

**AUMENTANDO EL VALOR SEMÁNTICO  
DE LA INFORMACIÓN JURÍDICA.  
UN ENSAYO INFORMÁTICO**

*Carlos Loría Sáenz<sup>(\*)</sup>*

---

(\*) El autor es consultor, investigador y docente en el área de la Informática (en especial en razonamiento automático y lógica computacional) por muchos años en su mayoría con el ITCR. Tiene un doctorado en Informática y actualmente es colaborador en investigación y desarrollo de la empresa costarricense ArtInSoft ([www.artinsoft.com](http://www.artinsoft.com)) está vinculado al centro de investigación en computación que promueve el CENAT. Dirección Electrónica: [cloria@artinsoft.com](mailto:cloria@artinsoft.com). Teléfonos: 247-3217 (ArtInSoft), 370-1537 ó 261-4971.

## **SUMARIO:**

1. Resumen
  - Estructura general
  - Limitaciones, propiedad y reconocimientos
2. Explosión de información digital
3. Información versus conocimiento
4. De la macro visión a la micro visión
5. Un caso de análisis: el impacto en el ámbito judicial
6. Tendencias aplicables en el SCIJ
7. Conclusiones
8. Referencias

## 1. RESUMEN

Los sistemas informáticos que se pueden denominar en este contexto de corte *tradicional* tienden a *procesar* información para propósitos operacionales y comunes en situaciones bien estructuradas y relativamente rígidas. Transacciones bancarias o procedimientos de consulta o certificación de información son dos ejemplos comunes de este tipo de interacción. Los sistemas de esta clase corresponden a tecnología que se enfoca en la simbolización cuantitativa o numérica de los datos más que en su representación *semántica* o contenido *natural*, en el sentido de los seres humanos. Coyunturalmente se observa, en el ámbito global, un crecimiento realmente explosivo en datos e información accesibles en medios y sitios digitales. Por encima de los obvios problemas de almacenamiento, este crecimiento representa además, en muchas situaciones interesantes, un verdadero reto para las capacidades del procesamiento de datos tradicional, pues las cantidades de *información cruda* pueden ser tales que el esfuerzo humano para *digerirla* sea una limitante a la eficacia de los sistemas, por lo menos para aquellos que quieran o deban sobrepasar los límites de procesamiento tradicional.

El incorporar en los sistemas aquellas funcionalidades que faciliten los procesos de búsqueda de información textual es ciertamente una evolución y ayuda muy importantes; aún así, en situaciones complejas, otras formas aún más refinadas pueden ser requeridas. Ejemplos de tales situaciones corresponden a tareas de análisis y síntesis de información no tan estructurada para brindarles a los seres humanos mayor agilidad en ciertos trámites y tareas decisorias en su interacción con las computadoras. Las tendencias actuales para atacar este problema giran alrededor de crear e incorporar una mayor cantidad de conocimiento digital (por encima de *información digital cruda*) en los sistemas computacionales. Para conseguir esto, los sistemas deben poder poseer y *manejar* más *información semántica* (es decir conocimiento) acerca de los datos que procesan y las *reglas de razonamiento* que deben seguir para procesarlos. Para que los sistemas, por encima de sólo procesar, puedan también ayudar a *razonar*.

Este ensayo es una reflexión introductoria sobre esta clase de contexto y tendencias con el fin de evidenciar y resaltar el valor potencial de que se pongan en marcha iniciativas concretas en lo que debe ser una transición hacia un mayor manejo semántico de la información digital; en especial en aquellas situaciones (locales) donde sea realmente importante y de impacto social. Si bien las bases del análisis son

generales y livianas, se usa el ámbito legal como punto de referencia por considerarse un caso indicativo para los propósitos de este ensayo.

### – **Estructura general**

El contenido de este ensayo se resume y estructura de la siguiente manera: Una breve visión y la problemática del acelerado crecimiento a nivel global de fuentes de información cruda es presentada como marco de contexto y de tendencias generales. La distinción entre información y conocimiento digital es clarificada así como las restricciones de los enfoques informáticos convencionales. Esta distinción resalta las limitaciones visibles aún en sistemas de búsqueda basados esencialmente en información textual. La visión de manejo semántico es analizada intuitivamente en el marco del ámbito de información legal y judicial, exponiendo la tesis acerca de la necesidad de un paradigma semántico en su concepción estratégica. El caso del Sistema Costarricense de Información Jurídica (SCIJ) es usado como plataforma para reforzar aún más esta tesis, lo que permite derivar conclusiones concretas para un posible aprovechamiento sobre las ideas ensayadas.

### – **Limitaciones, propiedad y reconocimientos**

El material acá expuesto fue elaborado, en la medida de lo posible, con un enfoque más filosófico que técnico, manteniéndose dentro de los límites usuales de un ensayo. Más detalles específicos en el campo técnico pueden ser solicitados directamente al autor o fuentes indicadas. El uso del ámbito legal o judicial como tema de referencia es simplemente ilustrativo, dado su gran valor e impacto; no debe presumirse que el autor posee conocimientos más allá que los de un nivel intuitivo en dicho campo. Se ofrecen las disculpas del caso por cualquier uso incorrecto en esa materia.

El contenido de este ensayo puede ser utilizado únicamente con fines académicos o no comerciales siempre que se conserve y se explicita la fuente.

