

El impacto de la informática en el lenguaje jurídico

INMACULADA RICO BONILLA

*Letrado Asesor Servicios Jurídicos Banco de España
(ESPAÑA)*

Antes que nada quisiera dar las gracias al Comité Científico y Organizador del III Congreso Iberoamericano de Informática y Derecho por la oportunidad que me da de poder dirigirme a Vds., para comentarles la experiencia llevada a cabo por los Servicios Jurídicos del Banco de España en el campo de la Informática aplicada al Derecho; experiencia que se ve reflejada en nuestra Base de Datos de Legislación Financiera.

Y con este «pretexto» hacer una reflexión en voz alta acerca de lo que, en mi opinión, se configura como una de las cuestiones más sensibles en los Bancos de Datos Jurídicos y que no es otra que el estudio del lenguaje. Por ello, he titulado mi ponencia «El Impacto de la Informática en el Lenguaje Jurídico».

Siendo ésta la principal preocupación del Banco de España a la hora de crear su Base de Datos, me gustaría exponer –si Vds. me lo permiten– lo que para nosotros han sido los pilares básicos de nuestro sistema, haciendo especial hincapié en aquella cuestión.

Estos pilares los podríamos resumir en los siguientes:

1. Delimitación conceptual y temporal.
2. Procedimiento de identificación del texto legal: unidad documental.
3. Técnicas de análisis documental: tratamiento del lenguaje jurídico.
4. Sistema de recuperación empleado.

1. DELIMITACION CONCEPTUAL Y TEMPORAL

Si bien su enunciado en teoría es tarea fácil, sin embargo, en la práctica no lo es tanto. Y esto porque en todo sistema de información jurídica automatizada resulta menos problemático un enfoque general del sistema que incluya todas las normas que componen el ordenamiento jurídico, que un enfoque parcial, en el que, inevitablemente, hay que seleccionar las disposiciones por razón de la materia escogida. La dificultad aumenta, en este caso, porque los puntos de contacto entre unas disciplinas y otras son cada día mayores.

En nuestro caso –aun siendo la solución más onerosa–, nos hemos atrevido a dar a nuestra Base de Datos un enfoque jurídicamente sesgado.

Así las cosas, podemos adelantar una definición que nos sirvió para fijar los extremos materiales de la Base de Datos y que es la de incluir aquellas disposiciones que reglamentan el Sistema Financiero e integran el conjunto de instituciones y operaciones destinadas a facilitar la actividad que es característica de los diferentes agentes económicos.

Quedan así referidos, aspectos tan diversos como los relativos a la estructura jurídica de las entidades de crédito, órganos de gobierno, controles formales que reglamentan su actividad –disciplina e inspección, riesgos, información financiera, auditoría, arbitraje, transparencia de operaciones, estadística, defensa del consumidor y de la competencia y publicidad–, mercados monetarios y de valores, control de cambios, operaciones activas, pasivas y de servicios, asignación de resultados, organismos internacionales de carácter económico, registros administrativos con incidencia bancaria, así como aspectos fiscales de ciertas operaciones bancarias.

Abundando en lo anterior, será preciso señalar que no sólo se incluyen las normas emanadas de la Administración del Estado, sino también aquellas que –afectando al sistema financiero– promulgan y publican las Comunidades Autónomas en virtud de las competencias asumidas en sus respectivos Estatutos de Autonomía.

Una vez estudiada la delimitación conceptual de la Base de Datos, con- vendrá referirse a su contenido. Se trataba de analizar si nuestro sistema de informática jurídica documental debería cubrir únicamente las normas vigentes del sistema financiero o también las que hubieran dejado de aplicarse, bien por haber sido derogadas expresamente o bien porque la sucesión en el tiempo del dictado de las normas ha provocado situaciones de incompatibilidad entre ellas.

La decisión ha consistido en emplear un criterio mixto. Así, en el tiempo en que se hizo el estudio de selección de normas –y retrocedemos al año 1984–, se dispuso que solamente integrarían nuestro sistema aquéllas que estuvieran vigentes, pero con la particularidad de que todas las que con el devenir del tiempo fueran acumulándose en nuestro archivo legislativo y pudieran sufrir cualquier alteración no se eliminarían de nuestra Base de Datos,

sino que habrían de configurar un archivo paralelo denominado por ello «Archivo Histórico».

Lo trascendental de esta decisión es hacer frente a cuestiones importantes como la intertemporalidad de las normas y las derogaciones tácitas.

La correlación entre los dos archivos de vigencias e histórico se ha llevado a cabo mediante el establecimiento de múltiples referencias cruzadas, que indican en todo momento el alcance de la alteración y su tipo, es decir, si se trata de una derogación total o parcial de la norma, de la derogación parcial de un artículo, así como las modificaciones, suspensiones o sustituciones de unas normas por otras.

