIAS

a) *Información y datos*

Se ha hablado en páginas anteriores (véase, *supra*, II, 2, b) que la sociedad contemporánea ha sido denominada “sociedad de la información” porque en ella la información adquiere un papel central en el funcionamiento de todas las actividades humanas. Pero, ¿qué es la información?

La teoría matemática de la información (y de la comunicación) considera a ésta como entropía negativa, esto es, como medida del orden o de la organización.¹ Se ocupa llanamente de las condiciones y limitaciones para la transmisión de mensajes a través de diversos canales, con independencia del significado de estos mensajes.² Este concepto de información tiene gran importancia para diversas ramas de la ciencia contemporánea, como la biología.³

En el ámbito de los fenómenos humanos, se puede entender la información como “un conjunto de mecanismos que permiten al individuo reformar y organizar los datos del medio para que, estructurados de una manera determinada, le sirvan de guía de acción”.⁴ O, en palabras de Marcos Kaplan, “en las sociedades humanas, la información es todo lo que permite el control, el mando, la conformación y organización: reglas, normas, prohibiciones, saber qué o conocimiento, saber cómo o técnicas, medios de masas”.⁵

¹ Liliensfeld, *Teoría de sistemas* [40], pp. 102 y ss.

² *Id.*, p. 104.

³ Bertalanffy, *Teoría general de los sistemas* [1], pp. 57 y ss.

⁴ López Ayllón, *El derecho a la información* [42], p. 36.

⁵ Kaplan, *Estado y sociedad* [37], p. 43.

Dentro de la función única que es la transferencia del conocimiento, la información cumple dos etapas:

- Es primeramente expresión, formulación (formal) destinada a hacer comunicable un mensaje.
- Es comunicación, transmisión activa a través de los signos adecuados para llevar un mensaje al otro.⁶

Recapitulando, la información en la vida social reviste las siguientes características:

- Es una medida de orden u organización.
- En cuanto estructuración de los datos del medio, permite la orientación de la acción. En este sentido, se encuentra estrechamente ligada a un contexto social específico.⁷
- Es expresión y formulación de ideas, conocimientos destinados a ser transmitidos como mensaje. Así, información y comunicación se implican mutuamente, aunque no sean lo mismo: la información es el contenido de la comunicación.⁸

Del concepto general de información puede deducirse que el dato es un signo simple, no estructurado con otros, de un hecho de la realidad. El dato es una unidad o elemento de información con sentido propio pero no se confunde con ésta: la información es genérica; el dato, singular, un hecho “en bruto”.⁹

b) *Documento e informática*

La primera imagen que viene a la mente cuando se piensa en un “documento” es la de un pedazo de papel en el que se asientan, en forma escrita, determinados hechos o se expresa un conjunto de ideas. Sin embargo, la noción de documento es más amplia, pues incluye fotografías, dibujos, planos, etc.; existen incluso documentos acústicos.

Entre los juristas, han sido los cultivadores del derecho procesal quienes se han ocupado más del concepto de documento, por su relevancia como elemento probatorio. De algunas de las definiciones que nos proporcionan¹⁰ se desprende primeramente como noción central la de “representación” de un hecho, idea, pensamiento, mediante los signos o símbolos apropiados y, en segundo lugar, que esa representación se asienta en una cosa, en algo material o soporte.

⁶ Catala, “Ebauche d'une theorie juridique de l'information” [83], p. 17.

⁷ Cfr. López Ayllón, *op. cit.*, [42], p. 36.

⁸ *Id.*

⁹ Véase Burch Jr., y Strater Jr., *Sistemas de información* [7], pp. 43 y ss.

¹⁰ Cfr. por ejemplo Chioyenda, *Instituciones de derecho procesal civil* [12], p. 265 y Devis Echañda, *Compendio de pruebas judiciales* [16], p. 491.

La informática ha venido a modificar este concepto, en al menos, dos sentidos:

— Resalta el hecho de que la representación significativa que constituye el documento es independiente, o no se identifica con el soporte material que lo contiene.

La información codificada en una cinta o disco magnético no es inmediatamente perceptible como “documento” si no es hecha visible en una pantalla mediante ciertos procesos electrónicos.

— Cambia también la imagen del documento como algo inmutable, cerrado, ya definido. Las capacidades propias de la computadora permiten “componer” o “definir” los documentos a través de la combinación *ad-hoc* de diversos elementos o unidades de información, con lo cual el documento se convierte en variable de una necesidad específica de información.

Por lo anteriormente dicho, un concepto de documento definido en función de la informática podrá ser el siguiente: toda representación objetiva y hecha perceptible mediante una computadora, de un hecho, idea, conocimiento a través de signos, símbolos o imágenes idóneas, que constituye una unidad de significado respecto de una necesidad particular de información.

c) *Sistemas de información*

Por sistema de información puede entenderse un “sistema utilizado para almacenar elementos de información que requieren ser procesados, buscados, recuperados y distribuidos a un conjunto de usuarios”.¹¹ Normalmente estos sistemas de información se emplean para manejar registros bibliográficos y textos.¹² Sin embargo, la noción de “sistema de información” es genérica y actualmente comprende diversos tipos de sistemas computarizados.

Estos tipos de sistemas son básicamente los siguientes:¹³

— Sistemas de información documental (*Information Retrieval Systems*). Como su nombre lo indica, estos sistemas procesan información consistente en documentos, esto es, se trata de la representación, almacenamiento y acceso a documentos o substitutos de documentos (véase, *infra*, 2).

— Sistemas de bases de datos. La información que manejan estos sistemas son datos estructurados en forma de tablas. Los “registros” tienen

¹¹ Salton, y McGill, *Introduction to Modern Information Retrieval* [59], p. XI.

¹² La expresión comúnmente utilizada en inglés es “*Information Retrieval Systems*”.

¹³ La clasificación está tomada de Salton, *Introduction*. . . [59], pp. 7-10, 354 y ss.

siempre la misma cantidad y clase de información y ésta sólo puede ser expresada de un modo. Los datos se caracterizan por un número limitado y definido de atributos y cada atributo sólo puede tener un valor previamente determinado (clasificación monotética). (Por ejemplo, una base de datos sobre el personal de una empresa, en la que se determine el nombre, edad, estado civil, salario, categoría, etcétera). Las características de las bases de datos permiten combinar la información y realizar con ella comparaciones lógico-matemáticas, operaciones de conjuntos y cálculos estadísticos.¹⁴

— Sistemas de información para la administración (*management information systems*) y de ayuda a la decisión. Estas categorías pueden considerarse un tanto artificiales, porque todo sistema de información es capaz de apoyar de algún modo la toma de decisiones. Sin embargo, la clasificación quiere indicar la posibilidad de incorporar a los sistemas mecanismos específicos para la toma de decisiones: generación de estadísticas, evaluación de alternativas, elaboración de gráficas, etcétera.

— Sistemas de “preguntas-respuestas” (*question-answering systems*). Estos sistemas pueden actuar en el ambiente del lenguaje natural, pero no ofrecen como respuesta un documento, sino datos específicos, de acuerdo con el conocimiento que tengan almacenado.

2. LOS SISTEMAS DE INFORMACION DOCUMENTAL

a) *Concepto*

“Todo sistema de información documental (*information retrieval system*) puede describirse como un conjunto de elementos de información (DOCS), un conjunto de consultas (CONS) y algún mecanismo (SIMILAR) que determine cuál elemento de información cumple con los requisitos de la consulta”.¹⁵ Si bien en teoría la relación entre los documentos y las consultas puede obtenerse mediante comparación directa, en la práctica los documentos o elementos de información son convertidos a una forma especial utilizando un lenguaje de clasificación o indexación (LENG), de acuerdo con el diagrama de la página siguiente.¹⁶

Para expresarlo de otra manera: “todo sistema documental es la representación (R) de un corpus de textos (T)”; el problema consiste en “la

¹⁴ Existe confusión entre los conceptos “base de datos” y “banco de información”, que frecuentemente se utilizan como términos sinónimos o intercambiables, no obstante existir diferencias de estructura entre ambos. Véase Salton, *Introduction*. . . [59], p. 357 y la tabla comparativa en Rijsbergen, *Information Retrieval* [56], p. 2.

¹⁵ Salton, *Introduction*. . . [59], p. 10.

¹⁶ *Id.*, p. 11.

