



Minería de textos

Análisis automatizado de jurisprudencia

Juan Manuel García Chamizo

Catedrático de Arquitectura y Tecnología de Computadores

Contenido

- ?? Motivación y objetivos 2
- ?? Estado del arte de la consulta jurídica 4
- ?? Capacidad de la tecnología actual 7
- ?? Especificación de un sistema de consulta 8
- ?? Consideraciones de diseño 10
- ?? Prototipo de consulta con ocultación de nombres 11
- ?? Conclusiones 23



Motivación y objetivos

La generación de jurisprudencia en un sistema judicial de una sociedad moderna de unos cincuenta millones de ciudadanos puede requerir entre cinco mil y diez mil folios para contener los textos de las sentencias jurídicas que son emitidas cada día.

El consumo de papel en este caso constituye una preocupación de orden menor. Lo relevante es que una organización social media produce jurisprudencia por cientos de miles de documentos cada año, sobre todo tipo de conceptos jurídicos, desde una amplia diversidad de jurisdicciones, para dar resolución a...

En fin, un volumen descomunal de información que deberá estar bien organizado si se aspira a poder llegar a utilizarlo eficazmente. El diagnóstico es muy simple en términos fenomenológicos: El volumen de datos puede llegar a imposibilitar el acceso a la información.

Como no es razonable reducir el volumen de los datos, la única vía de solución tiene que venir por la adecuación de las técnicas de organización y de consulta de la información jurídica; en otras palabras, hay que ingeniarse formas de escarbar en las pilas de papeles que garanticen el éxito de las búsquedas.

¡Escarbar! Lo mismo que en las minas. Es que, realmente estamos ante yacimientos de información y nuestro objetivo es separar la mena de la ganga. De ahí que en tecnología informática se llama "minería de datos" a esta clase de problemas caracterizados porque la información, aún existiendo, puede ser inaccesible por causa de su volumen. En términos técnicos, las soluciones a estos problemas tienen un requerimiento muy característico que satisfacer: El tiempo requerido para encontrar un dato no puede depender del tamaño del yacimiento. En caso contrario, siempre podrá concebirse un yacimiento donde fracasemos en el empeño. Es fácil intuir que las técnicas habituales de la informática, de inspiración determinista, provocan espantosos ridículos para quienes tratan de resolver con ellas problemas que son genuinamente de minería de datos. Hablar de fracaso informático en estos dominios se traduce en pérdidas económicas mayúsculas y daño al servicio difícil de asumir. Sin embargo, la tecnología informática sí que está en condiciones de resolver satisfactoriamente esta clase de problemas.

La responsabilidad principal para el éxito recae en el personal directivo del organismo o entidad, que es quién decide sobre el equipo técnico al que encarga el proyecto y quién debe establecer las políticas y los métodos. Esta afirmación lleva implícito una reflexión de suma relevancia. El staff directivo general es el que tiene que decidir sobre el sistema de información. Al staff informático sólo le compete la responsabilidad de asesorar y de producir.

La informática es la herramienta. El conocimiento es patrimonio de cada sector. A menudo, esos papeles están invertidos y ocurre que el responsable del negociado supedita sus planes a lo que le permite hacer la informática que tiene al alcance y, con ello, se sumerge en dinámicas en que la producción, la innovación y la modernización van a remolque de la disponibilidad informática; en una suerte de improvisación que sólo permite augurar fracaso o calidad deficiente. Donde eso se



detecte, lo primero que hay que hacer es cambiar la actitud del responsable máximo de la organización.

En esa línea, previo a hablar de lo informático, corresponde reflexionar sobre los requerimientos, sobre los objetivos que deben cubrirse desde el punto de vista de los sistemas de información.