Por lo que se refiere a la delimitación temporal, el señalamiento de un término «a quo» en todo sistema de información jurídica informatizada es importante para dar al usuario un elemento de referencia temporal precisa.

De la misma manera que antes se ha mencionado el año 1984 para realizar el estudio de las normas que estaban en vigor en ese tiempo, esta misma época nos sirvió para remontarnos en nuestro estudio a la disposición más antigua de la Base de Datos y que es en el ámbito estatal el Real Decreto de 22 de agosto de 1885, por el que se aprueba el Código de Comercio. Por tanto, todo lo promulgado desde esa fecha hasta hoy compone los límites temporales de nuestro sistema.

En lo que se refiere a las Comunidades Autónomas, se ha partido del primer Estatuto de Autonomía, que es el correspondiente al País Vasco, aprobado por Ley Orgánica 3/1979, de 18 de diciembre.

2. PROCEDIMIENTO DE IDENTIFICACION DEL TEXTO LEGAL: UNIDAD DOCUMENTAL

En primer lugar habrá que decir que la legislación, por su contenido, no es susceptible, a nuestro juicio, de resumen. El resumen implica la participación de una voluntad distinta a la del legislador, y si lo que se pretende es facilitar una información objetiva y uniforme, la introducción del texto debe hacerse de forma íntegra.

Sentado lo anterior, cabe preguntarse qué extensión debe abarcar la unidad de información. Podría pensarse que una sola unidad correspondiera al cuerpo legal en su totalidad. Sin embargo, entendimos que esto no era lo correcto.

Obsérvese que la inclusión de todo un cuerpo legal en un mismo documento dificultaría enormemente la composición de descriptores, si se pretende, claro está, un análisis exhaustivo de todos los aspectos relevantes contenidos en la norma; aumentaría la posibilidad de ruido informático, obstaculizaría la tarea de actualización y puesta al día del documento y, por último, perturbaría la visualización del mismo en pantalla, al no existir una

identificación entre un artículo concreto y las palabras claves que le corresponden.

De ahí que el Departamento de Informática Jurídica del Banco de España optara, como solución más ventajosa, aunque de una tremenda complejidad analítica y referencial, la de identificar la unidad documental con el artículo, unidad mínima pero básica, de cualquier texto legal.

En consecuencia, en nuestro sistema una norma tiene tantas unidades de información como artículos componen. Pero, además, habrá que añadir, como partes independientes: el título o encabezamiento de la disposición, la exposición de motivos, y el resto de disposiciones transitorias, finales, derogatorias y otras similares, si las hubiere.

3. TECNICAS DEL ANALISIS DOCUMENTAL: TRATAMIENTO DEL LENGUAJE JURIDICO

Tradicionalmente se distinguen dos tipos de análisis: la elaboración de resúmenes y la fijación de descriptores o palabras claves.

La elaboración de resúmenes persigue un doble objetivo, de un lado permite hacer una clasificación selectiva de los documentos pertinentes en una búsqueda y de otro obtener información con un importante ahorro de tiempo, cumpliendo dos finalidades complementarias: menor ocupación de memoria en el ordenador y la posibilidad de clarificar el texto.

El segundo método de análisis consiste en la representación del contenido del texto, mediante la fijación de términos o descriptores. Esta representación se orienta básicamente a destacar la noción principal de los documentos.

Expuestas las dos técnicas de análisis documental, es preciso considerar la utilidad de estas técnicas. La legislación, por su contenido, no es susceptible, a nuestro juicio, de resumen, a diferencia de otras fuentes del Derecho, como la Jurisprudencia. De ahí que el Departamento de Informática Jurídica descartara la primera de estas técnicas y optara por la segunda.

Llegados a este punto surge la cuestión más sensible en la creación de bases de datos jurídicas y que no es otra que el estudio del lenguaje jurídico. Lo realmente necesario en la informatización de los archivos legales y que va a condicionar críticamente la eficacia de una base de datos es, de una parte, la aplicación de técnicas que impidan la existencia de los denominados «silencio» y «ruido» informáticos y, de otra, la normalización del lenguaje jurídico, reduciendo la multiplicidad de sus formas a términos previamente establecidos, así como resolviendo determinadas cuestiones lingüísticas.

Estos aspectos son tan graves –y debemos destacarlo otra vez–, que de muy poco servirá tener un maravilloso paquete informático, una clara y concisa definición de campos o un diseño de funciones fácilmente comprensible

para el usuario, si todo ello no va acompañado de la resolución a las cuestiones antes apuntadas.

