# LOS SISTEMAS DE INFORMACIÓN, LA INFORMÁTICA JURÍDICA Y EL SISTEMA UNAM-JURE

Sergio L. MATUTE C.*



## I. INTRODUCCIÓN

En una sociedad moderna el acceso ágil y preciso a información de la más variada naturaleza constituye, cada vez más, una necesidad de primordial importancia. Lo mismo para el investigador universitario como para el industrial, para el administrador como para el abogado litigante, la información oportuna adquiere un incalculable valor para el eficaz desempeño de sus actividades.

Algunas instituciones de nuestro sistema social, de hecho, se han convertido en instancias de proceso y administración de información. En efecto, considérense por ejemplo las instituciones bancarias de nuestros días; en el pasado un banco podía considerar que la "mercancía" básica para sus actividades lo constituía el dinero. Sin embargo, en la actualidad un banco puede afirmar que el elemento básico de su actividad lo constituye la información.

Por otra parte, en nuestra sociedad la toma de decisiones requiere, para su ejercicio, cantidades variables de información cuya complejidad y volumen dependen directamente de la naturaleza de la decisión a tomar, así como del medio ambiente que rodea al sistema afectado. Comúnmente, conforme se hace complejo el medio ambiente en que reside el sistema, se reduce la posibilidad de allegarse oportunamente

* Jefe del Departamento de Proyectos Especiales de la Dirección General de Servicios de Cómputo para la Administración de la UNAM.

Para clarificar las diferencias entre sistemas de recuperación de datos y sistemas de información documentaria en el terreno más práctico, recurramos a una serie de elementos presentes en ambos tipos de sistemas, pero que se diferencian en sus características intrínsecas.

En cuanto al algoritmo de selección de los *ítems* requeridos como respuesta a una consulta, tendremos que en el caso de los SRD la selección se logra por medio de una decisión absoluta; un registro es seleccionado si cumple con los criterios planteados en la consulta, la cual tiene la forma de una expresión lógica que, evaluada para cada registro, determina la pertenencia o no de éste al conjunto respuesta. Por otra parte, en un SID, la naturaleza misma de la información está en contra de la posibilidad de establecer una división precisa de los documentos incluidos en el conjunto respuesta, y aquellos que no lo serán; para algunos documentos el problema de establecer su relevancia en función de la consulta realizada resulta indefinido.[3]

En términos similares, los *ítems* requeridos por una consulta en el caso de los SRD, serán aquellos que satisfagan los criterios de la expresión de consulta, en tanto que los *ítems* requeridos en un SID serán los que resulten relevantes en función de la consulta planteada. El concepto de relevancia es uno de los conceptos principales en la recuperación de información documentaria, la evaluación de la relevancia de un documento respecto a una consulta dada, no puede establecerse como una función binaria *(i.e.,* cada documento del banco de información es o no relevante), sino como una clasificación del acervo documental (muy relevantes, relevantes, regularmente relevantes, etcétera).

Por lo anterior puede afirmarse que el modelo de los SRD es de tipo determinístico; de todos y cada uno de los registros se puede afirmar, una vez planteada la consulta, su pertenencia o no al conjunto respuesta. Por otra parte, en los SRI corresponden a un modelo no determinístico[4] en los que frecuentemente son usadas técnicas probabilísticas o lógica borrosa.

[3] Al respecto debe tomarse en cuenta que el propio lector humano encontrará dificultades para seleccionar de modo absoluto en un acervo documental los documentos relevantes a una cuestión dada. Existirán, sin duda, documentos que son claramente relevantes y otros que en definitiva no lo sean; pero habrá algunos documentos cuya relevancia sea intermedia.

[4] C.J. Van Rijsbergen, en *Information Retrieval* (1979, Ed. Buttherworth & Co., Londres), sostiene que se trata de un modelo probabilístico; sin embargo, hemos preferido el término no determinístico, ya que modernamente se presentan tipos de indeterminación diferentes como el caso de la lógica borrosa.

En cuanto al tipo de inferencia involucrado en la selección de *ítems*, en SRD se tiene una inferencia de tipo deductivo simple, esto es: si *A* está relacionado con *B*, y *B* está relacionado con *C*, entonces se deduce que *A* está relacionado con *C*. En SRI, es más común utilizar una inferencia de tipo inductivo donde las relaciones sólo son especificadas con un cierto grado de certidumbre.

Una importante distinción existe en términos del esquema de clasificación que podría establecerse para los *ítems* incluidos en una colección. Los registros involucrados en un SRD pueden ser clasificados de una forma absoluta dado que cada uno de éstos presenta los mismos atributos. En efecto, en la información de tipo dativo se trata con colecciones de elementos homogéneos que permiten establecer una estrategia de diseño en la que el concepto variable-valor resulta de primordial importancia; tómese en cuenta que una base de datos está formada por registros que, a su vez, están divididos en campos (variables) que contendrán los datos (valores). En términos de clasificación, por lo tanto, una base de datos podría ser clasificada de modo total atendiendo a los valores que tienen un determinado campo en cada registro, ya que todos y cada uno de los registros están formados por los mismos campos; esto es una clasificación de tipo tradicional aristotélica llamada monotética.

En cuanto a la posibilidad de clasificar de igual modo un acervo documental, esto resulta imposible debido a que cada documento tiene una serie de atributos que pueden o no tener significado para otros documentos. Es factible, por otra parte, clasificar una colección de documentos a través de un tipo de clasificación distinta, esto es, a partir de un conjunto de atributos y con las siguientes características: cada una de las clases resultantes estará formada por elementos que compartan un alto (aunque no especificado) número de atributos y cada uno de éstos estará presente en la mayoría de los documentos incluidos en la clase (aunque no en todos); este tipo de clasificación es llamada politética.

Un criterio de diferenciación más lo constituye el tipo de lenguaje de consulta empleado. Los SRD, por su propia naturaleza, tienden al uso de un lenguaje de tipo artificial, con una sintaxis restringida y un frecuente empleo de elementos simbólicos y de operadores lógicos relacionales. Por el contrario, los SID presentan una franca inclinación al uso de lenguaje natural como lenguaje de consulta en correspondencia con el lenguaje usado en los documentos.