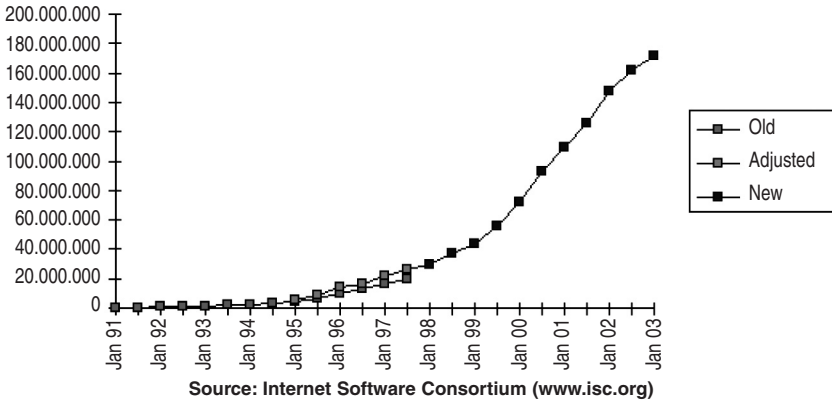
Se agradece a los funcionarios informáticos del Poder Judicial MSc. Fabio Muñoz y MSc. Fernando Ruiz por su apertura, sugerencias y motivación alrededor de este ensayo. La idea de probar el concepto acá expuesto alrededor del SCIJ por medio de un ‘asistente de sentencia’ proviene de conversaciones con ellos.

## 2. EXPLOSIÓN DE INFORMACIÓN DIGITAL

Es más que palpable en nuestra época el que realmente gigantescos volúmenes de información digitalizada estén disponibles a través de páginas en sitios de redes privadas o públicas en el *Web*. Diversas estimaciones estadísticas acerca de las tasas de crecimiento de la *Web* son evidencia de la magnitud de este gran crecimiento.

Por ejemplo, la gráfica de la Figura 1 muestra los volúmenes y tendencias de crecimiento en cantidad de direcciones accesibles en sitios de Internet (fue tomado de la encuesta del *Internet Software Consortium* <http://www.isc.org/>). Al inicio del 2003 se miden más de 170 millones de direcciones. Si bien la curva<sup>(1)</sup> de crecimiento no acelera de manera creciente como fue al inicio, es claro que los volúmenes alcanzados se traducen en millones de ventanas para potenciales enormes repositorios de servicios de información actuales y futuros.

**FIGURA 1**  
**Internet Domain Survey Host Count**



(1) Los tres tipos distintos de curvas se deben a cambios en la metodología de medición.

Aunque no se muestra directamente en los datos, es de suponer que, al menos, similares tendencias gobiernen el crecimiento de requerimientos de fuentes y servicios de información digital que las direcciones de sitios proporcionan ahora o deban proporcionar en el futuro.

Otras fuentes de datos interesantes relacionados con el crecimiento del acceso a información digital se pueden inferir indirectamente de estudios acerca de la adopción de servicios del denominado enfoque de *Gobierno Digital* (*Government online*, o *e-government*). Por ejemplo, entre otros datos, la encuesta de la firma consultora *Taylor Nelson Sofres* (disponible en <http://www.tnssofres.com/gostudy2002/>) propone que en el 2002, en promedio, un 30% de la población adulta de los países encuestados (el estudio no incluye Latinoamérica) accede a esos servicios (que representa un 4% de incremento sobre el promedio del 2001). En una buena parte de esos países la adopción refleja incrementos de entre un 13% hasta un 17% entre esos dos períodos de tiempo. La encuesta demuestra así la creciente adopción del acceso a sitios de trámite gubernamental que sugiere una mejora de la apreciación ciudadana por este tipo de facilidad informática.

### **3. INFORMACIÓN VERSUS CONOCIMIENTO**

No está con claridad establecido, dentro de las tendencias reseñadas sobre crecimiento de datos e información digital, cuál sea la correspondiente proporción de aprovechamiento *'inteligente'* que las organizaciones o los individuos estén logrando o puedan lograr en el futuro a partir de tales magnitudes de información; es decir, aprovechamiento por ejemplo, en procesos de ayuda a la toma de decisiones más eficaces, o incluso en el mejoramiento del entorno y la calidad del servicio informático sea hacia lo interno o externo de las organizaciones e individuos consumidores y proveedores de la información.

Tal aprovechamiento es entendido en este documento un nivel más allá del procesamiento de datos operativo y convencional que es usual en el quehacer rutinario informático, como puede ser por ejemplo, el asociado al usar una tarjeta de crédito o al consultar el estado de avance de un proceso administrativo o al realizar la declaración o pago de la renta electrónicamente; o aún entre otros procesos descentralizados electrónicamente entre sistemas computacionales. Todos estos son el resultado de una importante y valiosa infraestructura informática pero

esencialmente de *procesamiento* de datos e información. Como se quiere enfatizar reiteradamente en este ensayo, algo más que procesamiento e información es, en general, necesario en una informática basada en semántica.

Si aceptamos entonces que la información digital, digamos *cruda*, crece enormemente producto de la adopción del *Web*, es sin embargo, probable que la transformación de esta información en un mayor nivel de *conocimiento digital* no tenga lugar todavía tan uniforme y vertiginosamente aún. Esto sería así quizás porque, en general, los mayores esfuerzos informáticos históricamente se han concentrado más alrededor de la construcción de sistemas y bases de ‘datos’ y de ‘información’ que de ‘conocimiento’. Esto deriva en muchos casos en que, a pesar de haber más datos digitales, exista paradójicamente menos conocimiento digital en los sistemas; lo que significaría un margen de desaprovechamiento potencial de uno de los activos y paradigmas estratégicos organizacionales más valiosos de nuestros tiempos, cual es el *conocimiento digital* y la posibilidad de manejarlo computacionalmente. Tal desaprovechamiento será agravado aún más en función de la velocidad y los mismos volúmenes de crecimiento de la información digital cruda.