Por elemental que resulte, hay que recordar que el objetivo de toda base de datos jurídicos pretende siempre responder a dos interrogantes:

– La primera supone que el usuario conoce la norma que responde puntualmente a su pregunta, pero desconoce en qué estado se encuentra. En este supuesto, la respuesta es fácil. Se tratará de indagar por cualesquiera de los campos objetivos de la norma, fecha de la disposición o de publicación y, sobre todo, por el párrafo de referencias para que aparezca en pantalla la norma con sus posibles derogaciones o modificaciones. Esto sucederá siempre que los responsables del sistema hayan tenido especial cuidado en hacer un seguimiento puntual del acontecer de cada disposición.

– La segunda interrogante que se plantea en toda base de datos y que va a definir totalmente su eficacia es la conectada a la búsqueda conceptual. La cuestión ahora se centra en que el consultante desea conocer como está regulado un determinado supuesto. A lo más puede intuir que hay disposiciones que lo regulan, pero desconoce cuáles son y si son todas. Es en este momento cuando surgen los peligros antes enunciados, y que son el «silencio» y el «ruido» informático.

Existe «silencio» informático cuando el sistema no responde de manera exhaustiva a una pregunta. En este caso la respuesta comprende algunos documentos pertinentes pero no todos.

El «ruido» informático surge cuando el sistema ofrece junto a documentos pertinentes otros que no lo son, por corresponder a materias totalmente ajenas.

La solución a estos inconvenientes viene dada por el estudio de una serie de técnicas que contrarresten los efectos negativos provocados por ambos fenómenos:

Técnicas que nos preservan del silencio informático

a) Resolución de los significados equivalentes o sinónimos.

Pensemos, por ejemplo, en una búsqueda bastante común en nuestra disciplina, como es la de conocer qué disposiciones regulan el nivel del coeficiente de caja. Sencillamente se interrogará al sistema por las palabras «nivel», «coeficiente» y «caja» y, al momento, aparecerán en pantalla aquellos documentos que cumplan la triple condición. Ahora bien, ¿serán los documentos obtenidos los únicos que responden a la cuestión o habrá otros sobre los que, inexplicablemente, el sistema haya guardado silencio? No es aventurado que el legislador haya utilizado otros términos sinónimos a «nivel», cual son «porcentaje», «límite» o «cuantía» del coeficiente de caja.

La validez de la respuesta será nula o, cuando menos, incompleta, si no se han fijado de manera apriorística relaciones de sinonimia entre dichos términos o no se ha fijado un único término que englobe todos los demás.

b) Resolución de nociones implícitas.

Es muy frecuente que el texto aluda a conceptos, personas jurídicas o simplemente invoque otros artículos, sin mencionarlos expresamente. La consecuencia es que si no se hace frente a dichas nociones mediante su explicitación, la pérdida de información es igualmente inevitable.

c) Normalización de siglas.

Existe este caso de «silencio» informático cuando el legislador emplea abreviaturas en lugar de palabras completas. La solución puede venir dada estableciendo informáticamente una sinonimia absoluta entre el conjunto de palabras y su representación abreviada en siglas, de manera que sea indiferente al sistema la invocación por cualquier medio, ya que siempre habrá una respuesta.

d) Concordancias normativas.

Esta técnica consiste en el señalamiento de concordancias de unas disposiciones con otras. En una base de datos legislativa la existencia de un campo referencial es un instrumento indispensable de información. Y ello no sólo porque el juego de las concordancias permite tener una visión más completa del contenido de la base de datos, sino porque constituye un elemento auxiliar en la búsqueda de documentos. En nuestro caso, se ha creado un campo informático dirigido exclusivamente a señalar estas concordancias, tanto a nivel genérico como específico. Además, la aplicación informática posibilita, mediante el uso de determinadas funciones, visualizar la norma referenciada expuesta en el documento base, sin necesidad de preguntar directamente por ella.

Causas del ruido informático y técnicas para su resolución

Antes ya se ha comentado que surge «ruido» cuando, en una búsqueda, el sistema genera documentos pertinentes juntamente con otros que no lo son.

Dos son las razones por lo que esto puede suceder:

a) La primera tiene su origen en la combinación de términos no queridos por el consultante, por cuanto, si bien existen en el texto, no guardan, sin embargo, relación unos con otros.

Una de las soluciones que aminoran los efectos perniciosos de este tipo de «ruido» es el empleo de los llamados operadores de aproximación, de manera que la respuesta del sistema quede condicionada por la proximidad entre palabras, entre frases o entre párrafos.

Aun así seguirán apareciendo, en ocasiones, combinaciones no deseadas.

b) El segundo motivo viene constituido por la existencia de palabras polisémicas, es decir, aquéllas que expresan diferentes significados según el campo semántico en el que actúan. Por ejemplo, el término «intervención» varía claramente de sentido si nos referimos a la intervención de una letra de

cambio o a la intervención de una entidad de crédito como medida cautelar sancionadora.