En SRD la especificación de la consulta es casi siempre completa; en tanto que en el caso de los SID, ésta es invariablemente incompleta. Esto es debido en gran medida a que la búsqueda que se establece en este tipo de sistemas es por documentos relevantes, en contraposición a la selección completa que se practica para los SRD.

En la tabla siguiente se hace una lista de las propiedades distintivas entre los sistemas de recuperación de datos y los sistemas de información documental arriba discutidas.

| | SRD | SID |
|---|---|---|
| Selección | Exacta | Imprecisa |
| Inferencia | Deductiva | Inductiva |
| Modelo | Determinístico | No determinístico |
| Clasificación | Monotética | Politética |
| Lenguaje de consulta | Artificial | Natural |
| Especificación de la consulta | Completa | Incompleta |
| Ítems requeridos | Seleccionados | Relevantes |

Una primera preocupación de quien diseña sistemas de información deberá ser, sin duda, la caracterización del tipo de sistema a desarrollar. En algunos casos el sistema de recuperación de datos será suficiente, en tanto que en otros muchos deberá recurrirse a un sistema de información documentaria. En el primer caso, casi todos los problemas involucrados tienen una solución técnica factible y completa, en el segundo caso el diseñador se enfrentará a muchos problemas cuya solución óptima no ha sido alcanzada, antes bien, tendrá que recurrir a estrategias que resuelven en mayor o menor medida y con un variable grado de complejidad las dificultades planteadas, principalmente, por las características inherentes al lenguaje natural.

## 2. *Tipos de sistemas de información*

Desde la década de los años cincuenta, el problema de almacenamiento y recuperación de información documental ha atraído fuertemente la atención de los especialistas en ciencias computacionales; el planteamiento del problema, como ya se ha dicho, es simple: se cuenta en la actualidad con enormes acervos de información a los cuales un acceso rápido y preciso se vuelve cada vez más complejo. Con el advenimiento de las modernas computadoras una gran cantidad de esfuerzos han sido realizados para utilizarlas en la confección de sistemas de recuperación veloces e inteligentes. En las bibliotecas, muchas de las cuales realmente tienen problemas serios de acceso a la información, algunas tareas tales como la catalogación y clasificación han sido exitosamente llevadas a cabo por computadoras. Sin embargo, el problema de lograr una recuperación efectiva de la información ha permanecido en gran medida sin resolver.

El principio fundamental del proceso de almacenamiento y recuperación de la información es, en cierta forma, simple. Supóngase que existe un acervo de documentos y un individuo (usuario) que formula una consulta; ésta ha de ser respondida seleccionando un conjunto de documentos que satisfacen la necesidad de información expresada en la consulta. El usuario puede formar ese conjunto leyendo cada uno de los documentos del acervo y separando los que resulten relevantes. En cierto sentido esto constituye una recuperación de información "perfecta". Sin embargo, esta solución es, obviamente, impracticable. Un usuario seguramente no tendrá tiempo o ánimo para leer la colección entera, aparte del hecho de que, en muchos casos, resulta físicamente imposible lograrlo.

Cuando las computadoras modernas, con su alta capacidad y velocidad de proceso, se hicieron presentes en procesos de naturaleza nonumérica, muchos pensaron que sería posible hacer que la máquina pudiera leer grandes colecciones de documentos y extraer aquellos que resultasen de interés para una consulta dada.

Sin embargo, pronto resultó evidente que el uso del lenguaje natural en un documento obligaba a simular en la computadora el proceso intelectual de caracterización del contenido del documento. Éste resulta un problema sumamente complejo y, hasta la fecha, permanece sin una solución completa. El proceso de lectura inteligente involucra la extracción de información sintáctica y semántica, especialmente semántica, de un documento y el uso de esta información para decidir si el

documento es o no relevante en función de una consulta dada. El problema es, por lo tanto, doble; no sólo se debe determinar el modo de extraer información de un documento, sino establecer procedimientos automáticos de evaluación para determinar su relevancia. El desarrollo, comparativamente lento, de la lingüística en materia de semántica, así como el notable fracaso de la traducción automática, demuestran que este problema permanece, en gran medida, sin solución.

Ante la imposibilidad actual de simular de manera eficiente el proceso intelectual de lectura y comprensión de un texto, han sido planteadas diferentes estrategias para la elaboración de sistemas de recuperación de información que permitan tener acceso a un acervo documentario.

La adopción de una estrategia de recuperación de información en el diseño de un sistema debe ser motivo de una amplia reflexión por parte de sus diseñadores. Esta estrategia deberá corresponder en gran medida a la naturaleza misma de la información tratada, al propio lenguaje en que está expresada, e indudablemente a los objetivos que se persiguen con el sistema. Otras consideraciones deberán tomar en cuenta el medio ambiente en que residirá el sistema, el costo de implantación, el grado de especialización de los recursos humanos necesarios, etcétera.

Una estrategia de recuperación de información de tipo documentaria estará constituida por la determinación de un esquema de representación apropiado, por el planteamiento de instrumentos de tratamiento lingüístico y por el establecimiento de lenguajes de elaboración de los documentos y de las consultas. Todos los elementos antes mencionados están íntimamente relacionados con la forma de tratamiento del lenguaje natural y los conceptos abstractos expresados a través del mismo.

Por la forma de tratamiento del lenguaje los sistemas de información podrían ser clasificados en cuatro tipos básicos, no sin reconocer que muy frecuentemente en las implantaciones reales suelen practicarse estrategias combinadas.

Los sistemas del primer tipo serían considerados aquellos en los que la información contenida en el documento original es descrita a nivel superficial por medio de la asignación de descriptores, comúnmente tomados de colecciones preestablecidas. En este tipo de sistemas, el documento es analizado a nivel general y representado en la computadora, para efectos de recuperación, a través de un lenguaje artificial (asignación de descriptores). En todo caso, no se produce un trata-

miento del lenguaje natural propiamente dicho, antes bien éste se sustituye por la formación de colecciones de voces, frecuentemente *thesauri*, que pretenden estructurar los campos semánticos (normalmente temas y conceptos) presentes en el acervo documental a través de descriptores.[5] De modo similar, el usuario, al formular su consulta, lo hace seleccionando descriptores del mismo *thesaurus*. Esta práctica es realmente una catalogación de los documentos y es de destacarse el amplio arraigo y el desarrollo teórico que tiene este tipo de sistemas en el ramo de las ciencias bibliotecológicas.