El objetivo de este ensayo es proveer un conjunto de elementos de análisis que permitan una reflexión sobre esta hipótesis y tratar de describir alternativas y tendencias tecnológicas que están siendo impulsadas para comenzar a mitigar sus efectos.

Para comenzar, es necesario plantear una pregunta aclaratoria la que es central en este contexto: ¿Es el uso de la palabra ‘*conocimiento*’ en vez de ‘*información*’ un simple juego de palabras? En búsqueda de mayor claridad sobre este aspecto, puede abrirse un paréntesis para ilustrar la distinción que estará siendo empleada de forma tácita y frecuente en este documento.

Aunque se trata de conceptos abstractos no siempre distinguibles de manera precisa, es usual en el campo informático referirse a la *información* como aquellas representaciones de datos relacionados que describen *realidades* o *dominios* los que tienen algún *sentido* e interés para un observador que los percibe. Son usualmente *factuales*, *extensionales*, se relacionan con formas de procesamiento que permiten respuestas a preguntas como *qué*, *quién*, *cuándo*, *dónde*, *cuántos*, usualmente determinadas por la misma estructura de los datos. Por ejemplo la información de quién es o fue el juez de un juicio es información sobre el expediente de un procedimiento.

El *conocimiento* por su parte representa más el nivel de la intención (es decir el *sentido*) de cierta información disponible, el por qué de las relaciones, la causalidad entre la misma. El conocimiento permite la comprensión y el aprendizaje, la generalización o especialización. El procesamiento del conocimiento es deductivo, *inferencial*, lógico, hipotético, decisorio; contesta preguntas de la forma: *cómo, qué-pasa-si, por qué*.

Por ejemplo, caen en la categoría de conocimiento los *criterios* (reglas de razonamiento lógico) que permiten a un analista clasificar una conclusión asociada a una sentencia digamos como *nueva* jurisprudencia en el campo constitucional. Ese mismo conocimiento puede ayudar a comparar normativas en términos de sus coincidencias o contradicciones, por ejemplo.

El reflexionar sobre este fenómeno no es un asunto puramente filosófico y académico, sus implicaciones informáticas pueden ilustrarse por medio del siguiente simple ejemplo cotidiano en un ámbito escolar o familiar.

Suponga para eso que se quisiera consultar a un sistema buscador en el *Web* datos acerca de algo como el *Jefe de Estado de Costa Rica*. Un nivel de conocimiento *natural* debería permitir inferir que (en Costa Rica) el *Jefe de Estado* y el *Presidente* son conceptos equivalentes y por ende denotan el mismo conjunto de datos. Sin embargo, normalmente las denominadas máquinas de búsqueda textual no ‘razonan’ así, semánticamente sino sintáctica y estadísticamente; la Figura 2 muestra el producto de hacer la consulta con uno u otro término (que verifica que la indización de documentos no conocen la citada equivalencia).

En dicha figura puede apreciarse que la consulta de la izquierda (usando ‘jefe de estado’) arroja distintos resultados con relación a los de la derecha (usando ‘presidente’). De hecho, en este último caso algunos de los resultados (asociado con un hotel) no están (probablemente) relacionados con la intención (es decir la semántica) de la consulta que estaba tratando de evacuarse.

Si se quisiera reflexionar más usando este ejemplo, uno podría pensar que tal vez la intención específica del usuario de la consulta tal vez era, por ejemplo, indagar acerca de ‘*los deberes constitucionales del presidente de Costa Rica*’. Al hacer esta consulta por medios no



semánticos los resultados no son necesariamente todo lo esperado,<sup>(2)</sup> como se puede confirmar, ya sea usando ‘presidente’ o ‘jefe de estado’. Una consulta basada en razonamiento semántico debería identificar con mejor precisión los conceptos y relaciones (‘deber’, ‘constitucional’, ‘presidente’) en juego y de ahí moverse en un espacio de búsqueda semánticamente más específico.

En este caso (como en muchos casos similares pero normales), el no tener conocimiento acerca de una relación semántica entre los términos en juego, aún una relativamente simple, puede implicar obtener información incompleta o no relacionada si se opta por un camino o el otro. Las consecuencias en casos más relevantes y críticos pueden ser fácilmente inducidas.

Esta clase de ejemplo debe ayudar a visualizar mejor lo que se entiende en este ensayo por manejo semántico de información: uno que permita acceder, capturar y crear conocimiento digital en un sentido más amplio y natural. Se puede usar este enfoque para elaborar más sobre la evolución y tendencias importantes en el campo informático en las siguientes líneas.

FIGURA 2



(2) Se omite por razones de espacio la muestra de esos resultados.

En general, las *primeras generaciones* de sistemas de información digital (orientadas por razones pragmáticas<sup>(3)</sup> a información más que a conocimiento) abarcaron sustancialmente los procesos informáticos administrativos, financieros y similares tipos de operaciones fundamentales del quehacer propio de las organizaciones. Algunas organizaciones (por ejemplo en nuestro medio) de hecho todavía están superando esta etapa, es decir aumentando el ámbito su capacidad de procesamiento de información aún dentro de las fronteras de esta primera generación.

La tendencia informática actual es incrementar en lo posible el acceso a la información a través de la *Web* por medio de arquitecturas distribuidas, protocolos, formatos y mecanismos de representación estandarizados y portables. Esto ha dado lugar a una *segunda generación* de sistemas de información que exponen sus servicios (por ejemplo, de primera generación) a sus clientes o usuarios externos o internos (incluyendo otros sistemas computacionales) mediante sitios y aplicaciones del *Web*. Estos servicios habilitan, por medio de interacciones estructuradas, trámites y transacciones electrónicas *regulares* para la obtención de servicios o recursos bien definidos y establecidos.