La solución a este problema se encuentra en la creación de sintagmas forzosos y en la elaboración de relaciones asociativas.

4. Sistemas de recuperación textual

Tradicionalmente se distinguen dos sistemas de recuperación:

- a) Full-tex.
- b) Descriptores.

La diferencia entre ambos reside en que en el procedimiento full-tex el usuario puede utilizar, para su recuperación, todas las palabras contenidas en el texto del documento, en tanto que en el de descriptores solamente se utiliza un determinado grupo de palabras, previamente establecidas.

El sistema full-tex tiene dos ventajas fundamentales:

- a) la labor previa a la introducción del documento requiere poco trabajo, y
- b) las posibilidades de recuperación son grandes.

Sin embargo, junto a las ventajas expuestas, existen serios inconvenientes que desvirtúan en gran parte este sistema y lo hacen sumamente peligroso, como ya ha quedado reflejado anteriormente.

El sistema de descriptores o palabras claves consiste en el uso de ciertos términos o sintagmas previamente definidos en el documento. El texto se incluye, como en el caso anterior, de forma íntegra en la base de datos, pero simultáneamente se le señalan unos términos que son los que va a emplear el sistema para buscar el documento.

Si bien en la Base de Datos de Legislación Financiera conviven los dos procedimientos, se le ha dado prioridad al segundo de los apuntados. El trabajo ha consistido en la representación del contenido de los documentos (artículos) por términos acuñados por el equipo de analistas-juristas, siguiendo unas reglas preestablecidas, como son la utilización de singulares, sustantivar verbos, construcción de sintagmas forzosos, creación de siglas y definición apriorística de sinónimos.

Todo lo anterior ha conducido a la creación de dos diccionarios de voces, estatal y autonómico, que podrían definirse como monodisciplinarios: ya que sólo se ocupan del derecho público bancario; arborescentes: porque todos los términos están asociados entre sí verticalmente, siguiendo criterios de jerarquía jurídica y de asociación; y, por último, automatizados, por cuanto están incorporados a la aplicación.

Ambos diccionarios facilitan al usuario dos posibilidades de interrogación. La primera, que podría denominarse jerárquica, y la segunda, alfabética.

La búsqueda jerárquica consiste en la selección de entre los grandes temas que dividen nuestra disciplina el que comprende la cuestión planteada. Y una vez seleccionado dicho tema, el usuario puede descender a través de determinados conceptos supeditados unos a otros y, por lo tanto, ocupando diferentes niveles, hasta llegar al deseado. Una vez aquí, el usuario tendrá definida su pregunta y, por lo tanto, el sistema propiciará su respuesta.

La opción de búsqueda alfabética procura un listado de términos ordenados alfabéticamente, en unos casos unitérminos y en otros términos compuestos y que en ocasiones dicen poco por sí mismos. Lo que realmente facilita el sistema en estos casos es que, tras ejecutar una determinada función, una misma voz va a aparecer asociada a múltiples temas, no teniendo el consultante más que elegir de entre todos ellos cuál es el que se ajusta a su pregunta.

Téngase en cuenta que un mismo descriptor en nuestros thesaurus puede ocupar niveles o rangos diferentes según el campo estudiado, ya que el criterio definitivo ha consistido en dotar a cada descriptor de una determinada categoría no por sí mismo sino por la transcendencia jurídica que tiene dentro del contenido del artículo.

No me resisto a finalizar esta ponencia sin exponerles dos ideas que me parecen fundamentales respecto del futuro de los bancos de datos jurídicos.

La primera es que todo planteamiento maximalista sobre las bases de datos que pretenda la creación de un enorme archivo automatizado, tratando de abarcar todo el ordenamiento legal de un país, está abocado al fracaso, si no es con un enorme esfuerzo técnico, económico y documental. Y aún así la empresa se nos antoja extraordinariamente difícil.

Por contra, es preferible un planteamiento más realista que podría tener su desarrollo en la creación de múltiples bases de datos monodisciplinares y que fuera llevada a cabo por instituciones u organismos especializados en cada caso.

La segunda idea, que viene a complementar la anterior, es que en un estado ideal de la situación, sería menester una mayor homogeneización en la estructura de las diferentes bases de datos y, sobre todo, en el lenguaje jurídico. En este último sentido, sabemos que se está llevando a cabo en nuestro país la elaboración de un macrothesaurus jurídico, con el apoyo de numerosas instituciones, entre ellas el Banco de España, precisamente dirigido a la tan ansiada armonización del lenguaje.