El segundo tipo de sistemas adopta una estrategia de recuperación de información, cuyo soporte principal lo constituye el reconocimiento de las palabras contenidas en el documento. Independientemente de que se trate con el documento en su redacción original o con una reelaboración del mismo, éste es representado en la computadora por las palabras que lo constituyen. Una consulta será planteada al sistema en lenguaje libre de forma que una búsqueda, en los documentos, de las palabras usadas en la consulta resulte en la determinación de los documentos relevantes. A esta práctica se opone la inexistencia de una relación biunívoca entre los significantes (palabras) y los significados.[6] La existencia de accidentes lingüísticos como la polisemia (varios significados para una sola palabra), la sinonimia (varias palabras con el mismo significado), la hiponimia (palabras cuyo significado está subordinado a otro más general) y la analogía (palabras con significado similar, aunque no idéntico), por mencionar algunos, causa la posible omisión de documentos relevantes en el conjunto respuesta. Ante este inconveniente, diversos instrumentos de extensión lingüística suelen ser utilizados.

En el tercer tipo de sistemas de información, se desarrollan métodos de medición del grado de similitud (semático-sintáctica) entre los documentos para lograr respuestas de gran confiabilidad. Si bien, de nuevo se depende fundamentalmente de las palabras aparecidas en la representación documentaria utilizada, el cálculo del grado de semejanza atenúa en gran medida el efecto nocivo que causan los accidentes lingüísticos. Los sistemas de tercer tipo, aun en etapa experimental hoy en día,[7] de hecho aplican un interesante enfoque; no pretenden re-

[5] Una interesante exposición referente a la elaboración de *thesauri* catalográficos se puede encontrar en: Lancaster, F.W., *Thesaurus construction and use*, París, UNESCO, 1985.

[6] Lo contrario supondría la existencia de un solo significado para cada palabra, y de una sola palabra para cada significado, lo cual está lejos de ser cierto.

[7] La descripción de uno de los sistemas más conocidos de tercer tipo, el sistema

solver el problema netamente semántico, sino medir la semejanza entre los documentos. De este modo, una vez localizado un documento altamente relevante pueden ser recuperados otros similares, logrando un nivel de recuperación muy alto.

El cuarto tipo de sistemas de información está constituido por sistemas que no son propiamente sistemas de recuperación de información. En los tres tipos anteriores, en realidad no se informa al usuario, en el sentido más estricto *(i.e.,* cambiar el grado de conocimiento) de la materia consultada, sino de la existencia y localización de documentos relevantes a ésta. Por el contrario, en los sistemas de cuarto tipo se busca superar la sola localización de documentos ofreciendo al usuario respuestas concretas a sus consultas. Un caso particularmente estudiado por los especialistas en los últimos años lo constituyen los llamados sistemas expertos *(Expert Systems).* En un sistema experto, la información referente a un determinado tópico es preparada por un especialista en forma de reglas de decisión, éstas son alimentadas al sistema para constituir la llamada "base de conocimientos"; cuando un caso particular es presentado al sistema, éste aplica un mecanismo de inferencia a dichas reglas con los datos del caso particular hasta llegar a establecer una conclusión.

## III. La informática jurídica

La información jurídica tiene una especial importancia para la sociedad. El derecho cumple una función reguladora a través del establecimiento de reglas de las que derivan los derechos y obligaciones de las personas, tanto de las personas entre sí, cuanto en su relación con el poder público. Igualmente, el derecho tiene una función promocional por la que establece mecanismos y medios de los cuales pueden hacer uso las personas en su propio interés o de la sociedad. La existencia ubicua de un derecho ahí donde existe una sociedad, señala la importancia del derecho como factor fundamental de estabilidad y consolidación social.

Sin embargo, la existencia del derecho supone, para el cumplimiento de sus funciones, el conocimiento que la sociedad pueda tener del mismo, tanto el legislador como el poder público, el jurista como el ciudadano común. A este conocimiento se oponen diversas circunstancias,

SMART, se puede encontrar en: Salton, G., *et. al., Introduction to Modern Information Retrieval,* Singapore, McGraw-Hill, 1983.

muy especialmente el gran volumen de información jurídica que en la actualidad se produce.

El acelerado cambio social que caracteriza a la sociedad contemporánea se ve correspondido por un incremento en la actividad de los órganos de creación del derecho; si a este fenómeno se suma la creciente intervención del Estado en la vida social y económica, se podrá tener un panorama de los factores que motivan la incesante creación, modificación y supresión de normas jurídicas que actualmente se observa. Bastaría un breve examen de los órganos de publicación de disposiciones jurídicas de unas décadas a la fecha, para percatarse del desmedido aumento en la producción de normas.

Esta crisis de la información jurídica obliga a recurrir a modernos instrumentos de tratamiento de la información, en particular al uso de computadoras. Esta aplicación de la moderna informática a las características particulares de la información jurídica produce la aparición del campo, sumamente promisorio, de la informática jurídica.

## 1. *Orígenes*

La relación de la informática y el derecho, contra lo que pudiera parecer, se remonta a los primeros tiempos de la computación. Efectivamente, una aplicación que resulta muy probablemente la primera de una máquina de cálculo al ambiente jurídico, se dio en la Cámara de Representantes del Estado norteamericano de Ohio en 1938, con la utilización de una máquina de registro unitario[8] para el control y seguimiento de iniciativas de ley presentadas en la Cámara.

A nivel teórico, entre los primeros que consideran la aplicación de computadoras a la actividad jurídica se encuentran, por una parte, Norbert Wiener, el padre de la cibernética, quien consideraba al derecho como uno de los posibles campos de aplicación la cibernética,[9] y por otra, el jurista norteamericano Lee Loevinger, quien para referirse al uso de computadoras en el derecho, acuña en 1949 el término de jurimetría.[10]

[8] Las máquinas de registro unitario son el antecedente inmediato de las modernas computadoras electrónicas, fueron desarrolladas a partir de la experiencia de H. Hollereith en el cálculo del censo norteamericano de 1890, están caracterizadas por el uso de tarjetas perforadas.

[9] Wiener, N., *Cybernetics, or control and communication in the animal and the machine*, 1948.