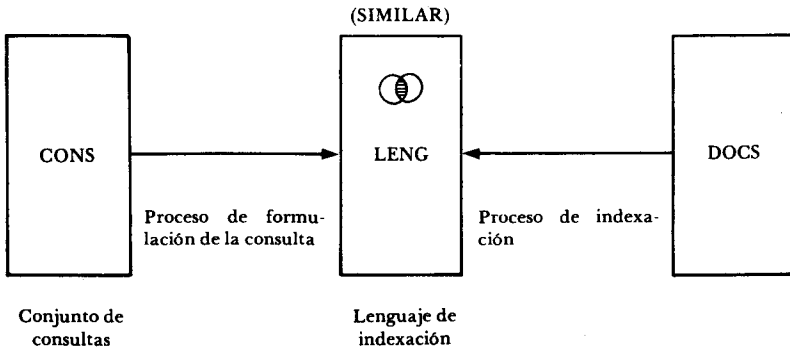
forma en que se pueda pasar de R a T e inversamente”,¹⁷ o más sencillamente, en hacer coincidir el lenguaje de una consulta dada con el lenguaje de los documentos.¹⁸ Todo sistema de información documental emplea diversas técnicas y procedimientos para resolver este problema central.

Volviendo a la figura anteriormente dibujada, podemos decir que representa un sistema documental en que el “sistema” se compone solamente de los elementos CONS, LENG y SIMILAR, DOCS y que constituyen un ámbito objetivo, un instrumento para el usuario, pero también podemos considerar al sistema documental desde una perspectiva más amplia y conceptualizarlo como una organización de comunicación que incluye al usuario como parte del sistema, en el cual se reproducen entonces los principios que regulan la comunicación entre emisores y receptores, correspondiente a un esquema usual de teoría de la comunicación.¹⁹

Se puede completar la figura de la manera ilustrada en la próxima página.²⁰

SISTEMA DOCUMENTAL DESDE EL PUNTO DE VISTA FUNCIONAL

Proceso de determinación de la semejanza de la representación de la consulta con la representación de los elementos de información



¹⁷ Bourcier, "Information et signification en droit" [79], p. 12.

¹⁸ Elmi, *Corso...* [19], p. 124.

¹⁹ Fiedler, "Strukturen und Wirkungen juristischer Informationssysteme in der Bundesrepublik Deutschland" [92], p. 146.

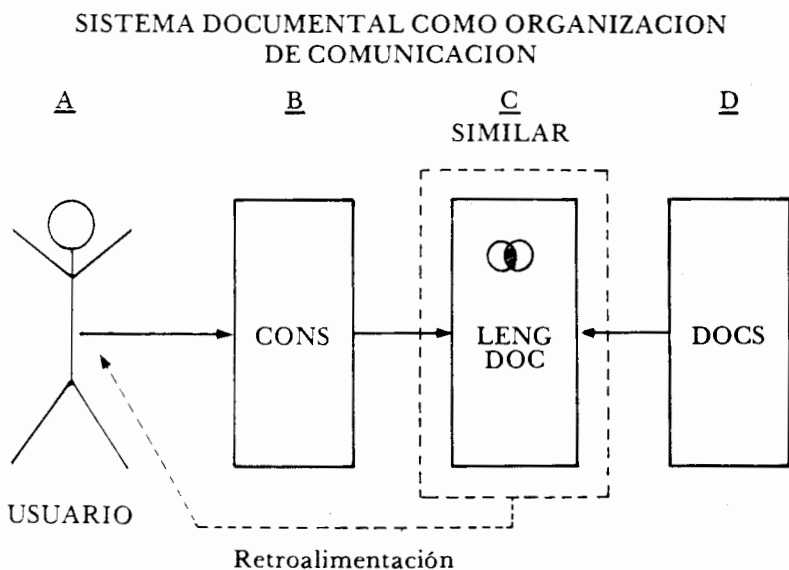
²⁰ *Cfr.* el esquema reproducido en Rijsbergen, *Information Retrieval* [56], p. 7.

Una vez ampliado el esquema, podemos examinar en una segunda aproximación las técnicas y procedimientos, tanto automáticos como manuales, que no tienen otra finalidad más que hacer coincidir los lenguajes de consulta y documental.

Se puede resolver el problema de esta coincidencia, en buena medida, controlando y definiendo previamente el lenguaje de indexación (C), ya sea tomándolo o no de los documentos fuente (D). En el caso de un lenguaje cerrado y previamente definido, se reduce o elimina el esfuerzo del usuario en la consulta (véase, *infra*, 3, b, ii).

Si se opta por no controlar totalmente el lenguaje, o no hacerlo en lo absoluto, se puede dejar el problema en manos del usuario (A), quien deberá hacer "suposiciones" sobre el lenguaje documental (C), tomando en cuenta que puede aprender de sus respuestas y afinar sus consultas (B) (véase, *infra*, 3, b, i).

Sin embargo, existe también la posibilidad de desarrollar ciertos instrumentos a partir del lenguaje documental (C) que incidan sobre B y muestren al usuario las relaciones semánticas y de otro tipo que se dan entre los conceptos contenidos en este lenguaje; tal es, por ejemplo, el caso de los *thesaurus* (véase, *infra*, 3, d).



Finalmente, la informática tiene a su disposición un arsenal de técnicas y procedimientos propios, que forman parte del mecanismo SIMILAR (C) y que, por diversas vías, contribuyen a solucionar el problema de la consulta y la relevancia de las respuestas que obtiene el usuario (véase, *infra*, 3, c).

Todo sistema documental hace uso, en realidad, de una mezcla de las opciones básicas anteriormente expuestas, en proporciones y con resultados diversos; de hecho, estas opciones se encuentran estrechamente relacionadas y su punto de referencia se encuentra en el mecanismo SIMILAR: la computadora.

b) *Sobre el problema lingüístico, en particular*

En última instancia, la problemática a la que se acaba de hacer referencia gira en torno del lenguaje. La cuestión lingüística, a su vez, parte de dos factores interrelacionados:

— La palabra como unidad característica de los sistemas documentales.

— La riqueza en la expresión de los conceptos a través del lenguaje (mayor número de significantes que de significados) y las relaciones que éstos guardan entre sí.

i) De acuerdo con Danièle Bourcier,²¹ las características principales de los sistemas de documentación automatizada en la actualidad son las siguientes:

— La palabra es la unidad característica (archivo de inversión).

— Estas unidades son clasificadas en palabras nulas (o funcionales) y palabras significativas, de acuerdo con su categoría gramatical (adverbio, preposición) y su frecuencia de uso en el idioma (verbos ser y haber).

— Toda palabra puede reconducirse a una forma fundamental (masculino, infinito).

— Estas unidades establecen entre sí una relación de concurrencia o de proximidad que determina su relación semántica o sintáctica.

Así, por ejemplo, el concepto "SEGURIDAD SOCIAL" puede ser "determinable como unidad de sentido, pero más allá de la adyacencia, dos entidades lingüísticas establecen entre sí una relación de significado sólo si pertenecen al mismo conjunto: la frase, el párrafo, el documento"²² (véase, *infra*, 3, c, ii).

Vale la pena detenerse un poco más en la cuestión de la palabra como unidad propia de los sistemas documentales.

²¹ Bourcier, "Information et signification en droit" [79], p. 12.

²² *Id.*

Por “palabra”, en lo relativo a su utilización por la computadora, debe entenderse una cadena finita de caracteres alfabéticos, numéricos y alfanuméricos que pueden definirse y tratarse en forma convencional (por ejemplo, raíces, sufijos y prefijos de las palabras) y que pueden coincidir o no con la palabra en sentido gramatical. Esta coincidencia se da, habitualmente, en la medida que se trate de unidades lexicales con sentido.

Sin embargo, la computadora no “entiende” o “comprende” las palabras, no al menos en el sentido en que pueda “comprenderlas” un ser humano; solamente las “reconoce” por su coincidencia con un patrón de reconocimiento que le haya sido previamente suministrado (por ejemplo, puede hacerse que reconozca las diferencias de género y número de los vocablos).