Un paso más allá es necesario cuando el servicio o las fuentes de información no están del todo o al menos suficientemente estructuradas, por ejemplo, cuando esas fuentes son en buena medida *textuales* (es decir en lenguaje natural) o heterogéneas y su volumen es alto y semánticamente complejo; el estado actual del desarrollo se concentra en poder facilitar la recuperación de la información por medio de búsquedas eficientes (usualmente de naturaleza indexada o estadística) para llegar a obtener posibles fuentes candidatas de información.

Aunque las técnicas de indización que dan soporte a la búsqueda textual son en verdad bastante acertadas, eficientes y valiosas, el tipo de procesamiento, o más bien dicho el razonamiento, requerido para poder determinar la (ir)relevancia de tales fuentes, o la (in)completitud o la fortaleza (debilidad) de su interrelación con el área o problema de interés sigue siendo fundamentalmente un espacio de esfuerzo y trabajo para el usuario de la información.

---

(3) Dado el estado tecnológico y de necesidades al momento de su concepción, usualmente.

Un razonamiento basado en semántica<sup>(4)</sup> proporciona un mejor soporte a esta tarea ‘intelectual’ de evaluar y filtrar la información en casos de búsquedas complejas, no frecuentes o no estructuradas. Dado un requerimiento de información que un usuario tiene y sobre el que quiere consultar a una base de datos, si un mecanismo de inferencia computacional ‘dispone’ de una estructura de conocimiento rica semánticamente, entonces no necesita forzar al usuario (uno quizás no especialista) a estructurar su consulta demasiado o a reformularla en un vocabulario determinado, a más aún reducirla a llenar ‘formularios de captura de datos’ predeterminados a los que se tiene que ajustar.

El sistema puede por sí mismo compilar su consulta en los términos y conceptos que él conoce aliviando al usuario de esta tarea.

Por ejemplo, un hipotético usuario podría querer formular una consulta (no estructurada) a una base de conocimiento legal, requiriendo exactamente lo indicado a continuación:

*“Sentencias en litigios de separación matrimonial en las que la potestad de los hijos haya sido concedida al padre”.*

En este tipo de ejemplo, si un sistema computacional no estuviese adecuado a un razonamiento semántico, el usuario debe primero *traducir* la consulta a la estructura rígida (usando formularios) o recurrir a consultas libres, con el problema de que tal vez esté usando un vocabulario de ‘sentido común’ (por ejemplo, si no es un experto en el vocabulario técnico concreto y en vez de ‘matrimonial’ debería usar un ‘hiperónimo’<sup>(5)</sup> como podría ser la palabra ‘conyugal’) y por esta razón no obtenga tan fácilmente todo lo que esperaría o debería poder obtener.

Como se puede extrapolar a partir de este ejemplo pensando en situaciones más realistas, un tipo similar de limitación puede ser realmente problemática y entabada en la medida que el quehacer natural de una organización esté girando alrededor de tareas basadas en razonamiento complejo, más analítico, por encima del procesamiento de información (como por ejemplo, podría ser el caso del ámbito jurídico).

---

(4) Por ejemplo, uno que tenga conocimiento de la materia que se busca y de su estructura de contenido.

(5) Queriendo decir una palabra que es más general pero de una misma categoría. Por ejemplo ‘moverse’ es hiperónimo de ‘correr’.

El siguiente paso en la evolución de los sistemas y las redes de información digital contemplaría el que la masa de información cruda pueda ser cada vez más accesible, convertible y digerible en la forma de conocimiento, de alto valor y sentido de oportunidad para las condiciones esperadas en una *situación informática* compleja o poco frecuente. Situaciones que son esperables en especial cuando cada vez más información rica semánticamente es abierta a una diversa variedad de usuarios en tareas de naturaleza intelectual.

Fuertes iniciativas al nivel mundial basadas en la visión presentada en la, ya famosa, propuesta de la “*Semantic Web*” de *Tim Berners-Lee*, pretenden compensar la desproporción (o brecha) entre información digital y conocimiento digital por medio de la introducción de mayor contenido semántico en el *Web*; esto es estandarizando, incorporando y aumentando la denominada ‘*meta información*’ digital, es decir aquella que recurrentemente sirva para describir la misma información digital disponible en las innumerables páginas y sitios del *Web*; para facilitar así su gestión, el acceso, localización, intercambio y evolución por encima de barreras semánticas; y he aquí lo más sustancial, que ese acceso se mejore en el mismo tanto para los seres humanos como también para *agentes computacionales* inteligentes.

Mucha tecnología e investigación alrededor de lenguajes para *Representación de Conocimiento, de Lógica Computacional, de Inteligencia Artificial, Procesamiento de Lenguaje Natural, de Computación Distribuida* entre otras muchas ramas, están siendo conjuntadas, adaptadas y (en especial) estandarizadas para poder disponer de repositorios de conocimiento globales distribuidos, interoperables, dinámicos e inteligentes. Con énfasis especial en que la disponibilidad de información debe existir tanto para los seres humanos como para los mismos sistemas computacionales.

Lo anterior se considera fundamental, pues debe ser evidente a esta altura que, dados los crecientes volúmenes de datos disponibles (como por ejemplo los que son observables en el *Web*) se estará en algún momento llegando a puntos donde, sólo por medio de la ayuda de la misma tecnología computacional, sea efectivamente posible seleccionar y filtrar (sin que por eso se incurra en pérdida de información valiosa) y realmente aprovechar aquellas fuentes que sean verdaderamente relevantes para ayudar a resolver un problema de interés particular, sea este de índole social, legal, comercial, académico o científico.