El razonamiento, por tanto, es desde lo más general hasta el detalle de cómo proceder. Por acotar el ámbito, las instituciones europeas están formulando recomendaciones a los estados de la Unión para que acometan cambios administrativos de base tecnológica. La recomendación (2001) 3 del comité de ministros a los estados miembros sobre los servicios de tribunales y otras instituciones jurídicas que deben proporcionarse a los ciudadanos mediante las nuevas tecnologías, traza como objetivos centrales los de facilitar la información legal en formato electrónico y proporcionarla a través de internet con los requerimientos de seguridad y privacidad necesarios. Los objetivos básicos de un sistema de consulta de jurisprudencia reflejados en la recomendación R(95)11 del comité de ministros son los siguientes:

- ?? Facilitar el trabajo de los profesionales proporcionando información rápida, completa y actualizada.
- ?? Facilitar la toma de decisiones, especialmente en áreas de legislación en desarrollo.
- ?? Proporcionar un mayor número de referencias.
- ?? Contribuir a la coherencia de la jurisprudencia sin introducir pérdida de flexibilidad.
- ?? Permitir a los legisladores analizar el uso de leyes.
- ?? Facilitar la investigación sobre jurisprudencia.
- ?? Proporcionar información con propósitos estadísticos.

En el caso de la Administración Española, el Centro de Documentación Judicial (CENDOJ) es un órgano técnico del Consejo General del Poder Judicial cuyas funciones son la selección, ordenación, tratamiento, difusión y publicación de información jurídica legislativa, jurisprudencial y doctrinal. Por otra parte, el reglamento del CENDOJ establece las siguientes actividades más concretas:

- ?? Ofrecer a los órganos judiciales documentación e información legislativa, jurisprudencial y doctrinal, permanentemente actualizada, en los términos que se establezcan a tal efecto.
- ?? Editar publicaciones unitarias o periódicas sobre materias jurídicas.
- ?? Proceder a la recopilación de la jurisprudencia del Tribunal Supremo y de las sentencias de los Tribunales Superiores de Justicia y otras resoluciones judiciales a través de un servicio central de jurisprudencia.
- ?? Realizar la edición de las publicaciones del Consejo General del Poder Judicial que le sean encomendadas.



- ?? Realizar los estudios que le sean atribuidos sobre la aplicación de técnicas y medios informáticos a la Administración de Justicia.
- ?? Realizar estudios de naturaleza estadística, económica y sociológica sobre la actividad judicial, recabando información de los órganos judiciales y promoviendo la suscripción de convenios relativos a las actividades y estudios del CENDOJ con otros servicios y centros de documentación que persigan análoga finalidad, así como con otras instituciones.
- ?? Gestionar, con la colaboración que se precise, la instalación y actualización de las bibliotecas judiciales, sistematizar su régimen y funcionamiento y asegurar su conservación.

Tiene, pues, encomendadas entre sus funciones más relevantes, las de recopilar, custodiar y poner a disposición de los órganos judiciales la jurisprudencia que se derive de la administración de justicia. El CENDOJ se constituye de facto en el servicio central de jurisprudencia.

Estado del arte de la consulta jurídica

La sistematización de la consulta jurídica se produce, como es previsible, con la implantación de la imprenta, tiempo desde el cual, el progreso del sector ha corrido paralelo al de generalización del uso de la documentación escrita. Durante el último cuarto del pasado siglo XX, las recopilaciones jurídicas han ido incorporando estructura, y los textos han sido analizados e indexados, lo cual ha facilitado considerablemente el acceso y la consulta. [REV,98], [ALV,04]

Desde el enfoque de la documentación se acepta que las bases de datos jurídicas son "conjuntos de documentos jurídicos básicos (legislación, jurisprudencia, interpelaciones parlamentarias y doctrina), almacenados en soportes magnéticos o de cualquier otro material y susceptibles de ser tratados, recuperados y transmitidos de forma total o parcial mediante procedimientos y medios informáticos, que con la aplicación de técnicas derivadas de la utilización de la informática jurídica pretenden ser utilizados con una finalidad divulgadora pública y generalizada de su contenido". Con ese criterio, se encuentra que las particularidades de la documentación jurídica son principalmente las siguientes: [LOP,84], [PAE,94], [PAE,95]

- ?? Se necesita recoger el texto íntegro para poder ser de utilidad a las necesidades reales de los usuarios y garantizar la seguridad jurídica.
- ?? Coexisten varios formatos documentales: texto completo, extractos, fragmentos especialmente significativos, etc.
- ?? Elevado volumen de documentos, lo que ha provocado la proliferación de productos temáticos específicos.
- ?? Conflicto de intereses entre la accesibilidad a jurisprudencia vigente y consolidada y la disponibilidad exhaustiva ignorando la antigüedad.
- ?? Requerimiento irrenunciable de accesibilidad en tiempo real de la jurisprudencia más reciente por ser la de mayor vigencia.