Esto no constituye el final del problema, en la medida en que el lenguaje no se compone de palabras independientes, sino, en la mayoría de los casos, de otras unidades, integradas por varias palabras y que se denominan sintagmas, cuyo significado puede ser distinto al de sus elementos.²³ En otros términos, las palabras adquieren sentido en un contexto, y rara vez se hallan aisladas. De esta forma, se plantea como reto a todo sistema documental el paso del nivel de reconocimiento de las *palabras* al de los *conceptos*. Para ello se han empleado diversas técnicas, con resultados variables y no siempre satisfactorios.

ii) Hace apenas poco más de un siglo que se iniciaron amplia y formalmente los estudios sobre el lenguaje humano y sus conexiones con la lógica y el pensamiento; surgieron así, como disciplinas autónomas, la lingüística y la filosofía del lenguaje.²⁴

Se puede definir provisionalmente al lenguaje como “cualquier sistema de signos simbólicos empleados para la intercomunicación social. . . cualquier sistema de signos que sirva para expresar y comunicar ideas y sentimientos, o sea, contenidos de la conciencia”.²⁵

Como todo fenómeno social, el lenguaje comporta distintos aspectos; en consecuencia, ha sido estudiado y valorado desde diversos ángulos y las conclusiones han sido igualmente divergentes.

Así, por ejemplo, algunos se han centrado en el aspecto “lógico” del lenguaje, es decir, la comunicación y las estructuras simbólicas. Otros parten de supuestos psicologistas, esto es, consideran al lenguaje como conjunto de hechos meramente psíquicos.²⁶ Mientras algunos han puesto

²³ Por ejemplo, CONSTITUCION POLITICA DE LOS ESTADOS UNIDOS MEXICANOS tiene un significado distinto al del sentido aislado de las palabras que lo constituyen.

²⁴ Coseriu, *Introducción a la lingüística* [11], pp. 10 y ss.

²⁵ *Ib.*, p. 10.

²⁶ *Ibid.*, p. 38.

énfasis en el carácter individual del lenguaje, otros han recalcado su impronta social.²⁷

No obstante, estas posiciones son parciales en la medida que acentúan unilateralmente algún aspecto del lenguaje. Una comprensión más amplia del fenómeno lingüístico tendría que ser complementada, además, por su adecuada consideración como fenómeno histórico y cultural.²⁸

El primer y más importante aspecto del lenguaje es su “no exactitud”, su “indeterminación”. Esto significa que el lenguaje como tal no contiene todos los elementos que le permitan realizar una comunicación completamente unívoca y no sujeta a interpretaciones. Esto es así porque toda comunicación reposa sobre supuestos implícitos;²⁹ el destinatario desempeña un papel activo que asigna significados, interpreta, en una palabra. El significado depende en buena medida, del contexto o campo lingüístico, así como de la situación dentro de la cual se desarrolla la comunicación.

La “indeterminación” que afecta al lenguaje tiene varios grados o clases, que van de la polisemia, a la inconsistencia y su significado en diferentes contextos.

Algunos de los problemas habituales que enfrenta un sistema documental en relación con el lenguaje, y que pueden ser mitigados pero no eliminados del todo, son los siguientes:

— La polisemia, consistente en la asignación de varios significados a una misma palabra o grupo de palabras.

— La sinonimia, es decir, la existencia de varias palabras o términos cuyo significado es igual. La sinonimia puede ser absoluta (muy rara) o relativa, es decir, limitada a un cierto contexto.

— La analogía, es decir, la relación entre palabras o términos que comparten algún núcleo de significado (por ejemplo, vehículo y automóvil). Se trata de un *continuum* de significados en que la relación es más próxima o más lejana.

— La vaguedad, es decir, la falta de precisión y rigor en la formulación de las ideas.

— La hiponimia, es decir, la inclusión del significado de un término en otro más extenso (relación de género a especie).

Algunos de estos problemas adoptan características particulares respecto de un lenguaje determinado, como lo es el del derecho (véase, *infra*, 4, c. i).

²⁷ *Ibid.*, p. 43.

²⁸ *Ibid.*, p. 45 y ss.

²⁹ Haba, “Apuntes sobre el lenguaje jurídico” (I) [96], pp. 23 y s.

3. LAS TÉCNICAS DE TRATAMIENTO Y RECUPERACION DE INFORMACION DOCUMENTAL

a) *El control del lenguaje*

Existe una cierta confusión, o más bien, inconsistencia en la utilización de términos como “lenguaje documental” y “lenguaje cerrado”; “lenguaje abierto” y “lenguaje natural”; “indexación”,³⁰ *full-text*, *free-text*. Es preferible hablar de “lenguaje controlado” y de “lenguaje no controlado”, tomando en cuenta que estos conceptos constituyen los polos de un *continuum* que comprende varias clases y grados. Sin embargo, el lenguaje no controlado, que comúnmente se identifica con el lenguaje natural, es más difícil de tratar desde el punto de vista de una mayor eficiencia en las respuestas, por lo que casi todo sistema documental emplea alguna o varias técnicas de control del lenguaje; la diferencia se encuentra en el punto o fase en que se aplique predominantemente este control.

En este apartado se tratarán las técnicas de control del lenguaje en la fase de indexación, esto es, las técnicas de representación documental en cuanto esquemas “puros”, así como sus ventajas e inconvenientes.

b) *La representación documental*

Por representación documental debe entenderse al conjunto de técnicas o esquemas mediante los cuales se obtiene, a partir de un documento o fuente de información originales, el documento o los elementos de información que serán objeto de tratamiento informático. Al procedimiento de obtención de esta información se le denomina genéricamente *indexación*, y el lenguaje que “representa” a la información o documento fuente, *lenguaje documental* (véase, *supra*, 2, a).

i) *el full-text*

El *full-text* consiste en la representación del documento fuente u original mediante todas las palabras significativas contenidas en él, es decir, el documento es recuperable mediante todas las palabras del texto excepto las llamadas “nulas”. Esto equivale al empleo de un “*thesaurus* negativo”³¹ y a la ausencia de casi todo tratamiento del lenguaje en la generación del documento que va a informatizarse.

³⁰ En realidad “indexación” o “indización” deben significar la generación de un índice, más que el procedimiento de asignación de descriptores a un documento.

³¹ Elmi, *Corso*. . . [19], p. 97.

Las principales ventajas de este esquema son las siguientes:

– Proporciona al usuario no sólo la indicación de la localización del texto, sino el texto mismo, que aparece en pantalla como producto de la consulta y es susceptible de imprimirse en un equipo periférico.

– Objetividad en la selección e indexación de los conceptos interrogables.

– Nula necesidad de un trabajo manual de elaboración del documento, pues éste se indexa por sí mismo

Algunos inconvenientes son los siguientes:

– Gran espacio de memoria para el almacenamiento de los textos.

– No existe una definición clara, o ésta es arbitraria, de la unidad-documento (véase, *infra*, 4, a, ii).

– No hay explicitación de conceptos.

– Alta posibilidad de “ruido”,³² cuando los términos o conceptos del documento no van calificados.

– No uniformidad en la indexación de las palabras y conceptos, cuando el lenguaje de los documentos fuente es inconsistente (por ejemplo, el lenguaje del legislador).

– Como consecuencia de lo anterior, la necesidad de elaboración de instrumentos lingüísticos que controlen algunos aspectos del lenguaje y auxilien la consulta.

ii) análisis de contenido

Todo análisis, en cuanto técnica de representación documental, conlleva una intervención humana previa en la selección y reformulación de la información del documento-fuente.

a) mediante lenguaje cerrado o palabras-clave

Esta técnica consiste en representar el contenido del documento-fuente mediante una serie de palabras-clave (key-words) que se toman de una lista previamente elaborada, cerrada y obligatoria (“*thesaurus* de clasificación positivo”).

Las ventajas de este esquema residen en:

– Utilización de un menor espacio de memoria.

– Control absoluto del lenguaje.

– Posibilidad de explicitación de conceptos.

– Clara determinación de la unidad-documento.

³² En informática se entiende por “ruido” la obtención de información que no es relevante respecto de una consulta; “silencio” es el fenómeno contrario: no se obtiene toda la información que contiene un sistema y que es relevante para una interrogación.

Como inconvenientes no pueden dejar de mencionarse:

- El “silencio”.
- La necesidad de contar con un analista especializado.
- La subjetividad e inconsistencia* en la asignación de los descriptores.

Esta técnica es la más tradicional, pues no es más que la aplicación informatizada de los sistemas manuales de catalogación bibliohemerográfica, que ha quedado reducido actualmente a la indexación de libros.

b') mediante lenguaje abierto

Mediante la expresión “lenguaje abierto” se quiere decir que no existe una lista previa, cerrada y obligatoria de descriptores que deba utilizar el analista para indexar el documento, lo cual no significa, por otro lado, que no puedan aplicarse otros controles del lenguaje al proceso de obtención del documento que se va a informatizar.