#### **4. DE LA MACRO VISIÓN A LA MICRO VISIÓN**

Aceptemos que toda esta visión macro, acerca de la necesidad de un manejo más semántico de la información digital y que ha sido delineada en párrafos previos, sea en verdad una iniciativa que debería ser considerada como relativamente inminente en el futuro de las sociedades digitales, en la escala mundial; resulta en ese contexto tecnológico altamente interesante realizar el ejercicio de evaluar y proyectar las ideas (aunque en nuestras escalas menores) hacia nuestro medio por ejemplo, en búsqueda de patrones organizacionales favorables o desfavorables hacia este tipo de visión.

Esto es, en los ambientes más locales, específicos y concretos de aquellas organizaciones que por su complejidad inherente dependen de un manejo computacional mucho más *inteligente* (menos operacional) de su información; aquellas en las que las computadoras podrían ofrecer formas de *razonamiento computacional* por encima de un nivel procesamiento de información; uno que ofrezca a sus usuarios la disponibilidad de vías más eficaces y flexibles de acceder a los servicios y beneficios propios de la función que le da sentido funcional a toda organización.

No es difícil imaginarse que los entes y organizaciones públicas (estatales) aparecen en este sentido como excelentes candidatos para el ejercicio propuesto; esto es así por la cantidad y calidad de información y conocimiento vital de interés público y social que por su naturaleza administran o custodian, por los procedimientos regulatorios que deben garantizar que se cumplan y muchas otras razones similares de peso.

Es evidente que, por las limitaciones de este trabajo, no se puede pretender profundizar para pensar en llevar a cabo un diagnóstico formal desde esta perspectiva del sistema estatal; eso puede ser con seguridad motivo suficiente para investigaciones derivadas específicas. Sin embargo, dado que el valor potencial de un manejo semántico de información puede ser mejor visualizado en los términos concretos de un caso, se considera valioso tocar superficialmente uno en particular: el objetivo se limita a tratar de resaltar las ventajas y oportunidades que presenta el ámbito de información jurídica para el tipo de análisis de este documento.

#### **5. UN CASO DE ANÁLISIS: EL IMPACTO EN EL ÁMBITO JUDICIAL**

Como persona común uno obviamente considera que el Sistema Jurídico de una sociedad organizada representa, por su definición, un pilar fundamental para los alcances e ideales de la justicia.

Bajo tales principios (y aceptando el sesgo que acompaña a una quizás *simplista* perspectiva '*techo-informática*') la búsqueda de la justicia debería poder ser aún más eficaz y eficiente en la medida que mayores niveles –basados en *conocimiento digital*– de la ley, su función y administración sean creados y brinden mejor acceso a aquellos que son encargados de hacer posible la difusión, administración y aplicación eficiente de la ley. No sea posible esto *sólo* con este tipo de tecnología, pero probablemente sea imposible *sin* la introducción de formas cada vez más avanzadas de ella.

Sin ignorar todas las limitaciones coyunturales del contexto que rodea a una sociedad como la nuestra, las tecnologías computacionales actuales pueden ser, sin duda, una importante pieza en la promoción del objetivo de continuar el mejoramiento del acceso inteligente a la información legal, pero asimismo incrementando la creación de conocimiento digital de naturaleza legal. Esto hacia lo interno de la organización judicial como a hacia el entorno de sus usuarios externos en los niveles que a estos les compete.

Bajo esta premisa, y por la alta sensibilidad social que ostenta la información jurídica, sería evidentemente más que deseable poder empezar a evolucionar la informática judicial hacia un aprovechamiento computacional cada vez más refinado e inteligente de su invaluable inventario cognoscitivo. Un enorme reto estratégico y tecnológico al que seguramente le esté llegando su mejor momento.

Esto en especial en tiempos en que las demandas sobre cualquier sistema de administración de la justicia son muy elevadas y quizás las capacidades de mejoramiento informático no estuviesen totalmente orientadas a los propósitos acá presentados. Tanto por su volumen como por su compleja naturaleza semántica, la que presupone capacidades de razonamiento de los sistemas basadas en tecnología del conocimiento.

La materia legal y jurídica, en general, resulta especialmente interesante desde el contexto del mejoramiento informático por muchas excitantes razones. Varias de esas razones evidencian concordancias con los sistemas computacionales, en especial aquellos que aspiran a ser basados en conocimiento digital. Sobre estas concordancias vale la pena reflexionar un poco por medio de una analogía.

Por ejemplo, vistas sólo de manera *funcional* (que es la usual en la mente del informático) las leyes constituyen formalmente *modelos*

que buscan abstraer, capturar y describir los propósitos, los razonamientos y lógica, las vías, los alcances, las responsabilidades, etc. que son convencionalmente aceptados como *válidos* o *justos*, mediante los cuales los entes de una sociedad deberían interactuar racionalmente.

Los sistemas computacionales son también modelos (en la misma dimensión que lo son las leyes) reflejan partes de lo que conocemos de cierta realidad en forma digital. Algunos, y cada vez más, sistemas computacionales y sus bases de datos también regulan, limitan o autorizan ciertas acciones que ocurren entre los individuos u organizaciones. Son en muchos casos implementaciones de regulaciones y conocimientos legales, que son denotados y *compilados* en el lenguaje computacional.<sup>(6)</sup>

En la actualidad en las organizaciones informatizadas cualquier regla operativa (o '*del negocio*' que se estila denominar) o política o procedimiento, usualmente va acompañada de un sistema computacional (o al menos un cambio al mismo) que debe ayudar a su realización y puesta en práctica de forma eficiente. Esto se considera *natural* en muchos campos convencionales organizacionales, pues esas *reglas* son conocimiento humano que distingue y mejora el ejercicio organizacional. Estas reglas deben también ser parte inherente del conocimiento digital.