- ?? Existe delimitación de carácter territorial y jurisdiccional.
- ?? Necesidad de correlacionar las resoluciones en coherencia con las relaciones que establecen o rompen una línea jurisprudencial, como interrelaciones entre legislación, jurisprudencia y doctrina.

En cuanto a los hitos y la cronología de desarrollo y explotación de las grandes bases de datos jurídicas, por completitud ya que no es el objetivo de este documento, pueden consultarse numerosas referencias. Baste decir que empezaron a utilizarse en Estados Unidos durante los años 70 y que, progresivamente, fue extendiéndose su uso por Europa (Reino Unido, Francia, Bélgica, Alemania, etc.) y que la evolución seguida ha conducido a que los requerimientos de infraestructura tecnológica y la capacitación requerida en el personal para acceder a las bases de datos en línea, limitaba grandemente su utilización. [BIN,91], [ALV,04], [MAC, 96], [CAR,93], [GAR,99]. La evolución más reciente viene en la línea de la aparición de gran número de publicaciones basadas en las nuevas tecnologías de la información y las comunicaciones; lo que está mejorando notablemente la difusión, sobre todo debido a la gratuidad de los servicios que vienen proporcionándose. Aún así, los sectores profesionales siguen documentándose mediante servicios esencialmente de pago o proporcionados colegiadamente en formato restringido.

Aunque el acceso a la jurisprudencia es libre, no lo es la explotación comercial por un tercero de los trabajos de recopilación, selección, análisis documental y jurídico, tratamiento informático, etc.; los cuales, a su vez, aportan valor añadido, mediante el análisis y la elaboración que incorporan. Las editoriales, pues, realizan una valiosa aportación económica, técnica e intelectual sobre la materia prima que constituyen los documentos jurídicos originales. Es ahí donde se justifica la labor que realizan.

Por otro lado, los organismos propietarios de la documentación judicial tienen planteado el reto de proporcionarla en soportes y formatos orientados a la prestación de servicio público que tienen encomendada.

La relativamente reciente creación del propio órgano de documentación judicial es una de las razones para que, hasta la fecha, la gestión documental de la jurisprudencia se encuentre todavía en los preliminares de su automatización. La realidad es que la jurisprudencia se sigue proporcionando en virtud de antiguos acuerdos y costumbres a empresas del sector editorial especializado, las cuales, a partir de los documentos de las sentencias, elaboran productos editoriales de utilización mucho más amable que la documentación original. Siendo irreprochable el beneficio que para la administración de justicia cumple la producción editorial, algunas carencias limitan la vigencia de estos usos:

- ?? La cantidad de sentencias que se proporcionan es extremadamente limitada respecto del volumen total de jurisprudencia que se produce. En consecuencia, sigue sin estar satisfecho el objetivo de proporcionar la mayor cantidad posible de referencias, necesario para garantizar la corrección de la práctica judicial. Otro tanto ocurre con la elaboración de estadísticas que debieran servir de ayuda para la toma de decisiones tanto al judicial, como al ejecutivo y al legislativo.
- ?? La selección de la jurisprudencia que se publica es realizada externamente al poder público de administración de la justicia. Con ello se renuncia



implícitamente a los supuestos de independencia y de responsabilización. Téngase en cuenta que la transcripción implica a un proceso de interpretación de las sentencias a cargo del personal de la empresa editora.

- ?? La elaboración se realiza mediante procesos basados esencialmente en intervención humana. El efecto es que los tiempos de operación son los que corresponden a la actividad humana y la consecuencia es que la jurisprudencia llega así procesada llega a manos de los profesionales tras una latencia temporal que resta vigor, y puede que hasta vigencia, a la propia jurisprudencia.