El análisis mediante lenguaje abierto constituye en realidad una mezcla de las características del *full-text* y del análisis mediante palabras-clave, cuyos inconvenientes trata de superar.

Las características positivas que definirán a esta clase de análisis son:

- Normalización del lenguaje.
- Posibilidad de explicitar conceptos.
- Posibilidad de definición objetiva y técnica de la unidad-documento.
- Reducción del texto (evitando repeticiones y redundancias).
- Reorganización de la información.
- Menor ruido y silencio.

Los inconvenientes principales son los que se mencionan a continuación:

- Necesidad de un analista especializado.
- Subjetividad e inconsistencia en el análisis.
- Necesidad de instrumentos lingüísticos para la consulta.

c) las “técnicas informáticas”

Por tales se han de entender aquellas que derivan del uso o intermediación de la capacidad propias de la computadora. A continuación se explican las principales utilizadas en sistemas de información documental.

i) archivos de texto y archivos de inversión³³

Casi todos los sistemas de información documental utilizan el esquema de archivos de texto y archivo invertido (*inverted file*) o de inversión.

El archivo de textos contiene los documentos que son el resultado del proceso de indexación, esto es, la representación documental de los documentos-fuente.

El archivo de inversión es un índice de las palabras no nulas contenidas en los documentos de los archivos de texto, junto con su dirección en dichos documentos. La inversión consiste en que, de la misma manera en que un documento "x" apunta o permite identificar las palabras que lo representan, en el archivo de inversión son las palabras las que apuntan hacia los documentos que las contienen.

La figura de la página hace visible esta idea.

De esta manera, la computadora recupera los documentos a partir de la dirección de las palabras en el archivo de inversión, sin tener que recurrir a la lectura del archivo de textos.

ii) La unidad-documento

Uno de los problemas que enfrenta la recuperación automatizada de información documental consiste en la definición de la "unidad-documento", es decir, aquella parte del documento (en ocasiones se trata del documento completo) dentro de la cual deben encontrarse las palabras que forman parte de una consulta, pues se supone que las concurrencias de las palabras tienen sentido sólo si su combinación aparece dentro de partes definidas y limitadas de un texto.

ARCHIVO DIRECTO				ARCHIVO INVERTIDO				
Documentos				Descriptoros				
				a	b	c	d	e
D e s c r i p t o r e s	1	2	3	X	X			
					X			
	c			X			X	
	d				X		X	
	e			X			X	
				D o c u m e n t o s	1	2	3	
				X		X		X
				X	X		X	
						X		X

³³ Véase Salton, *Introduction*. . . [59], p. 16 y ss.

La definición de una o varias unidades-documento dentro de un texto tiene por objeto evitar el ruido que se produce cuando hay “combinaciones no relevantes”, es decir, que las palabras consultadas se encuentren en el texto, pero su significado no sea equivalente al contenido de la consulta.³⁴ Así, mientras más estricta y objetiva sea la definición de la unidad-documento, habrá menos posibilidades de ruido.

Por lo general se define a la o las unidades-documento agregando a la representación documental signos o símbolos ortográficos previamente determinados para tal fin. En ocasiones sirven como tales los signos redaccionales comunes de un texto, tales como puntos, comas, etcétera.

Los operadores booleanos y la llamada “restricción de distancia” de los datos son instrumentos complementarios de la técnica de la unidad-documento (véase, *infra*, iv y v).

iii) campos de longitud fija y variable

La información de un documento puede estar contenida en campos de longitud fija o variable.

Son campos de longitud fija aquella organización de datos que se encuentra situada siempre en la misma posición relativa del documento y que contiene siempre la misma clase de información.³⁵ Los campos de longitud fija contienen habitualmente los datos formales, externos y objetivos de un documento fuente, tales como fecha, procedencia geográfica, emisor del documento, que pueden utilizarse en forma codificada.

Los campos de longitud fija, por su naturaleza (organización monotética) pueden ser utilizados para la recuperación mediante operadores relacionales,³⁶ esto es, pueden ser manejados como una base de datos (véase, *supra*, 1, c).

En cambio, son campos de longitud variable o libres aquella organización de datos que puede encontrarse en cualquier parte del documento e incorporar texto libre (descriptores).

iv) Operadores booleanos³⁷

Los sistemas de información documental no sólo permiten la búsqueda de palabras aisladas dentro de un documento, sino que una consulta puede contener varias palabras y el sistema realiza la búsqueda con base

³⁴ Por ejemplo, que a una consulta sobre OBRAS PUBLICAS el sistema recupere una unidad-documento donde se encuentren simultáneamente las palabras OBRAS ARTISTICAS y SECRETARIA DE EDUCACION PUBLICA.

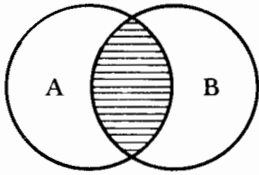
³⁵ Elmi, Corzo. . . [19], p. 108.

³⁶ López-Muñiz Goñi, *Tratamiento y recuperación de la información jurídica* [43], pp. 251-252.

³⁷ Véase *El Sistema UNAM-JURE* [31], pp. 93 y ss.

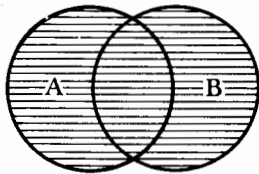
en los criterios de su presencia simultánea, alternativa o excluyente. Esto se logra mediante el empleo y combinación de los llamados operadores lógicos o booleanos. Las tres operaciones básicas son las siguientes:

– Intersección: $A \cap B$



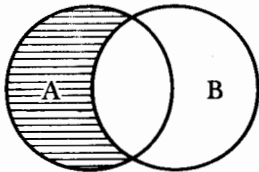
Significa que los términos A y B deben encontrarse en la misma unidad-documento, para resolver la consulta.

– Conjunción: $A \cup B$



Significa que resuelven la consulta todas las unidades documento en donde se encuentren los términos A o B (tiene por resultado la suma de ambos).

– Diferencia: $A \cap \bar{B}$



Significa que resuelven la consulta las unidades-documento en las que se encuentra el término A pero no el B, por lo que se excluyen aquellas unidades donde se encuentran en forma simultánea. (A o B pueden ser también conjuntos de palabras).

v) Restricción de distancia de los datos

El control de distancia de los datos se realiza mediante operadores especiales que indican a la máquina que las palabras que forman parte de una consulta (y cuya relación se define mediante los operadores lógicos) deben encontrarse a cierta distancia dentro del documento: dentro de la misma unidad-documento (frase, párrafo, documento) o a otras distancias, como por ejemplo, que sean adyacentes.

d) *Los instrumentos lingüísticos: la recuperación por medios semántico-conceptuales*

En este apartado se tratarán aquellas técnicas o instrumentos que tratan de resolver algunos problemas lingüísticos de los sistemas documentales.

i) enmascaramiento, truncamiento³⁸

Esta técnica consiste en proporcionar a la computadora sólo una parte de la palabra de consulta, de manera que la máquina extienda automáticamente la búsqueda a aquellas otras que cumplan ciertos requisitos; así, por ejemplo, todas las palabras que tengan la misma raíz, o bien se dejan abiertos los sufijos, prefijos, etcétera.

Ejemplo:

MAR

MAR

MAR

MAR

pero también:

MAR

MAR

LEGALIDAD

LEGALIDAD

³⁸ Cfr. Elmi, *Corso*. . . [19], pp. 103 y ss.

Las dificultades de esta técnica radican en que no dan los mismos resultados en idiomas diferentes. Por lo pronto, es más apropiada para el inglés que para el español, que tiene mayor número de variantes a partir de la raíz de las palabras.

ii) el tratamiento de las variables alotácticas y la normalización del lenguaje

Se llaman términos alotácticos a aquel conjunto de palabras que normalmente tiene una misma raíz morfológica (que es indicio de un parentesco semántico) y cuyo uso depende de la sintaxis que una frase, esto es, se trata de palabras que pueden ser equivalentes semánticamente con sólo variar la sintaxis de una frase.