[10] Loevinger, Lee, "Jurimetrics, the next step forward", *Minnesota Law Review*, XXXIII, 1949.

Sin embargo, no es el propio Loevinger quien realiza la formalización de la jurimetría como disciplina;[11] esta tarea sería llevada a cabo por Hans Baade hasta 1963, para quien la jurimetría contempla tres aspectos fundamentales:

*a)* Aplicación de modelos lógicos a normas jurídicas establecidas según criterios tradicionales.

*b)* Aplicación de la computadora a la actividad jurídica.

*c)* Prevención de futuras sentencias de los jueces.

En el primero de sus aspectos, la jurimetría pretende el establecimiento de métodos similares a los de la informática, en la creación, aplicación y estudio de las normas jurídicas; la generación de modelos lógico-matemáticos, así como el planteamiento estructurado de las normas jurídicas, forman parte de los estudios que comprende este aspecto.

La aplicación de la computadora al quehacer jurídico resulta el campo más amplio de los planteados por Baade, y sin duda el más práctico de ellos. En este terreno las realizaciones han sido muy importantes, como lo demuestra el surgimiento de la informática jurídica.

El último aspecto formulado por Baade resultó de hecho el más problemático de todos; se fundamentaba en las características que tiene la creación del derecho en un país de *Common Law,* como los Estados Unidos. En este sistema jurídico, el principio fundamental lo constituye el precedente jurisprudencial vinculante, según el cual la decisión de un juez sobre un caso concreto debe atenerse a las decisiones precedentes que han resuelto casos análogos. La idea de los jurimetristas era que el almacenamiento de un número estadísticamente significativo de sentencias de jueces en el pasado podría llevar a determinar cuál será la decisión del juez en un caso actual.[12]

En materia de realizaciones prácticas, la primera es el casi legendario sistema realizado en 1959 por el profesor John Horty, director del Health Law Center, en la Universidad de Pittsburgh. El profesor Horty, al compilar un manual jurídico sobre la legislación en materia de sanidad de los Estados Unidos, encuentra sumamente difícil conciliar los índices manuales, existentes para cada uno de los 50 estados de la Unión; por ello recurre al Centro de Cálculo de la Universidad de Pittsburgh creado apenas cuatro años antes. De esta manera se desarrolla el primer sistema de recuperación de información jurídica

[11] Se dice, incluso, que al preguntársele lo que era la jurimetría, Loevinger respondía diciendo: "es lo que hacen los jurimetristas".

[12] Una amplia discusión de los factores que hacen fallar esta tesis se puede encontrar en: Lozano G., Mario, *Introducción a la informática jurídica,* Facultad de Derecho de Palma de Mallorca, 1982 (traducción del italiano por M. Atienza).

en el mundo, al que en las décadas siguientes se vendrían a unir un número, cada vez mayor, de desarrollos tanto en los Estados Unidos como en el resto del mundo.

El sistema de Horty recurre al almacenamiento a texto completo de cada una de las leyes sanitarias de los estados de la Unión Americana, perforándolas en tarjetas y alimentándolas a la computadora. Como mecanismo de acceso recurre a un modelo de tabla de inversión similar a los utilizados hoy en día. El acceso al sistema se hacía igualmente perforando la consulta en tarjetas, el sistema operaba fuera de línea. El éxito del sistema fue tal que, en 1960, fue presentado a la barra americana de abogados y con el tiempo se comercializó a través de la Aspen System Corporation, creada para tal efecto.

## 2. *Conceptos*

La insatisfacción por los resultados concretos ofrecidos por la jurimetría y la presencia de instrumentos teóricos atractivos, como los ofrecidos por la cibernética teórica, hicieron que en Europa los estudios puramente empíricos de tipo loevingeriano se unieran con estudios del tipo netamente teórico, teniendo como resultado el surgimiento entre 1966 y 1969 de una disciplina que, a falta de un nombre específico, fue conocida simplemente como "Cibernética y derecho".[13]

Con la idea de llevar un poco de orden a una situación que metodológica y conceptualmente se volvía caótica, Mario G. Lozano propone, en 1968, sustituir el término jurimetría por el de "jusciberné-tica", la cual comprendería, según Lozano, dos ramas fundamentales: la modelística jurídica y la informática jurídica.

En términos prácticos, la informática jurídica ha sido el principal foco de atención de los interesados en la materia y el término es usado, cada vez con más frecuencia, para designar a nivel genérico la aplicación de la computadora al quehacer jurídico.

Atendiendo a la naturaleza de los implementos desarrollados, podemos establecer una clasificación de la informática jurídica en tres grandes divisiones:

1) Informática jurídica de gestión y control.
2) Informática jurídica documentaria.
3) Informática jurídica metadocumentaria.

La primera de las ramas que configuran la informática jurídica es la llamada de gestión y control. Son desarrollos propios de la misma,

[13] Lozano G., Mario, *op. ult. cit.*, p. 31.

aquellos productos informáticos especializados que apoyan la práctica del profesional del derecho. La creación de bases de datos, así como la elaboración de sistemas de cálculo y de clasificación, pertenecen a esta rama. Este es el caso, por ejemplo, del *software* desarrollado para llevar el control de asuntos contenciosos en una institución.

La segunda rama de la informática jurídica está constituida por la llamada informática jurídica documentaria; en ella es preocupación primordial el almacenamiento y recuperación automática de grandes acervos de información jurídica. La experiencia pionera de Horty recae en este campo. Las líneas de investigación de la informática jurídica documentaria se dan, principalmente, en desarrollos propios de *Information Retrieval* (ver *supra*), muy especialmente en el estudio del lenguaje propio de las ciencias jurídicas, con objeto de elaborar mejores instrumentos de apoyo al almacenamiento y recuperación de documentos de esta naturaleza.

Por último, la rama de más reciente desarrollo en el ámbito de la informática jurídica es la informática jurídica metadocumentaria. Uno de los casos más conspicuos en esta rama lo constituye la elaboración de sistemas expertos. Según muchos especialistas la información jurídica presenta un atractivo especial para la elaboración de sistemas expertos; de hecho, las reglas de decisión, que constituyen el núcleo de un sistema experto, encuentran un símil en las normas jurídicas. Los avances recientes en cuanto a la formalización de una lógica especial del derecho, la lógica deóntica,[14] resultan muy alentadores en este sentido.