Los sistemas clásicos de bases de datos textuales proporcionan acceso a legislación, jurisprudencia general y jurisprudencia especializada:

- ?? Legislación. Las disposiciones de carácter general publicadas por los servicios públicos oficiales; ya sea acceso a los textos completos o a información referencial.
- ?? Jurisprudencia general. Emanada de los altos tribunales (Constitucional, Supremo, Superiores de Justicia y otros).
- ?? Jurisprudencia especializada: Tributaria, social, civil.

Estas bases de datos estructuran su contenido en campos (cod-ident, resolución, jurisdicción, disp-estud, resumen, ponente, ..., texto libre), con alguno de ellos de texto libre que suele contener el texto íntegro de la sentencia.

- ?? cod-ident: Número que identifica la resolución.
- ?? jurisdicción: Tribunal que dicta la resolución, la sede, la sección y la jurisdicción.
- ?? resolución: Tipo de resolución, la fecha, el número de sentencia, rollo, expediente, etc.
- ?? disp-estud: Referencias a artículos estudiados en la resolución.
- ?? resumen: Contenido principal de la resolución.
- ?? ponente: Nombre y apellido del ponente que prepara la resolución.
- ?? voces: Palabras jurídicas clave, resultado del análisis de la resolución. Cada voz desarrolla un concepto jurídico compuesto por conceptos más sencillos, en organización jerárquica.
- ?? texto: Texto completo de la resolución, antecedentes de hecho, fundamentos de derecho, y el fallo.

Es frecuente proporcionar la posibilidad de búsqueda mediante palabras clave en alguno de los campos o en el texto libre y búsquedas jerárquicas. Para ello se suele disponer de un tesoro.

La otra alternativa, la de tener acceso a la documentación bruta sobre jurisprudencia, tiene los conocidos inconvenientes de que la cantidad convierte en inaccesible, de manera práctica, a la información ya que:

- ?? El tiempo que se requiere para efectuar una consulta depende directamente del volumen de jurisprudencia disponible.



- ?? No siempre se dispone de personal experto para ayudar en las tareas de consulta de jurisprudencia.
- ?? La calidad de la consulta realizada directamente sobre la documentación jurídica decae vertiginosamente a medida que crece el volumen de la documentación debido a factores humanos de cansancio, desánimo, motivación y rutina, principalmente.

La superposición de ambos métodos de consulta, la búsqueda de documentación preseleccionada, completada con la consulta sobre la documentación bruta, lógicamente mejora los resultados y establece un equilibrio que es el que, en la práctica, se admite como adecuado. Sin embargo, esa aceptación adolece de los avales de completitud y objetividad deseables. Se admite como válida porque es lo mejor que se puede hacer con los medios disponibles.

El papel que se ve obligada a desempeñar la pericia de los profesionales para garantizar la calidad de utilidad de la jurisprudencia en un contexto así es desproporcionadamente relevante.

La consecuencia global es que la consulta de jurisprudencia, tal cual se realiza en la generalizadamente en la actualidad, no se corresponde con la capacidad técnica que la tecnología informática tiene. Se hace necesario que los servicios de documentación jurídica profundicen en las posibilidades de la herramienta informática como vía para contribuir a proporcionar importantes mejoras de coherencia, completitud y transparencia de la consulta jurídica.

Capacidad de la tecnología actual

Las consideraciones precedentes sugieren un cambio en la tecnología y en los modos de trabajo. Las líneas maestras son las siguientes:

- ?? Profundizar en la automatización de los procesos de gestión de la jurisprudencia mediante el empleo de técnicas modernas de organización de información masiva. Los beneficios que se conseguirán serán fundamentalmente los de racionalización de las operaciones, ganancia en seguridad, reducción de los costes y optimización de los tiempos.
- ?? Universalizar la disponibilidad de la jurisprudencia para que pueda ser consultada de forma fácil y asequible tanto a los estamentos judiciales como a las demás instituciones, sectores profesionales y a los ciudadanos en general mediante la consulta a través de internet.
- ?? Crear el archivo nacional de sentencias, de modo que permita dar servicio a profesionales, instituciones y ciudadanos, preparándolo para los usos que se la asignen en materia de certificación, conocimiento público, etc.