A la luz de esta analogía, entre marco legal y sistema computacional, se vislumbra una base analítica interesante en el contexto del manejo semántico de información legal. Esta base permite levantar una tesis muy interesante sobre la clase de estructura cognoscitiva que debe usarse para representar el conocimiento legal del ámbito jurídico, para que se facilite su manejo computacional inteligente. Podemos entonces reflexionar y establecer algunos cuestionamientos:<sup>(7)</sup>

- ¿Deberían la materia legal y los sistemas computacionales del ámbito jurídico coincidir cada vez más, y así los marcos legales disponer de un correspondiente marco *computacional referente*, uno fuertemente basado en conocimiento digital, que sirva para

---

(6) Usual, y lamentablemente, más en una forma procedimental que en una declarativa.

(7) Que corresponden a expectativas y enfoques ya maduras en los campos legal e informático por varios autores, por ejemplo ver la referencia de J. Breuker (basado en M. Sergot).

umentar su consulta, comprensión, aplicación, difusión, validez, cohesión, su acceso e interrelación con y desde otros marcos legales o no legales incluso?

- ¿Debe una mayor formalización computacional de la semántica y conocimiento legal ser empleada para esta clase de propósito?
- ¿Sería esta una forma efectiva, y bajo cuáles condiciones, de ayudar a agilizar los procesos de la administración de la justicia?

Sin intentar entrar en detalles técnicos de realización, bajo una visión idealizada el usuario (especializado o no) de una fuente de información legal debería poder interactuar con ella, por ejemplo, con una sentencia digital, como si esta fuese una base de *conocimiento digital* (¿como de hecho lo es!): poder *interrogarla* alrededor de las leyes que la sustentan y la forma en qué en el caso concreto lo hacen; o ayudar a analizar la jurisprudencia que de ella se deriva (entre otras posibilidades) de una manera natural y flexible no estructurada; a través de formas en las que el sistema computacional sea capaz de evidenciar en sus funcionalidades reglas de razonamiento legal que le dan contenido a una fuente de conocimiento jurídico.

Aunque es evidentemente una afirmación subjetiva, existen muchas pistas en la naturaleza de la realidad y la problemática asociada a la información judicial las que sugieren una respuesta afirmativa a estos cuestionamientos y necesidades jurídico-informáticos planteadas anteriormente; y que por ende las condiciones técnicas para este tipo de enfoque deban empezar a ser evaluadas, creadas o fortalecidas al corto o mediano plazo al menos en ámbitos y situaciones concretas. Antes de que el costo de moverse a un enfoque de manejo semántico resulte muy alto o se pierda parte o todo el control sobre el valioso activo que representa el conocimiento legal alrededor del ámbito judicial.

Algo es evidente de toda esta analogía presentada y es que la mayor oportunidad, calidad y cantidad de conocimiento y razonamiento del y sobre el dominio (o realidad contextual) que esté reflejado, integrado y formalizado en el *modelo* (sea este una ley o un sistema computacional o preferiblemente ambos) mayor será su capacidad para cumplir con su propósito funcional y poder desencadenar todos sus beneficios.

Si el ámbito de información judicial es uno en el que sea natural y quizás hasta imperativo disponer de una tecnología computacional



cada vez más sustentada en conocimiento digital, resulta evidente que el proyecto organizacional que conllevaría modelar, crear, mantener y evolucionar el conocimiento digital debe ser adecuadamente inducido en las esferas decisorias y estratégicamente concebido siguiendo expectativas realistas y sustentadas en la debida planificación, la capacidad *ingenieril* y la *expertisia* técnica, administrativa y legal: orientar los esfuerzos de cualquier organización compleja hacia este enfoque envuelve en sí mismo un cambio de paradigma en la visión y el manejo informático y su rol estratégico en la cultura institucional.

Esto en particular confronta el análisis con la pregunta de cómo inducir este tipo innovador de ideas que por su misma naturaleza técnica son, en ocasiones, difíciles de visualizar. Al fin y al cabo los entes judiciales no tienen por misión la innovación en el campo *tecnoinformático*. Con certeza no, pero tal vez sea imperioso velar porque el acceso al conocimiento que encierra el sistema judicial crezca en la proporción y con la debida anticipación que la sociedad (y por ende el mismo sistema operacional de la justicia) pueda requerir en el futuro cercano, o quizás hoy mismo.