## IV. El sistema UNAM-JURE

### 1. *Introducción*

La UNAM, consciente de la importancia que reviste el desarrollo de instrumentos apropiados de conocimiento del derecho, ha abordado la realización de proyectos en materia de informática jurídica. En efecto, desde el año de 1981 se ha desarrollado, a través del Instituto de Investigaciones Jurídicas y la Dirección General de Servicios de Cómputo para la Administración,[15] el proyecto UNAM-JURE, cuyo objetivo es

[14] Un excelente tratado de lógica deóntica puede encontrarse en: Von Wright, G.H., *Un ensayo de lógica deóntica y teoría general de la acción* (traducido al español por E. Garzón), México, UNAM, Instituto de Investigaciones Filosóficas, 1976.

[15] Cuando el proyecto fue iniciado, esta dependencia se denominaba Centro de Servicios de Cómputo (CSC); merced a las reestructuraciones de los servicios de cómputo

brindar un acceso ágil y preciso a la información jurídica nacional, a través de la elaboración de un sistema de información adecuado.[16]

El sistema UNAM-JURE actualmente comprende la información legislativa, tanto federal cuanto de cada uno de los estados, aparecida en los diferentes órganos de publicación legislativa desde diciembre de 1976 a la fecha, lo cual representa más de 18,000 fichas de información en línea con alrededor de 250,00 puntos de interés.

Igualmente se encuentra en desarrollo un proyecto complementario para la elaboración de un banco de información conteniendo las fichas de información legislativa del periodo comprendido de 1917 a 1976; los trabajos en este sentido se encuentran muy adelantados y pronto este nuevo desarrollo será puesto en servicio.

Además de los bancos legislativos se trabaja en la actualidad en la creación de un nuevo banco de información que contenga tesis de jurisprudencia, así como en la puesta en línea del *Diccionario Jurídico Mexicano.*

El sistema UNAM-JURE forma parte integral del Sistema Nacional de Información Legislativa que coordina la Secretaría de Gobernación y está siendo puesto a disposición de los gobiernos estatales y de las instituciones públicas para su consulta. Por otra parte, UNAM-JURE es consultable a través del Servicio de Consulta a Bancos de Información (SECOBI) dependiente del CONACYT.

## 2. *Estrategias para el tratamiento de la información*

Para elaborar los bancos de información mencionados, el primer problema a resolver fue la determinación de una estrategia y un esquema de representación documentaria adecuado al tipo de información a tratar. A este efecto se hizo una revisión de algunos de los bancos de información jurídica existentes en diversos países del mundo, especialmente de Estados Unidos, Francia e Italia.

Finalmente se decidió hacer una adaptación del esquema que el Institut du Recherche et d'Etudes pour le Traitement de l'Information Juridique (IRETIJ) de Montpellier, Francia, había aplicado exitosamente al tratamiento de la jurisprudencia francesa.

habidas en la Universidad, cambió su denominación, primero a Dirección de Cómputo para la Administración Central (1981) y después a su denominación actual (1985).

[16] Una descripción más completa del proyecto puede encontrarse en: Matute C., Sergio, *et al., El sistema UNAM-JURE, un banco de datos jurídicos,* Instituto de Investigaciones Jurídicas, UNAM, 1985.

En este esquema, llamado análisis de contenidos y caracterizado por la elaboración de *abstracts,* se pretende obtener ventajas por sobre un sistema de primer tipo (ver *supra)* y el método del texto integral[17] en un sistema de segundo tipo. En la redacción del *abstract* se lleva a cabo una reelaboración del documento original de forma que se logre tener toda la información, aunque no todo el texto original. Para ello, se persiguen tres metas básicas al realizar un *abstract:*[18]

1) *Reducción:* Se reduce el texto al mínimo posible, se eliminan redundancias, texto irrelevante, etcétera.

2) *Reordenación:* Los temas e ideas contenidos en un documento, suelen no estar ordenados conforme a un criterio metodológico estricto. Al realizarse el análisis, se elabora el *abstract* respetando una estructura lógica desarrollada especialmente para el tratamiento de información jurídica.

3) *Explicitación:* El redactor del documento original, con frecuencia hace referencia a un concepto sin llegar jamás a utilizar su denominación técnica precisa; un sencillo ejemplo lo es el uso de "Primer mandatario" para referirse al presidente de la República, o el uso de "Ley fundamental" por "Constitución política". Es tarea del analista el transmitir al *abstract* la explicitación de esos términos usados en forma implícita en el original.[19]

El tratamiento del lenguaje documentario, que resulta también un elemento de importancia primordial en un sistema de esta índole, se hizo a través de un léxico de dos niveles, de forma que se lograra, en el primer nivel de agrupamiento, amortiguar al máximo el efecto nocivo que causa la presencia del fenómeno llamado "variantes alotácticas"[20] en el lenguaje natural, y en el segundo nivel extender el cam-

[17] En el texto integral *(Full-Text)* se utiliza todo el texto del documento para lograr la representación, en la computadora, del mismo. Cada una de las palabras del documento fuente es tomada como punto de acceso al momento de recuperar los documentos relevantes a una consulta.

[18] Para una explicación a detalle se sugiere consultar: Cáceres N., Enrique, *El* abstract *legislativo* (tesis profesional), Facultad de Derecho, UNAM, 1983.

[19] Esta tarea no debe confundirse con la interpretación del documento original, lo que haría del *abstract* un documento subjetivo; debe ser una explicitación que conserve con toda objetividad los conceptos usados por el redactor original.

[20] Variantes de concordancia contextual en género, número y tiempo, pero cuyo sentido es comúnmente el mismo, sobre todo para efectos de recuperación de documentos; p.e., las frases "Se creó un instituto" y "Un instituto fue creado", son simples variantes redaccionales cuyo sentido semántico es el mismo, las palabras "creó" y "creado" son variantes alotácticas entre sí.

po semántico de las palabras usadas en la consulta a su sentido más amplio.[21]

El primer nivel de agrupamiento de las palabras se hace de forma que todos los sinónimos estrictos de una palabra queden agrupados en un solo grupo. Así, por ejemplo, si el usuario utiliza en su consulta la palabra "Creación", el sistema, a través del léxico, extiende automáticamente la búsqueda a palabras como: "Creado", "Creada", "Creará", "Creándose", etcétera, que son variantes alotácticas de la palabra original.