En cuanto a las posibilidades de la tecnología actual, muchos de los objetivos que se han planteado pueden cubrirse; mientras que para alcanzar otros de ellos, el conocimiento de que se dispone puede proporcionar mejoras y ayudas. Algunas de las necesidades actuales y otras que se plantearán a medida que limitaciones previas



vayan siendo superadas, requerirán todavía de un esfuerzo de investigación y desarrollo.

Las bases de datos textuales y sus mecanismos de indexación proporcionan la plataforma metodológica más comúnmente utilizada en estos contextos. Son numerosos los desarrollos que facilitan mecanismos de consulta de las bases de datos con cierta flexibilidad; sin embargo, la selección de jurisprudencia mediante esquemas de consulta próximos al lenguaje natural constituye un problema complejo y en su solución se requieren técnicas avanzadas. Estas técnicas utilizarán relaciones semánticas y características contextuales extraídas de los documentos. Otro tanto puede decirse en relación con la obtención de criterios de mantener la vigencia y obsolescencia de jurisprudencia debida a la aparición de documentación posterior.

La evolución de los requerimientos se está produciendo en dos líneas fundamentales: hacia el acceso a los servicios a través de Internet y hacia el desarrollo de esquemas de consulta inteligentes y flexibles próximos al lenguaje natural que sean capaces de acotar de forma certera el resultado de las consultas. Un buen ejemplo de este tipo son los sistemas de búsqueda de páginas web donde el problema de la dimensión de la información y la falta de estructura son especialmente agudos. Google, por ejemplo, utiliza tecnología PageRank de meta-búsqueda para medir objetivamente la importancia de las páginas web y, por ese criterio, las muestra. Se calcula que resuelve una ecuación de 500 millones de variables y más de 2.000 millones de términos. La técnica de meta-búsqueda interpreta un vínculo de la página A hacia la página B como un 'voto' de la Página A por la Página B y establece la importancia de una página a partir de los votos que recibe. También analiza la página que emitió el voto. El éxito del buscador se apoya en 54.000 servidores, 100.000 procesadores y 261.000 discos duros repartidos por Estados Unidos y Europa, lo que lo convierte en uno de los sistemas informáticos más grandes del mundo.

La búsqueda de jurisprudencia tiene planteada la necesidad de desarrollar sistemas de computación de altas prestaciones y esquemas de consulta sofisticados basados en técnicas de inteligencia artificial y minería de datos. Estos esquemas deben ser capaces de acotar los resultados de las búsquedas: utilización de tesauros con relaciones semánticas, históricos de consultas similares, esquemas de votación de jurisprudencia para consultas previas, estudios de frecuencia de términos de los documentos para su agrupación por categorías... Deberá poder prestar servicios en línea a un gran número de usuarios y cumpliendo las especificaciones de calidad de servicio orientadas a garantizar la imposibilidad de identificación sistemática de las personas inculpadas, optimizar los resultados de las consultas en cuanto a cantidad de referencias proporcionadas y selección inteligente de dichas referencias, así como garantizar tiempos de respuesta y calidad de la documentación accedida en línea.

Especificación de un sistema de consulta

La información recogida en las sentencias judiciales se puede considerar como una mezcla de información estructurada y no estructurada. El primer tipo de información se caracteriza por la existencia de unos atributos claramente definidos (palabras clave) que pueden ser utilizados para la extracción de conocimiento (juizado,



tribunal, número de sentencia, etc.). Por el contrario, se considera información no estructurada a aquella consistente en lenguaje jurídico (cuerpo de la sentencia) donde se mezclan datos relevantes con datos irrelevantes. Es importante tener en cuenta las características propias de los documentos para poder establecer mecanismos de clasificación que proporcionen elevada fiabilidad, adaptación al usuario y estabilidad ante pequeñas variaciones en la descripción de los términos (mayúsculas, tildes, etc.).

La información estructurada facilita la clasificación automática ya que se dispone de una agrupación natural correspondiente a los diferentes atributos y los valores que éstos pueden tomar. Si el conjunto de valores que pueden tomar los atributos es elevado o desconocido en parte, se deben utilizar estrategias flexibles de búsqueda y clasificación incorporando, por ejemplo, la posibilidad de no diferenciar entre mayúsculas y minúsculas o no tener en cuenta las tildes. Dentro de estas estrategias también cabe la detección de pequeños errores tipográficos como la repetición de la misma letra ("lettra" por "letra"), la existencia de alguna letra incorrecta añadida ("polítizca" por "política"), la escritura inversa de pares de letras ("cnación" por "canción").