Un ejemplo podrá aclarar la idea:

SE CREA UNA EMPRESA

CREACION DE UNA EMPRESA

UNA EMPRESA HA SIDO CREADA

Los vocablos CREA, CREACION, CREADA son equivalentes alotácticos, pues su contenido semántico es el mismo, aunque sólo cambie la sintaxis de la oración.³⁹

Este instrumento lingüístico presupone la elaboración manual de listas de equivalentes alotácticos, pero con ello se logra reducir la gran cantidad de términos variables del lenguaje natural, esto es, la determinación de estas equivalencias tiene por efecto una de normalización del lenguaje.

iii) los instrumentos semántico-conceptuales: el *thesaurus*

En forma general se puede definir un *thesaurus* como “un léxico jerarquizado que comprende una red de interconexiones, de exclusiones, de discriminaciones y de proximidades semánticas bajo la forma de listas de substitutos, de contrarios, de términos vecinos o los genéricos”.⁴⁰

Las tres notas básicas definitorias de un *thesaurus* son: que se trata de un vocabulario organizado y estructurado; que entre los términos que lo componen se establecen determinadas relaciones y que se refiere a un campo o disciplina particular (el derecho, la medicina, etc.).⁴¹

³⁹ Véase *El Sistema UNAM-JURE* [31], pp. 52 y ss. e IRETIJ, *Le traitement par ordinateur des variables allotaxiques* [35].

⁴⁰ Linant, *L'informatique*. . . [41], p. 84.

⁴¹ *Id.*, pp. 115 y ss.

El *thesaurus* proviene del ámbito de la catalogación bibliográfica, en el sentido de ser un vocabulario estructurado previamente que sirve para clasificar documentos, tomando de él las voces o descriptores relevantes y normalizando el lenguaje. Por su función, se trata de un *thesaurus* de clasificación.

Sin embargo, el *thesaurus* puede tener otras funciones.

En un sistema de información documental es posible elaborar un *thesaurus* a partir del lenguaje de los documentos ya almacenados. Dicho *thesaurus*, al reflejar el vocabulario y las relaciones semánticas entre sus términos constituye un instrumento de gran ayuda para la consulta, aunque su función no se agote ahí.

Si bien existen diversas clases de *thesauri*, dependiendo a la materia que se apliquen, baste decir que se trata de un instrumento que busca dar cuenta de las relaciones semánticas en el nivel de los conceptos y las ideas, mas no sólo entre las palabras.

Algunas de estas relaciones son las siguientes:

- Sinonimia y antonimia.
- Polisemia.
- Analogía o vecindad (normalmente indicada por la abreviatura RT = *Related Term*).
- Hiponimia o relación de jerarquía (indicada por las abreviaturas NT y BT = *Narrower Term* y *Broader Term*).⁴²

Estas relaciones pueden ser definidas formalmente de acuerdo con características como la transitividad, doble implicación, etcétera, lo que las hace aptas para su tratamiento por computadora.⁴³ Sin embargo, debe tenerse en cuenta que el lenguaje natural y las relaciones entre sus conceptos se hallan lejos de constituir un sistema de axiomas y relaciones formales.

B. Los sistemas de información jurídica documental

1. *Problemas particulares en la concepción y desarrollo de sistemas de información jurídica documental*

En una primera impresión parecería que la concepción y elaboración de un sistema de información documental en materia de derecho es un problema exclusivamente de técnica informática, es decir, que basta con aplicar a la documentación jurídica las técnicas y los instrumentos informáticos explicados en el apartado anterior.

⁴² Ciampi, Fameli y Trivisonno, "THES/BID: A Computer-bases Thesaurus of Terminology in Computers and the Law" [86], pp. VII y ss.

⁴³ *Id.*, p. X.

Sin embargo, esto no es así. Los documentos jurídicos poseen características particulares, reflejo de la propia estructura del derecho, que requieren un diseño y adaptación de acuerdo con una necesidad jurídica. En este sentido, la informática jurídica demuestra ser verdaderamente interdisciplinaria: la aplicación de los principios y técnicas de la informática a un objeto, el derecho (o la información jurídica) que da, en razón de su naturaleza y estructura, un perfil particular al producto resultante.

En segundo lugar, la elaboración de un sistema de información documental jurídica significa *más* que el almacenamiento y automatización de lo que antes se encontraba en papel.⁴⁴ Se trata de automatizar la información jurídica de acuerdo con su estructura y dinámica propias, no una publicación. Tal es lo que significa también el paso de la “recuperación de los documentos” a la “recuperación de información” de que habla Layman E. Allen.⁴⁵

a) *Algunos problemas particulares*

i) la selección de la información

La cuestión de la selección de la información (¿qué se quiere automatizar?) va íntimamente ligada a la de sus destinatarios (¿para quién y para cubrir cuáles necesidades se va a automatizar?), y en función de la respuesta se definen las características y contenido del sistema.

De acuerdo con lo anterior, habría tantas clases de sistemas como necesidades específicas de información.

— Por un lado se encuentran los sistemas generales de información jurídica, que intentan abarcar todo lo relativo a una o varias fuentes de derecho. No obstante, estos sistemas no pueden prescindir de una selección de la información (en el sentido de discriminación) o de su tratamiento diferencial. La selección y tratamiento de la información debe guiarse por criterios racionales y prácticos, en vista de resolver rápidamente una necesidad concreta. Como es de suponerse, la utilidad e interés de los sistemas generales varía de acuerdo con las características y necesidades de los usuarios.

— Por otro lado se desarrollan ya sistemas de información temáticos y especializados, dirigidos a un sector particular de los juristas y a la medida de sus necesidades. Estos sistemas tendrán una perspectiva más amplia sólo si se basan en los sistemas generales (véase también, *infra*, IV, 1, d, i).

⁴⁴ Bibent, *Compte-rendu d'une mission au Mexique* [3], § 6.

⁴⁵ Cfr. Allen, “Beyond Document Retrieval, Toward Information Retrieval” [75].

ii) La representación documental

Comúnmente se acepta que la indexación mediante lenguaje cerrado es la representación documental más adecuada para un catálogo bibliográfico; el *full text* para la legislación y el análisis para la jurisprudencia y los artículos de doctrina jurídica.⁴⁶

El caso de la legislación requiere quizá una discusión aparte.

A primera vista, la elección del *full-text* parece indiscutible: representa de una manera objetiva y completa el contenido de la ley (que tiene casi carácter sacramental) y proporciona al usuario de manera inmediata el texto mismo de la disposición.

Sin embargo, el principal problema del *full-text* aplicado a la legislación es, además del costo de captura y almacenamiento, la posibilidad de silencio por falta de explicitación y la producción de ruido debido a la definición un tanto arbitraria de la unidad-documento.

Habitualmente los sistemas de *full-text* aplicado a la legislación adoptan como unidad-documento el artículo o alguna de sus subdivisiones, partiendo de la suposición de que el artículo es una unidad de significado dentro de la ley, reglamento, etcétera. Esta suposición, empero, pierde sustento en los casos en que, por una deficiente técnica legislativa, el artículo no es tal unidad de significado.

Si lo importante es obtener el documento en su forma original, nada impide utilizar otro esquema de representación documental como medio de recuperación de la información y alguna otra técnica para el almacenamiento de los textos en su forma original. En otras palabras, la posibilidad de obtener el documento fuente como producto de una consulta no necesariamente debe ir acompañado del *full-text* como su representación documental.

Por otro lado, el principal problema del análisis mediante lenguaje abierto es la subjetividad y falta de uniformidad del trabajo del analista al asignar los descriptores, lo cual puede tener efectos negativos particularmente si se trata de la legislación.

No obstante, este problema se puede resolver en buena parte elaborando "estructuras" para el análisis. La posibilidad de establecer dichas estructuras depende de la organización misma de las ramas del derecho, de sus elementos invariables. Así, por ejemplo, son elementos estructurales de los delitos el tipo, la pena, la forma de persecución, etcétera.⁴⁷ Las estructuras de análisis tienen entonces un efecto de normalización y control objetivos del lenguaje.

⁴⁶ Véanse las opiniones que recoge Cáceres, "La metodología del 'abstract' legislativo" [81].

⁴⁷ Díaz Aranda e Islas Colín, "Notas para el análisis de los delitos" [89].

b) Las "relaciones internas" de la información

Se encuentra en la naturaleza del derecho el ser un orden, es decir, una estructura cuyos elementos establecen ciertas relaciones entre sí. Sin entrar en discusión sobre cuáles son estos elementos y de qué clase sus relaciones internas, resulta claro que todo sistema de información jurídica debe tomar en cuenta y dar un tratamiento adecuado a estas relaciones. Un ejemplo de ello sería la existencia de referencias internas de las normas jurídicas emanadas de una misma fuente del derecho (la remisión de una determinada ley a otra) o bien, de normas producto de diferentes fuentes (la jurisprudencia que interpreta un artículo de una ley).