Estos conjuntos de palabras agrupadas en el primer nivel, son a su vez agrupados en familias que a nivel global comparten su significado. En este caso se encontrarían las familias de palabras "ley", "legal", "legislador", "legalizar", etcétera. La extensión de una palabra usada en la consulta del sistema a sus variantes de segundo nivel se hace a petición del usuario.

Una gran cantidad de sistemas de recuperación documentaria utilizan una estrategia de agrupamiento de variantes alotácticas por medio de operaciones de truncamiento de palabras. Esta práctica, aunque es mucho más simple, resulta más aplicable a idiomas anglosajones que a lenguas romances como el español o el francés.

El documento fuente del sistema UNAM-JURE lo constituye la ficha de análisis, la cual es elaborada por los especialistas a partir de los documentos fuente publicados en el *Diario Oficial.*

En cuanto a su estructura, la ficha de análisis está constituida de tres zonas básicas:

1) Zona de datos.
2) Zona de textos.
3) Zona de representación documentaria.

La zona de datos, que originalmente habíamos denominado "zona de campos fijos", contiene la información descriptiva externa del documento original, es decir, los datos que describen al documento sin atender propiamente a su contenido; estos datos son: número de referencia, asignado a la ficha especialmente para su control; fecha de publicación en el *Diario Oficial;* tipo de documento (ley, código, etcétera); nombre de la publicación, pues éste varía en los diferentes estados; procedencia geográfica, estado de la República del que proviene el documento; accidente de publicación, describe la condición del do-

[21] Una amplia exposición de estos conceptos puede encontrarla el lector en: *El sistema UNAM-JURE, documentos,* México, UNAM, Instituto de Investigaciones Jurídicas, 1984.

cumento de aparecido en un número extraordinario, en un alcance u otra variante.

La zona de textos contiene, en el caso de las fichas de información legislativa, un resumen, normalmente breve, del documento, de manera que el usuario, al leerlo, pueda darse una idea precisa de la posible relevancia del documento recuperado. Por lo anterior, esta zona había sido llamada originalmente "zona de resumen". En el caso de las fichas de información jurisprudencial esta zona resulta todo lo opuesto a un resumen, ya que ahí se ha incluido el texto completo de la tesis de jurisprudencia tal como fue publicada. Una estrategia similar será implantada para el caso del *Diccionario Jurídico Mexicano*. La zona de textos, para efectos de recuperación, no es una información "activa", el proceso de selección de los documentos relevantes no toma en consideración el contenido de esta zona al resolver una consulta.

La zona de representación documentaria está constituida por el *abstract* propiamente dicho. Constituye la zona de información "activa" para efectos de recuperación de información, pues el contenido de esta zona será el que determine la inclusión del documento en el conjunto respuesta inducido por una consulta. El *abstract* está formado por párrafos que a su vez se dividen en frases; esta estructura permite reconstruir el contexto de uso de las palabras y, en todo caso, evitar combinaciones no pertinentes al momento de la consulta (ver *infra*).

La que sigue es una ficha de información legislativa típica del sistema UNAM-JURE. La zona de datos puede identificarse en las tres primeras líneas; cada uno de los códigos correspondientes a procedencia geográfica, nombre de la publicación y tipo de documento han sido substituidos por el texto correspondiente. La zona del resumen viene a continuación de los campos fijos e incluye información respecto a las páginas del *Diario Oficial* de donde fue tomado el documento, así como la fecha de promulgación del mismo. La zona del *abstract* se puede identificar por la secuencia de caracteres '*/' que la inicia y la secuencia '/*' que la termina, la división de párrafos se señala por medio de los caracteres '/.' y la separación de frases por una diagonal simple. El *abstract* de la ficha presentada como ejemplo está compuesto por dos párrafos,[22] el primero dividido en 4 frases y el segundo en 5. Puede apreciarse que en la zona del *abstract* existe texto situado entre paréntesis, este texto es tratado como información "inactiva" a semejanza de la zona de textos y constituye un comentario al texto del *abstract*.

[22] Por restricciones de diseño el *abstract* puede constar de un máximo de 256

850001 14/ENE/85

FEDERAL DIARIO OFICIAL
CONSTITUCIÓN

CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS, ARTÍCULO 20 FRACCIÓN I, DECRETO DE REFORMAS. 1 ARTÍCULO, 1 TRANSITORIO. PÁGINAS 3 Y 4. 171284.

+ DETERMINA LOS REQUISITOS PARA LA OBTENCIÓN DE LA LIBERTAD PROVISIONAL EN EL PROCESO PENAL Y LAS AUTORIDADES IMPLICADAS EN EL PROCEDIMIENTO.

*/SG/CONSTITUCIÓN POLÍTICA DE LOS ESTADOS-UNIDOS-MEXICANOS, REFORMAS, ARTÍCULO 20, FRACCIÓN "I" (VIGENTE A LOS 6 MESES DE SU PUBLICACIÓN)/ JUICIO DE ORDEN CRIMINAL, PROCESO PENAL, ACUSADO, PROCESADO, GARANTÍAS INDIVIDUALES, DERECHOS MÍNIMOS/ LIBERTAD PROVISIONAL BAJO CAUCIÓN (OMITE MENCIÓN DE LA FIANZA), REQUISITOS, CIRCUNSTANCIAS PERSONALES, GRAVEDAD DE DELITO, MODALIDADES, SANCIÓN APLICABLE PARA SU PROCEDENCIA (QUE CORRESPONDA A UNA PENA NO MAYOR A 5 AÑOS LA MEDIA ARITMÉTICA)/.