La organización de los documentos basándose en términos que aparezcan en el cuerpo de la sentencia resulta más compleja por el carácter no estructurado de esta sección. A pesar de ello, una propuesta inmediata es la clasificación basada en criterios estadísticos (tomando como partida un diccionario, se pueden organizar grupos, de diferentes grados de exactitud, en función de las ocurrencias de las palabras del diccionario). También se puede utilizar algún sistema de filtrado, destinado a la clasificación, como la indexación de términos por frecuencias, técnica que establece un sistema de pesos en función de la frecuencia relativa de cada término en cada documento. Para mejorar la organización de grupos significativos de documentos relacionados con una idea o concepto se pueden utilizar técnicas de localización de radicales lexicográficos. Estas técnicas consisten en utilizar un lema o raíz léxica que identifique el tipo de documento, de manera que se incorporen al grupo los documentos que presenten algún término formado por la raíz y los sufijos y prefijos permitidos. Por ejemplo: asesin + (o) (a) (ar) (ado) (ada) (os) (as). Este método se puede combinar con algoritmos estadísticos utilizando varias raíces relacionadas que identifiquen un mismo grupo y que su ocurrencia indique determinado grado de fiabilidad en la clasificación. Otra estrategia que puede aumentar la calidad de las clasificaciones realizadas es la utilización de diccionarios de sinónimos que incorporen un mayor grado de automatismo al proceso de clasificación. Estos diccionarios permiten que el análisis estadístico relacionado con varios términos referentes al mismo concepto se realice, bien de forma automática, bien bajo supervisión pero ofreciendo comodidad al usuario al presentar una lista de sinónimos ya confeccionada. En cualquier caso, estos diccionarios pueden ser modificados o ampliados por el usuario de forma sencilla. También se pueden utilizar combinaciones de palabras teniendo en cuenta su posición respectiva. De esta forma se pueden organizar grupos de documentos que contengan determinados términos distanciados una cantidad máxima de palabras. Siguiendo con esta misma estrategia de localización de términos en el documento, podría resultar interesante clasificar en función de la aparición de un término significativo en determinada sección de la

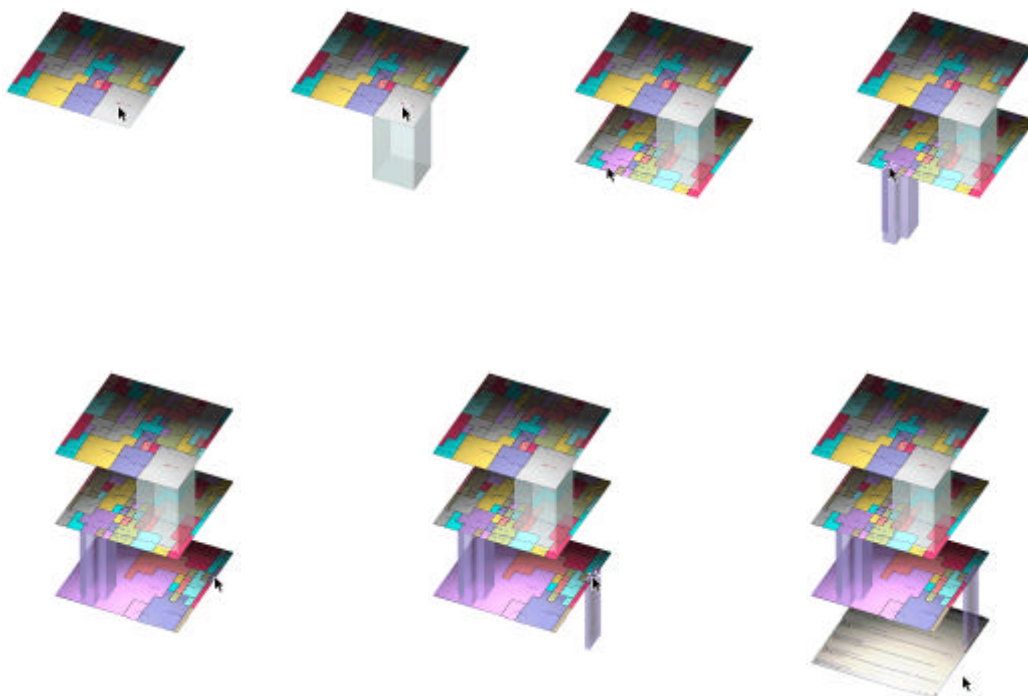


sentencia; sección que podría ser identificada mediante palabras clave o según el número de líneas que le antecederan o le siguieran.