La situación se complica por el hecho de que el orden jurídico, y por tanto también las relaciones entre sus elementos, se modifican constantemente en el curso del tiempo.⁴⁸ Para el sistema de información, esto implica la necesidad de prever la actualización de su contenido.

Dependiendo de su estructura interna y de la información que contiene, un sistema de documentación jurídica automatizada deberá reflejar y tratar estas relaciones internas en forma más sencilla o más compleja.

Así, por ejemplo, en un sistema de información legislativa que se actualiza en forma acumulativa (los documentos nuevos — como las reformas — simplemente se suman a los anteriores), las relaciones entre las disposiciones jurídicas (reformas, abrogaciones) se incluyen en los documentos informatizados, pero se deja al usuario la interpretación de los efectos que estos actos de modificación puedan tener sobre el orden jurídico.⁴⁹

En otras ocasiones, la estructura de la información documental es, en sí misma, de mayor complejidad, lo cual requiere un tratamiento informático especial al interior del sistema de información.

Un ejemplo puede aclarar la cuestión de las relaciones internas:

Las resoluciones del poder judicial federal mexicano son obligatorias en ciertas circunstancias para el propio poder judicial y otros tribunales. La Ley de Amparo establece, en su artículo 192, que "las ejecutorias constituirán jurisprudencia, siempre que lo resuelto en ellas se sustente en cinco sentencias no interrumpidas por otra en contrario. . ." y en determinadas condiciones de votación.

Así, una tesis de jurisprudencia tendrá, al menos, cinco precedentes. Si se quiere obtener éstos a partir de la tesis, no existe garantía de que

⁴⁸ Cfr. Bauer-Bernet, "Les aspects temporels dans la formalisation et l'automatisation du droit" [67], pp. 110-130.

⁴⁹ En ocasiones la simple reconstrucción y actualización de un texto legislativo presenta dificultades casi insuperables.

pueda lograrse con toda efectividad a través de las palabras que representan el contenido de los precedentes, sino que debe establecerse un mecanismo de encadenamiento especial. De esta manera, el sistema debe ser capaz de reconstruir la historia de una tesis de jurisprudencia o de un precedente, a partir de sus precedentes o de sus consecuentes.⁵⁰

Un sistema de información documental que incluya la legislación, la jurisprudencia y la doctrina no puede basarse solamente en el lenguaje documental como medio de recuperación de toda la información relativa a una cuestión jurídica, pues ese lenguaje varía notablemente en los conceptos que utiliza. En consecuencia, la relación entre la legislación, la jurisprudencia y la doctrina debe darse también a partir de otros elementos formales que funcionen como puente, como pueden ser las referencias legislativas implícitas o expresas.

c) *Características del lenguaje jurídico y construcción de instrumentos lingüísticos*

El lenguaje del derecho posee características particulares que deben tomarse en consideración en la creación de un sistema de información documental, particularmente en el aspecto de la representación documental y en el de la elaboración de los correspondientes instrumentos lingüísticos.

i) características del lenguaje jurídico

Casi al mismo tiempo que se iniciaron los estudios de lingüística, comenzó también el estudio de las conexiones entre el derecho y el lenguaje porque, ante todo, el derecho es discurso o, si se prefiere, expresión lingüística. El lenguaje jurídico forma parte del lenguaje natural y, por ello, comparte sus características generales, así como los problemas que enfrenta todo proceso de comunicación. Al mismo tiempo, el lenguaje del derecho posee rasgos propios que derivan de su función normativa en la vida social.

El lenguaje del derecho tiene dos niveles que hay que distinguir: el lenguaje *del* derecho, de las normas, y el lenguaje *sobre* el derecho.

El lenguaje de las normas se presenta, en primer término, como discurso, como una imponente y compacta construcción lingüística, pero es más que eso pues se trata de “un lenguaje peculiar dotado de extrañas propiedades operativas”.⁵¹

Una de estas extrañas propiedades del lenguaje del derecho es su carácter “performativo”, es decir, que sus palabras no dicen o describen

⁵⁰ Cfr. Bibent, *Compte-rendu*. . . [3], §§ 35 y ss.

⁵¹ Giménez, Gilberto, *Poder, estado, discurso* [25], p. 70.

algo, sino que “hacen algo”: decir equivale a hacer.⁵² “El discurso del derecho no es meramente constataivo, descriptivo o valorativo, sino un discurso operativo, un discurso-acción”.⁵³ En términos más usuales, ahí reside el carácter “normativo” del derecho.

Otras características del lenguaje jurídico que son de particular interés para la creación de sistemas de documentación automatizada son las siguientes:

— El discurso jurídico es un *discurso implícito*, porque hace referencia a cosas sin nombrarlas expresamente.⁵⁴ En este sentido, el proceso de “explicitación” forma parte del proceso general de asignación de significados por parte del destinatario. Comprendería lo mismo la elaboración doctrinal como la interpretación judicial y, en general, todo tipo de construcción conceptual a partir del derecho, cualesquiera que sean sus medios u objetivos.⁵⁵ La explicitación tiene un sentido concreto y técnico en la indexación de documentos jurídicos.

— El discurso jurídico es un *discurso referencial*, esto es, contiene expresiones “que permiten al locutor señalar al destinatario uno o varios objetos particulares del universo del discurso”⁵⁶ o lo que es igual, la referencialidad del discurso jurídico consiste en designar, señalar cosas que forman parte del mismo derecho.⁵⁷ “Esta función referencial del discurso permite al derecho autodefinirse y deviene un modo discursivo implícito”;⁵⁸ ello no es más que una derivación del carácter sistémico del derecho.

Existe, no obstante, un tipo de referencialidad “señalada por la presencia del artículo definido delante de los grupos nominativos”,⁵⁹ que implica una cierta ambigüedad que, tratándose de los bancos de información, no puede ser resuelta si el que consulta no obtiene todos los documentos ligados a un universo de información pues, según el caso, no podrá identificar ciertos sujetos (por ejemplo, la “autoridad competente”) ni saber si se trata de un referente definido o de una indicación de existencia, inconveniente que se intensifica por la usual división de las leyes en artículos, fracciones, incisos, etcétera.⁶⁰

⁵² *Id.*, p. 71. Así se explica que el título de la traducción francesa de la obra de John Austin que trata esta cuestión (*How to do things with words*) sea *Quand dire c'est faire*: cuando decir algo, significa hacer algo.

⁵³ *Ibid.*, p. 71.

⁵⁴ Bourcier, “Information. . .” [79], pp. 14-15.

Así, por ejemplo, el artículo 1910 del Código Civil que establece la RESPONSABILIDAD CIVIL, no la menciona expresamente.

⁵⁵ *Id.*

⁵⁶ *Ibid.*, p. 16.

⁵⁷ *Ibid.* Sobre las clases de referencialidad véase la p. 16.

⁵⁸ *Ibid.*, p. 17.

⁵⁹ *Ibid.*, p. 16.

⁶⁰ *Ibid.*, p. 17.

— El discurso jurídico es un *discurso conceptual*, característica que es complementaria de las anteriores. Decir que el derecho “reposa actualmente sobre el concepto” significa decir “una organización cognoscitiva y lógica del lenguaje” determinada.⁶¹

Para Kant, el concepto es una “representación general, o la representación (*Vorstellung*) de aquello que varios objetos tienen en común”;⁶² la formación de los conceptos, es decir, la conjunción y definición de estos caracteres se guía por el interés cognoscitivo, es decir, se trata de una cuestión de adecuación o conveniencia.⁶³

El derecho no siempre ha tenido el mismo grado de elaboración y abstracción conceptual.⁶⁴ La evolución de sus conceptos es reflejo de las necesidades del tráfico jurídico moderno, de una mayor precisión técnica y de mayor seguridad en la regulación de la vida social. El concepto como centro de gravedad del derecho tiene también una función política, aparte de la meramente “técnica”. Así, por ejemplo, la llamada “jurisprudencia de conceptos” proponía que el derecho se obtuviera exclusivamente mediante técnicas lógico-conceptuales como medio para sujetar estrictamente al juez a la ley y no dar lugar a la valoración de ésta.⁶⁵

¿Cuál es el origen lingüístico de los conceptos del derecho?