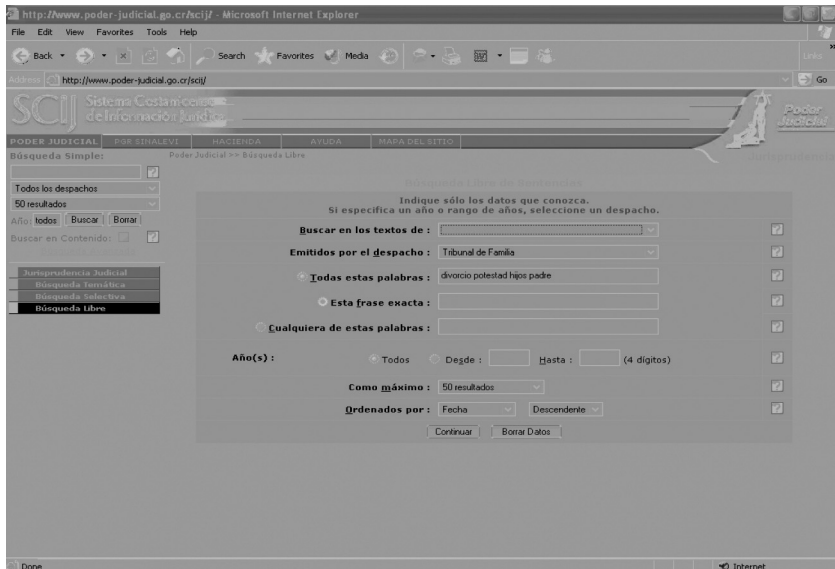
En el caso del Poder Judicial en Costa Rica existe un altamente interesante y pionero antecedente, que es coincidente con los principios básicos y tendencias acá expuestas, el que resulta valioso considerar como una base de partida para los propósitos y reflexiones de este ensayo. Este es el caso del *Sistema Costarricense de Información Jurídica*, o *SCIJ* en sus siglas (más detalles sobre el mismo en <http://www.poder-judicial.go.cr/scij/>, o por ejemplo en <http://www.hess-cr.com/publicaciones/dereinfo/scij.html>)

Presentado sólo muy someramente, este sistema permite en su frente principal al mundo exterior, y mediante una aplicación accesible en el *Web*, acceder por medio de búsquedas de distinta naturaleza a la información de jurisprudencia y normativa en forma interlazada y distribuida entre distintas entidades del sistema jurídico (Procuraduría General de la República, Ministerio de Hacienda entre ellos).

La formulación de consultas por el *Web* es basada en un modelo estructurado (sobre la base de formularios; para un ejemplo ver la Figura 3) dependiendo del tipo y naturaleza de la consulta: incluye libre, selectiva y temática con base a un vocabulario jurídico controlado por el mismo sistema. En otro frente del sistema, hacia adentro del Poder Judicial, en el Centro Electrónico de Información Jurisprudencial (CEIJ), o en su caso el Centro de Información de la Sala Constitucional (CISC),

el sistema permite a analistas sintetizar (extraer, clasificar, construir e indexar) sentencias por su distinto valor y relación jurídica para la formulación de consultas, creación e indización de jurisprudencia derivada de las sentencias y relación con los temas de normativa y doctrina. Un tesoro en Derecho está disponible para la definición de palabras clasificadoras (denominadas *descriptores* y *restringidores*) de modo que permitan la clasificación, tipificación, recuperación de las distintas materias. Es posible acceder hasta el texto de la sentencia y extractos.

**FIGURA 3**



## 6. TENDENCIAS APLICABLES EN EL SCIJ

Volvamos a una pregunta planteada anteriormente acerca de cómo demostrar la validez y utilidad del enfoque de manejo semántico, antes de dar un paso mayor que implique la elaboración de un proyecto de la correspondiente envergadura que esta clase de problema implica (técnica y administrativa) como ya se mencionó antes.

El SCIJ es un primer (ciertamente enorme y pionero) paso hacia facilitar el acceso a la información jurídica. Si se asume también que las necesidades estratégicas de conocimiento son una realidad que debe enfrentar el Sistema Judicial, y que un impacto aún mayor sería posible si

este sistema *evolucionara* hacia uno que permita el razonamiento basado en semántica, entonces es probable que dar un segundo paso sea una vía política y estratégica correcta en el ámbito informático del Poder Judicial; esto para hacer más tangible y ponderable cuáles son los beneficios que una tecnología basada en semántica pueda ofrecer, para demostrar que no se trata de *‘las ropas del emperador’* o de simple ciencia-ficción.

Por sus limitaciones y dadas las bases sobre las que se construye este ensayo no se puede sin embargo, determinar, con absoluta seguridad, cuál sea la forma mejor y la más concreta de materializar una iniciativa alrededor del SCIJ. Para poder hacerlo se tendría previamente que haber analizado más profundamente sus detalles técnicos internos; tampoco se dispone de estadísticas o la retroalimentación de sus usuarios para afinar áreas específicas donde se pudiera introducir tecnología basada en conocimiento. Otros detalles y consideraciones organizacionales no han sido evaluados tampoco. Este tipo de pregunta requiere un estudio que no es parte de este material por limitaciones obvias; tal vez la difusión de este documento sirva para este propósito y sea un disparador que evidencie esta posible necesidad y su eventual atención.