/CONSTITUCIÓN POLÍTICA DE LOS ESTADOS-UNIDOS-MEXICANOS, REFORMAS, ARTÍCULO 20, FRACCIÓN "I"/ AUTORIDAD JUDICIAL, JUZGADOR, FACULTADES EN EL PROCESO PENAL/ DETERMINAR Y RECIBIR LA SUMA DE DINERO Y OTRA CAUCIÓN (NO LO RESTRINGE A HIPOTECA O CAUCIÓN PERSONAL COMO EL TEXTO ANTERIOR) PARA LA LIBERTAD PROVISIONAL DEL PROCESADO EN EL ORDEN CRIMINAL, MONTO MÁXIMO EN RELACIÓN AL SALARIO MÍNIMO VIGENTE EN EL LUGAR DEL DELITO (A 2 AÑOS) AMPLIACION (HASTA 4 AÑOS)/ DELITOS INTENCIONALES CONTRA EL PATRIMONIO, DETERMINA LA CAUCIÓN SUFICIENTE. REQUISITO PARA LA LIBERTAD PROVISIONAL CON BASE EN EL BENEFICIO, DAÑO O PERJUICIO ECONÓMICO OBTENIDO (3 VECES EL BENEFICIO)/ DELITO PRETERINTENCIONAL, O IMPRUDENCIAL, LIBERTAD PROVISIONAL,

párrafos, cada uno de los cuales de hasta 256 frases, las que pueden contener hasta 1,024 palabras. El banco de información admite poco más de un millón de fichas como máximo.

REQUISITOS, GARANTÍA DE LA REPARACIÓN DEL DAÑO O PERJUICIO, DETERMINACIÓN/*



### 3. *Modelo del sistema*

El sistema UNAM-JURE, en lo referente al aspecto técnico-computacional, ha sido estructurado en forma de tres subsistemas básicos: altas, consulta y edición y control.

El subsistema de altas se encarga de la inclusión de nuevas fichas en los bancos de información. Sus dos tareas principales son: analizar las fichas de información para determinar su ingreso o rechazo y, una vez aceptadas, realizar su inclusión.

En cuanto al análisis de las fichas de información, el trabajo principal es llevado a cabo por medio de un autómata aceptor[23] de la gramática, conforme a la cual son estructuradas las fichas; esto permite realizar una detección muy detallada de los errores que pudieran presentarse en la estructura de una ficha y realizar el reporte correspondiente. La inclusión de las fichas en la estructura del banco de información se lleva a cabo a través de mecanismos de actualización de cada una de las estructuras de datos de que consta el sistema.

El subsistema de consulta permite la selección de fichas relevantes a las consultas planteadas por parte del usuario. Para esto se ha establecido un lenguaje de consulta estructurado de forma que la construcción de cada comando resulte lo más lógica y cercana al lenguaje natural posible. El subsistema contiene los elementos necesarios para el control de usuarios, por lo que cada sesión es registrada en sus bitácoras.

El subsistema de edición y control permite, por su parte, el mantenimiento del banco de información, con la creación de nuevas familias de palabras en el léxico, eliminación de fichas, captura en línea de las mismas, etcétera.

[23] En la implantación del sistema se utilizó una máquina virtual llamada máquina de Mealy. Este autómata se caracteriza por la existencia de un alfabeto de salida además del alfabeto de entrada; la transición del autómata de un estado a otro conlleva la emisión de un elemento del alfabeto de salida. Haciendo que las salidas mencionadas gobernaran las acciones a realizar por el sistema, se logró un interesante modelo de programación dirigida por sintaxis; esta estrategia permite que sea la propia estructura de la ficha la que fije la secuencia de actividades del sistema, reduciendo en gran medida el grado de complejidad de la programación a realizar.

Por otra parte, en cuanto a las estructuras de datos del sistema UNAM-JURE, éste ha sido diseñado a partir de 5 archivos físicos básicos: GRID, ITEMS, LEXICAL, POINTERS y FIXEDF.

El archivo GRID es de hecho una base de datos[24] y contiene un registro por cada documento incluido en el sistema; los campos de esta base de datos están constituidos por los campos fijos definidos para el banco de información,[25] más otros valores especiales (apuntador al *abstract,* apuntador al resumen, etcétera).

En el archivo ITEMS se almacenan todos y cada uno de los textos de los documentos dados de alta, la implantación de este archivo es la de un archivo de textos. Es decir, un texto que se desee incluir en el archivo es insertado a continuación del último previamente insertado, usando para ello todos los registros que sea necesario. El texto puede ser reconstruido usando para ello los datos de dirección de inicio y longitud de texto que se almacenan en el registro correspondiente de GRID en forma de apuntador.

El archivo LEXICAL contiene el léxico del sistema en forma de registros en que han sido capturadas las palabras reconocibles por el sistema y registros que almacenan los datos de interrelación de las palabras para formar nociones y subnociones lingüísticas.

El sistema UNAM-JURE es un sistema de acceso por tabla de inversión; esta tabla reside en el archivo POINTERS, donde a cada conjunto de sinónimos en LEXICAL corresponde una colección de apuntadores. Cada apuntador es una identificación completa de cada ocurrencia de la palabra de que se trate en el sistema. Es decir, cada vez que aparezca, por ejemplo, la palabra "ley", un nuevo apuntador se agregará a los ya presentes para esa palabra, con los datos de documento, párrafo, frase y orden en que haya aparecido.

En el archivo FIXEDF se almacenan diferentes tablas de importancia para el sistema (tablas de valores de un campo, tablas de acceso, palabras nulas, etcétera).

[24] De hecho, existen tablas de acceso que permiten localizar los registros en la base de datos, así como tablas de equivalencias para los códigos empleados en algunos de los campos. El diseño de este archivo como una base de datos permite, entre otras cosas, la realización de procesos estadísticos y de control en relación con los documentos dados de alta.

[25] En el caso del banco de información legislativa, éstos son: número de referencia, tipo de documento, procedencia geográfica, fecha de publicación y accidentes de publicación (número extraordinario, alcance, etcétera).

4. *La consulta del sistema*

Sin lugar a dudas, la función más importante del sistema UNAM-JURE es permitir la consulta del material documental integrado a los diferentes bancos de información que lo componen. Esto se logra, principalmente, a través del comando de consulta, identificado por el uso del carácter "arroba" (@) al principio de la línea.