La coincidencia del lenguaje jurídico con el común, del cual proviene, es parcial. Las funciones propias del derecho requieren que su lenguaje tenga un carácter técnico y especializado.⁶⁶

Podemos identificar al menos tres clases de conceptos del derecho, dependiendo de su mayor o menor grado de coincidencia con el lenguaje común:

- Conceptos de hecho, de la vida cotidiana, que el derecho toma para darles relevancia jurídica (ejemplo: “automóvil”).⁶⁸
- Conceptos del lenguaje corriente que asumen un significado especial y técnico en el derecho (ejemplo: “solidaridad”).
- Conceptos propios y exclusivos del derecho (ejemplos: “subrogación”, “usufructo”).⁶⁹

⁶¹ *Ibid.*, p. 18.

⁶² *Lógica*, § 1, citada por Zippelius, *Rechtsphilosophie* [73], p. 5.

⁶³ Zippelius, *Rechtsphilosophie* [73], p. 6.

⁶⁴ Bourcier, “Information. . .” [79], p. 18.

⁶⁵ Zippelius, *Rechtsphilosophie* [73], pp. 199-203.

⁶⁶ Mazet, y Bernard, “Aspects Linguistiques du traitement automatique de l’information juridique” [115], p. 73.

Después de todo, si el derecho es una técnica para controlar el comportamiento social (Kelsen), su lenguaje también deberá ser técnico.

⁶⁷ *Id.*

⁶⁸ Esta clase podría subdividirse en dos grupos: conceptos técnicos no jurídicos (pertenecientes a otras ciencias o disciplinas) y conceptos no técnicos del habla común.

⁶⁹ Muchos de estos conceptos pasan después al lenguaje común, donde asumen un significado más amplio y vago.

El lenguaje del derecho posee también características propias en cuanto a su estructura y sintaxis, entre las que pueden señalarse:

— Su grado variable de abstracción (dependiendo de la fuente y jerarquía de la norma).

— Su impersonalidad.

— La denominada derivación regresiva o cambio de categoría gramatical: “se trata, mediante una sustantivación a partir de un adjetivo o de un participio, de evitar el recurso a una proposición relativa; ejemplo: el cedente, el gerente, el adoptante. . .”.⁷⁰

— Distintas posibilidades de formulación de una idea, tanto desde el punto de vista sintáctico como semántico.

Estas características formales, estructurales y normativas del discurso del derecho han permitido el surgimiento de una corriente del pensamiento jurídico que ha sometido este lenguaje a un análisis lógico.⁷¹ Félix Oppenheim sostiene que el derecho “visto como un lenguaje, puede ser el objeto de dos tipos de ciencia jurídica: la *ciencia del derecho empírica*. . . que estudia la relación entre los enunciados jurídicos y los seres humanos que los crean, interpretan y aplican” y el *análisis lógico del derecho* que “se ocupa de los enunciados jurídicos en su aspecto lógico”.⁷²

Las posibilidades del análisis lógico-formal de los enunciados jurídicos tienen gran importancia para la informática jurídica, sin olvidar los límites, ya mencionados, que tiene este análisis.

ii) La construcción de instrumentos lingüísticos

Los instrumentos lingüísticos en un sistema de información jurídica documental tienen por finalidad “reconstruir” las relaciones semánticas entre los conceptos⁷³ del derecho expresados en una base documental, como medio de control del lenguaje y ayuda a la consulta. Entre ellos, el principal es el *thesaurus*.

Por sus características, el lenguaje del derecho presentará problemas distintos para la construcción de un *thesaurus de los que puede plantear, por ejemplo, la medicina*.⁷⁴ Como regla, puede decirse que la estructura

⁷⁰ Mazet y Bernard, “Aspects. . .” [115], p. 74.

⁷¹ Véase Bobbio, *Derecho y lógica* [4], pp. 22 y ss.

⁷² Oppenheim, *Lineamientos de un análisis lógico del derecho* [52], p. 15.

⁷³ También es posible construir instrumentos lingüísticos automatizados o no, sobre palabras, con la misma finalidad de controlar el lenguaje de todo un campo del conocimiento. Véanse los trabajos correspondientes del IRETJ, *Traitement par ordinateur des variables allotaxiques* [35] y Mazet y Bernard, *Le thesaurus de mots*. . . [47], especialmente pp. 3-5.

⁷⁴ Para una introducción a estos problemas véase Knapp, “Alcuni problemi relativi alla costruzione di un thesaurus giuridico” [97].

del thesaurus será un reflejo de la estructura del lenguaje que pretende automatizarse.

En el derecho, no es igual el lenguaje de la legislación al de la jurisprudencia y la doctrina, ni son iguales los conceptos que utilizan y, por tanto, tampoco las relaciones que éstos establecen entre sí.

En la legislación, el lenguaje es más general, mientras que el de la jurisprudencia es la puerta para que muchos conceptos concretos de hecho reciban una calificación jurídica y entren en el mundo jurídico;⁷⁵ existen, por otro lado, conceptos que son exclusivamente de elaboración doctrinal.

En segundo lugar, tanto la estructura como las relaciones de los conceptos varían entre las distintas ramas del derecho. En las más técnicas y elaboradas, como el derecho penal, existe un mayor grado de “clasificación” de los conceptos y las relaciones entre éstos son más definidas y formales. Esto permite, por ejemplo, el establecimiento automático de relaciones entre los conceptos. Por ejemplo, todas las especies del concepto genérico DELITOS tienen una relación asociativa definida entre sí.⁷⁶

Por estas razones, resulta casi imposible elaborar un *thesaurus* único para las distintas fuentes y ramas del derecho. La validez de las relaciones de los conceptos se extiende sólo a uno o a varios sectores jurídicos, pero no a todos.

Por lo que se refiere a la metodología para la construcción de un *thesaurus* jurídico, no existe una “receta”, sino que la elaboración varía según la clase y función del *thesaurus* y según la rama o la fuente del derecho que se trate. Sin embargo, las etapas generales por las que debe pasar este trabajo de elaboración son las siguientes:

- Selección del campo o área del *thesaurus*.
- Selección y normalización del lenguaje y los conceptos pertenecientes a este campo.
- Determinación y definición de las relaciones semánticas que incluirá el *thesaurus* (sinonimia, asociación, hiponimia, antonimia, etcétera).
- Establecimiento de las relaciones entre los conceptos seleccionados, de acuerdo con su significado en el campo tratado.
- Automatización (que permite incluso la expansión automática de las relaciones formalmente definidas).⁷⁷

⁷⁵ Así, la estructura de un *thesaurus* de lenguaje jurisprudencial reflejará fundamentalmente esta relación entre los conceptos de hecho y su calificación jurídica. Véase por ejemplo IRETIJ, *Constitution d'un thesaurus en droit social*. . . [32].

⁷⁶ Véase por ejemplo Ciampi, Fameli y Trivisonno, THES/BID: A Computer-based Thesaurus of Terminology in Computers and the Law [86], pp. IX-X. La expansión automática de relaciones tuvo que ser revisada en los casos en que era semánticamente inadecuada.

⁷⁷ Como bibliografía adicional sobre el tema pueden consultarse Elmi, *Corso*. . . [19], pp. 110-145; Abba, Romano, Cammelli, Antonio y Elmi, “Definizione del Thesaurus GIUR in am-

2. LOS SISTEMAS DE INFORMACIÓN JURÍDICA DOCUMENTAL

a) *Evolución*

Los orígenes de los grandes sistemas de información jurídica documental suelen situarse en el sistema que el profesor John F. Horty, de la Universidad de Pittsburgh, Pennsylvania, creó en 1959 y que presentó el año siguiente ante la *American Bar Association*. El motivo concreto para la creación del sistema fue el encargo que recibió el profesor Horty de elaborar un proyecto de ley en materia de salud, que requería la comparación de la numerosa legislación de los estados. Tras fracasar en el intento de realizar el trabajo de comparación en forma manual, Horty se dirigió al Centro de Cómputo de la Universidad de Pittsburgh.⁷⁸

El sistema de Horty utilizó el esquema de archivos de texto y archivo de inversión (véase, *supra*, A, 2, c, i) para la recuperación de los documentos, esquema que casi todos los sistemas posteriores han seguido, con variantes.⁷⁹

Constantino Ciampi⁸⁰ considera que los sistemas de información jurídica documental se dividen en cuatro tipos, ligados en buena medida al desarrollo de la tecnología de las computadoras.

El primer tipo corresponde a los sistemas “batch”, en los que las consultas son procesadas en forma de “lotes” individuales. El sistema del profesor Horty era típicamente un sistema *batch*.