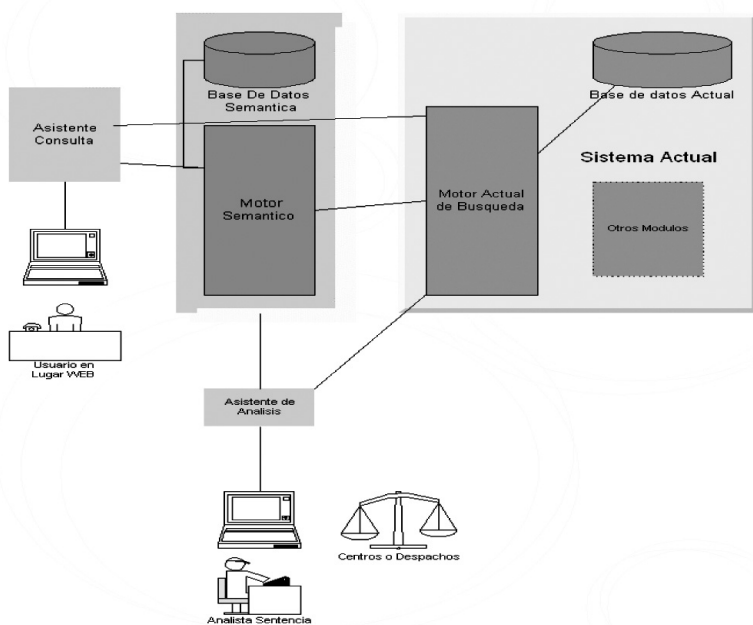
Salvadas estas limitaciones, se podría sin embargo, como un ejercicio estratégico, al menos especular en las siguientes posibilidades demostrativas: Tal como se ha ilustrado al inicio de este documento, una alternativa podría ser adaptarle, mediante un prototipo, una forma de manejo textual más inteligente (no sólo estructurada) por ejemplo durante el proceso de análisis y síntesis en alguno de los *frentes internos*, es decir, en los centros CEIJ o CISC. O quizás más agresivamente aún (y evidentemente con más esfuerzo técnico) durante la elaboración o redacción de una sentencia, como una herramienta *‘asistente’* al redactor en el despacho mismo. Esta propuesta conlleva en particular la posibilidad derivada de plantear consultas (no estructuradas por formularios sino más flexibles) al sistema, por lo que el *frente al Web* también se podría ver ejercitado bajo este enfoque con el mismo *mecanismo de razonamiento* y por ende debería haber beneficios para los usuarios externos, un diagrama hipotético se muestra en la Figura 4. En un *diagrama de bloques* (y omitiendo muchos detalles de bajo nivel) se muestra a grandes rasgos una posible vía de interacción entre la arquitectura actual de SCIJ y los potenciales nuevos componentes basados en conocimiento que se pudieran integrar con aquel.<sup>(8)</sup>

---

(8) El término ‘motor’ usado puede entenderse por simpleza como sinónimo de mecanismo o procedimiento.

Basado en un análisis del contenido textual, el hipotético asistente de redacción podría ayudar al proceso de síntesis y valoración. Esta propuesta requiere la elaboración (o disponibilidad si ya existiese uno) de un *vocabulario semántico* más general y estructurado<sup>(9)</sup> que el se supone acá ya dispone actualmente el tesoro actual del SCIJ, y probablemente sea por esto al inicio adecuado restringir la aplicación sólo a cierta clase de materia legal, para facilitar la evaluación de estas ideas. En la elaboración de este vocabulario puede requerirse asesoría en el campo del Derecho según corresponda. El mecanismo de inferencia debería poder integrarse con el de búsqueda actual, manteniendo al menos en esta funcionalidad el mismo performance que actualmente se ofrece una vez que las deducciones han sido procesadas por el motor semántico. Bajo estas condiciones debería ser posible desarrollar una *'prueba del concepto'* suficientemente clarificadora sobre alternativas para introducir manejo semántico en el SCIJ.

**FIGURA 4**



(9) No en el sentido del Derecho sino de acuerdo a los requerimientos técnicos más generales que puedan requerirse en este contexto.

No se ha elaborado mucho sobre este punto en este documento, pero se supone que herramientas y técnicas del campo de la ‘minería de datos’ (que es distinto pero cercano al de razonamiento semántico) podrían ser usadas también; esto puede ser útil en el caso de querer escalar aún más esta idea sin tener que invertir demasiado trabajo humano en la tarea de ‘*semantización*’ de la información legada, pensando en la futura posibilidad de ir moviendo la información a las estructuras que el modelo semántico demandara. Otras aplicaciones de este campo pueden también ser empleadas, como es usual, para inducir reglas de razonamiento estadísticas entre los datos.

## **7. CONCLUSIONES**

La visión de manejo semántico expuesta ofrece un marco global sobre necesidades, tendencias y estrategias informáticas las que, debidamente entendidas, pueden ser adoptadas en organizaciones donde la introducción de tecnologías basadas en conocimiento y manejo semántico de la información pueda ser de alta relevancia (o eventualmente de urgencia), donde el simple procesamiento de datos convencional o rígidamente estructurado ya no sea suficiente y los volúmenes de información cruda estén creciendo vertiginosamente. El espacio de información y conocimiento de índole legal y por su naturaleza misma se presenta como muy adecuado como modelo para este tipo de propuesta. En el caso del Poder Judicial, el SCIJ parece ofrecer una plataforma muy adecuada y natural para probar y evaluar la utilidad y potenciales beneficios de este tipo de paradigma. Dar ese paso puede crear y abrir nuevos espacios de innovación y mejoramiento para el Poder Judicial y otras entidades del Sistema Judicial. O más aún a otras iniciativas más generales alrededor del esfuerzo denominado Gobierno Digital. La investigación académica y el desarrollo de software nacional también pueden resultar altamente beneficiados con este tipo de iniciativa.

## 8. REFERENCIAS

- BERNERS-LEE T., HENDLER, J. y LASSILA, O. '*The Semantic Web*'. Scientific American, mayo del 2001.
- BREUKER, J. On Legal Information Serving. En '*Legal Knowledge based System*'. Jurix 92: Inf. Tech. And Law'. 1994.
- HESS, C. y QUESADA, C. '*El Sistema Costarricense de Información Jurídica*'. Ponencia al VII Congreso Iberoamericano de Derecho e Informática; Lima, Perú; enero del 2000.
- SAIAS, J. y QUARESMA, P. '*Using NLP techniques to create legal ontologies in a logic programming based Web IR retrieval system*'. ICAIL. 2003.
- VALENTE, A. '*Legal Knowledge Engineering. A Modelling Approach*'. Dissertation. IOS Press, Amsterdam. 1995.
- WEIGAND, H. '*A multilingual ontology-based lexicon for news filtering the TREVI project*'. En IJCAI Workshop on Ontologies and Multilingual. NLP. Agosto, 1997.