En la línea de consulta el usuario define la materia de los documentos que desea recuperar por medio de palabras cuya interrelación es establecida por medio de operadores especiales. Los operadores utilizados por el sistema son los comúnmente utilizados en sistemas de este tipo: disyunción, conjunción y diferencia (intersección, unión y diferencia en álgebra de conjuntos). Así, un usuario que desee formular una consulta sobre el tema del comercio internacional, lo hará de la manera siguiente:

@ COMERCIO * INTERNACIONAL

La operación, en este caso una disyunción, se lleva a cabo tomando como marco de referencia las frases de los documentos que integran el acervo, es decir, serán recuperados los documentos que contengan al menos una frase en que las palabras COMERCIO e INTERNACIONAL aparezcan simultáneamente. Siendo las consultas como la anterior las más usuales, se ha establecido como opcional el uso del asterisco (*).

Las operaciones de conjunción y diferencia se han implantado para ampliar o precisar la materia consultada; así, en una consulta como la siguiente:

@ CIRCULACIÓN DE (VEHÍCULOS, AUTOMÓVILES)

El sistema recuperará los documentos que, en alguna de sus frases, contenga las palabras CIRCULACIÓN Y VEHÍCULOS o las palabras CIRCULACIÓN y AUTOMÓVILES. Como ha podido observarse, el operador empleado es representado por una coma (,); sin embargo, existe la alternativa de utilizar un signo de más (+) con idéntico significado. La palabra "DE" se considera nula para efectos de recuperación [26] y los paréntesis son utilizados para establecer el orden correcto de ejecución de las operaciones. Por otra parte, la consulta siguiente:

[26] Las partículas como de, por, del, para, a, al, etcétera, son consideradas nulas

@ COMERCIO INTERNACIONAL - ADUANAS

El conjunto respuesta estará integrado por los documentos que en una de sus frases contenga las palabras COMERCIO e INTERNACIONAL, pero no la palabra ADUANAS. Se ha utilizado un guión (-) para representar la operación de diferencia.

Hasta ahora las operaciones del sistema han sido definidas a nivel de las frases contenidas en una ficha de información; no obstante, en el sistema UNAM-JURE es posible combinar operaciones a nivel párrafo con las ya mencionadas. De esta manera es posible realizar consultas como la siguiente:

@ CONSTITUCIÓN POLÍTICA AND DERECHOS HUMANOS

En esta consulta existen dos conceptos relacionados a través del operador de disyunción a nivel párrafo (AND); una consulta así produciría la recuperación de aquellos documentos que contengan un párrafo en donde las palabras CONSTITUCIÓN y POLÍTICA aparezcan en una misma frase y las palabras DERECHOS y HUMANOS ocurran en esa misma u otra frase, siempre dentro del mismo párrafo.

Los operadores a nivel párrafo son representados por palabras (AND disyunción, OR conjunción y EXCEPT o BUT para diferencia) a diferencia de los operadores a nivel frase que son representados por signos (* + -).

## V. Conclusiones

La experiencia producida por la elaboración del sistema UNAM-JURE abre un amplio espacio de reflexiones en torno a los sistemas de información y muy particularmente en materia de información jurídica.

La importancia que reviste hoy en día el control de grandes masas de información conlleva la necesidad de crear mejores instrumentos de acceso a la información. En la práctica, esto plantea el imperativo de establecer mejores mecanismos de procesamiento del lenguaje natural, lo que establece un estrecho vínculo con el desarrollo de la inteligencia artificial.[27] Ante este reto, al estudio específicamente lingüístico

dado que no tienen un efecto significativo en la determinación del campo semántico consultado, es decir, se parte de la hipótesis de que la respuesta del sistema sería la misma si estas partículas fueran incluidas.

[27] En la inteligencia artificial se reconocen tres sectores básicos: análisis y com-

deberá sumarse el desarrollo de instrumentos prácticos que permitan su procesamiento. En este terreno el desarrollo de la lógica borrosa, entre otros, resulta promisorio.

El desarrollo de modernos dispositivos de almacenamiento masivo, como el CD-ROM y otros, posibilita la construcción de bancos de información mixtos en los que el texto completo de un documento aparezca en la pantalla, a petición del usuario, independientemente de que la caracterización del documento en la máquina se haga mediante un esquema de representación documentaria diferente del texto integral. A este respecto cabe mencionar que en materia de informática jurídica con frecuencia se afirma que para la creación de bancos de información legislativa el texto integral es lo apropiado, arguyendo que el usuario querrá observar el texto tal cual fue publicado. Esta afirmación, sin embargo, no toma en cuenta la posibilidad anteriormente expuesta, es decir, la separación del elemento de trabajo para realizar la selección de documentos (la representación documentaria) y el texto a desplegar como respuesta.

El análisis y formalización de la lógica propia del derecho abre un interesante panorama en materia de informática jurídica metadocumentaria y enriquece, al mismo tiempo, la perspectiva de análisis de documentos jurídicos para la creación de bancos de información. El desarrollo de la lógica deóntica aparece como un factor muy promisorio en este sentido.

La informática jurídica, como disciplina joven que es, no ha sido formalizada del todo. Si bien los diferentes desarrollos prácticos que han sido realizados en esta materia le conceden una identidad, también es cierto que ha faltado un análisis y formalización de sus métodos y objetivos que le permitan ocupar un lugar entre las disciplinas formales. En los próximos años un trabajo teórico de esta especie permitirá que la informática jurídica tome un cuerpo definido, lo que facilitará su inclusión en la curricula del estudiante de derecho, así como la creación de estudios de posgrado en este sentido.

La informática jurídica, a pesar de ser un instrumento específico del derecho, supone para su desarrollo la concurrencia de especialistas en diferentes ramas, particularmente informática y derecho, desarrollando una labor llena de puntos de contacto. De hecho, la ignorancia de esta naturaleza interdisciplinaria de la informática jurídica ha constituido

prensión del lenguaje natural; resolución automática de problemas y demostración de teoremas; reconocimiento de formas y objetos en espacios no-dimensionales, con aplicación a la robótica.

el elemento subyacente a muchos de los fracasos más espectaculares que conocemos.

Esta labor interdisciplinaria ha sido una de las principales experiencias que nos ha dejado la realización del proyecto UNAM-JURE. El encuentro de especialistas en derecho e informática ha tenido, una vez establecidos los espacios de reflexión, un efecto superior a la suma de sus partes. El autor de estas líneas ve en ello la expresión de uno de los componentes principales del ideal universitario: el diálogo creativo entre diferentes ciencias, lo que rendirá en el futuro, estamos ciertos, mayores éxitos.