El segundo tipo está representado por los sistemas “on-line” (1971-), cuya innovación consiste en que la consulta puede ser sucesivamente afinada por el usuario en un verdadero diálogo con la máquina, con la ventaja adicional de poder interrogar la información a distancia y con mayor rapidez. Además, estos sistemas tratan de resolver algunos problemas sintácticos y lingüísticos mediante, por ejemplo, el uso de palabras truncadas y otros instrumentos que también son consultables en terminal.

El tercer tipo son los “sistemas on-line evolucionados” (1975-). Se trata de sistemas interactivos que incorporan nuevas y más poderosas funciones de consulta y que se vuelven más comprensibles para el usuario, el cual tiene a su disposición diversas herramientas (ordenación, formatos de presentación, etcétera) que hacen más eficaz su búsqueda.

biente TLS” [67], pp. 15-29; Giannoni, y Novaresio, “I nessi semantici e sintattici e la loro rilevanza nella costruzione dei «thesauri» (un esperimento)” [94].

⁷⁸ Ciampi, “La evolución de los sistemas de cálculo y de las técnicas de obtención automática de la información jurídica” [114], pp. 36-37.

⁷⁹ *Id.*, p. 39.

⁸⁰ *Ibid.*, pp. 34 y ss.

Una innovación interesante de estos sistemas consiste en que la búsqueda a través de la identidad entre las palabras de la consulta y las de los documentos es sustituida por una búsqueda de aproximación o semejanza entre la pregunta y los documentos, de acuerdo con un criterio lingüístico.

El cuarto tipo (1980-) estaría representado por los "sistemas expertos" o "inteligentes", que en el fondo no son ya sistemas documentales. Sin embargo, es posible imaginar la combinación de los sistemas documentales "tradicionales" con algunas funciones inteligentes de auxilio a la consulta.

b) *Principales sistemas operativos en la actualidad*⁸¹

En este apartado se expondrán, muy brevemente, las características de los principales sistemas de documentación jurídica que existen, aclarando que se trata en su mayoría de bancos de información de gran tamaño.

i) Estados Unidos de Norteamérica

— LEXIS es un sistema comercial operado actualmente por la Mead Data Central. Es el sistema más grande del mundo. Ofrece sentencias y leyes tanto federales como de los estados. Incluye asimismo información jurídica del Reino Unido, Francia y Brasil, países donde también opera.

— WESTLAW es otro sistema comercial, de la *West Publishing Company*, introducido como servicio a partir de 1975. Contiene sobre todo resoluciones judiciales federales y estatales, además de otros archivos especiales en algunas áreas jurídicas.

Entre los sistemas públicos puede citarse a JURIS, sistema desarrollado y operado por el Departamento de Justicia como servicio de apoyo a los procuradores federales, ofreciéndoles sentencias y derecho legislado, codificado o no, así como resoluciones de otras autoridades como la Comisión Reguladora Nuclear; y FLITE (*Federal Legal Information Through Electronics*), que es más bien un sistema *batch*, inaugurado en 1963 y que pertenece al *Air Force Accounting and Finance Center*.

⁸¹ Para la redacción de este inciso, se hizo una síntesis de la información contenida en los siguientes trabajos:

Ciampi, "La evolución. . ." [114], pp. 34-49; Nunn-Price, "Diseño, construcción y resultado de un servicio a gran escala de información jurídica on-line" [114], pp. 147-155; Stewen, y Busse, "Juris: El sistema de información legal de la República Federal de Alemania" [114], pp. 157-170; Martino, "La informática jurídica hoy" [101], pp. 559-561; Fiedler, "Strukturen und Wirkungen. . ." [92], pp. 148 y ss.; Mazet, "Les systemes informatisés de documentation juridique. . ." [102], pp. 776-779; Mason, *An Introduction to Using Computers in the Law* [45], pp. 106 y ss.

ii) Francia

En Francia, después de la reorganización de la informática jurídica de los años 1984 y 1985, la producción y desarrollo de los sistemas de información se realiza alrededor de tres polos: el sector público emisor de textos de interés nacional (tratados, leyes, reglamentos, etc.) y de resoluciones de los tribunales supremos; el sistema nacional de la jurisprudencia, que tiene asignada la informatización de las resoluciones de las jurisdicciones inferiores, y la enciclopedia electrónica, que reagrupa el conjunto de bancos de información especializados.

Los principales bancos de información que contribuyen en esta organización son:

— CNIJ (antes CEDIJ): *Centre National d'Informatique Juridique*; Juris-Data, producido por una agrupación de editores; Sydoni, que depende del Notariado y la Caja de Depósitos y Consignaciones; Lex: archivo legislativo y reglamentario del Secretariado general del Gobierno; Juridoc: banco de jurisprudencia y doctrina producido por el IRETIJ de Montpellier; Resagri: Red de organismos agrícolas; Jurinpi: banco del Instituto Nacional de la Propiedad Industrial; Sigad: producido por la unión de industrias mecánicas y metalúrgicas; CDJO: *Centre de Documentation Juridique de l'Ouest*, producido por la Universidad de Rennes y el CNRS.

iii) Alemania Federal

En Alemania Federal existen varios sistemas, aunque la mayoría son pequeños y muy especializados. Los más importantes que pueden mencionarse son:

— JURIS (Sistema de Información Jurídica). Su creación se remonta a un encargo que el gobierno federal alemán hizo en 1973 al Ministerio de Justicia. Actualmente es un sistema on-line que ofrece bancos de información de jurisprudencia, doctrina, leyes federales y ordenanzas, por un total de más de 400 000 documentos.

— DATEV, es un sistema de desarrollo privado, de acceso limitado a los socios de la asociación que lo creó, especializado en el campo del derecho fiscal. Contiene textos y referencias sobre legislación, jurisprudencia, lineamientos administrativos y doctrina.

iv) Italia

El sistema más importante es el ITALGIURE/FIND, de la Corte de Casación. Comprende archivos de legislación, jurisprudencia, doctrina (tratada por el Instituto para la Documentación Jurídica —IDG— de

Florenca), así como otra documentación como contratos colectivos nacionales y datos de los abogados y procuradores.

v) Bélgica

— CREDOC es el sistema más antiguo de Europa. Funciona desde 1969 y es patrocinado por la federación de notarios y abogados belgas. Una particularidad interesante es que cuenta con un *thesaurus* bilingüe francés-flamenco.

vi) Comunidad Europea

— CELEX contiene en inglés y francés el derecho comunitario europeo, así como los tratados y acuerdos con terceros países, la jurisprudencia de la Corte Europea, las resoluciones del Parlamento Europeo, etc.

vii) Brasil

— PRODASEN es el sistema de apoyo a la Cámara de Senadores, que contiene información jurídica sólo a nivel federal.

viii) Inglaterra

— EUROLEX es un servicio comercial que se inició en 1979, que incluye archivos de diversas clases y ofrece ya acceso a usuarios de la parte continental europea.

c) *Perspectivas*

Las perspectivas de desarrollo de los sistemas de información jurídica documental parecen encaminarse en varias direcciones:

— Una extensión de los sistemas existentes para tratar de incorporar la mayor cantidad posible de documentos jurídicos. En este sentido, algunos sistemas tienen la pretensión no sólo de auxiliar las labores cotidianas de los juristas, sino convertirse en una especie de “memoria jurídica” similar a cualquier biblioteca o archivo. Esto implica un posible problema de coordinación si existen varios sistemas en competencia.

— Al mismo tiempo seguirá ampliándose la posibilidad del acceso a los diversos sistemas por vía telemática, lo cual quizá tenga por resultado un impulso al desarrollo de interfaces multilingües y una integración y homogeneización de los instrumentos y lenguajes de consulta.

Según se apuntó anteriormente, los sistemas documentales, como los conocemos hoy, tienen limitaciones intrínsecas que requieren un nuevo punto de partida para ser superados. Uno de ellos sería el desarrollo de

un mecanismo “inteligente” que tuviera como base de conocimiento el lenguaje natural de los documentos,⁸² es decir, el mecanismo inteligente realizaría en el momento de la consulta las operaciones que actualmente hay que introducirle previamente. Este posible paso en la evolución de los sistemas de información documental requerirá también nuevos avances en el diseño de las computadoras.

⁸² Véase Heather, “Future Generation Computer Systems in the Service of the Law” [67], vol. II, pp. 172 y ss.