Consideraciones de diseño

Realizado el análisis preliminar, se plantea utilizar un método híbrido de clasificación basado en un tesoro ponderado por utilidad y optimizado en varios aspectos léxicos y sintácticos. La correlación de documentos por el grado de similitud semántica se basará en técnicas clasificatorias conexionistas no supervisadas.

El entorno de usuario incorporará una interfaz gráfica análoga a un mapa geográfico. Las regiones del mapa corresponderán a los conceptos semánticos y la vecindad entre regiones se organizará por criterios de correlación semántica. Las vistas del mapa serán escalables desde una vista general conteniendo todo el universo semántico hasta el detalle de cada sentencia y sus vecinas. Como ejemplo para comprenderlo mejor puede pensarse en una pantalla conteniendo la península Ibérica (una distribución coloreada de los conceptos jurídicos) y etiquetas en los puntos Lisboa, Sevilla, Madrid, Donosti, Barcelona y Valencia (alternativamente, conceptos como, penal, administrativo, auto, etc.) Mediante un zoom podría pasarse a otra vista más detallada donde la región etiquetada 'penal' estuviera, a su vez, organizada en zonas (homicidio, asesinato, etc.) y así sucesivamente, mayores niveles de detalle hasta cada sentencia y las más parecidas, a las cuales se podrá acceder. Así mismo, se podrá efectuar consulta mediante interfaz de texto, prescindiendo del soporte gráfico.





Para cada sentencia podrá consultarse su ficha resumida o el texto completo. En ambos casos, el formato será de hipertexto para proporcionar al usuario la facilidad de navegación interactiva a otras sentencias del mismo magistrado o con atenuantes parecidos, por ejemplo.

La estructura y el tamaño del clasificador será independiente del volumen de sentencias (en realidad, dependerá de la cantidad de conceptos) por lo cual, la consulta será rápida. A pesar de ello, las características del paralelismo a bajo nivel de los algoritmos previstos, permiten establecer la velocidad de operación según necesidades. El sistema tendrá, así mismo, prestaciones de monitorización continua, clasificación en línea de nuevas sentencias y creación de conceptos semánticos nuevos.

Como mejora inmediata a incorporar, se propone la de impedir que puedan efectuarse consultas basadas en los nombres de los inculpados pero sin que por ello los nombres correctos de los inculpados dejen de poder leerse en los textos de las sentencias. Ello garantizará la ausencia de errores debidos a efectos semánticos secundarios de consanguinidad y similares. Piénsese, por ejemplo, en la confusión que puede provocar el hecho de utilizar nombres ficticios para los inculpados que aparecen en el texto de una sentencia sobre un contencioso por haber facturado las cargas tributarias correspondientes a una propiedad, a un ciudadano cuyos apellidos y la inicial del nombre coincidan con los del propietario legal. Consideramos además que hay un factor estético relevante: en la medida que la tecnología lo puede resolver fácilmente, es difícil justificar que se pierda el carácter de público que tiene el texto literal de la sentencia cuando lo que se persigue es cumplir el precepto de impedir la identificación sistemática automatizada de los inculpados.

Prototipo de consulta con ocultación de nombres

El servicio de documentación de jurisprudencia se fundamenta en las siguientes características:

- ?? **Accesibilidad.** Una característica muy importante si se tiene en cuenta que los usuarios del sistema no tienen necesariamente por qué tener grandes conocimientos en informática pudiendo incluso mostrarse, en primera instancia, reacios a su utilización. Se debe considerar la facilidad de uso como una característica primordial. La accesibilidad del sistema no debe ser justificación para reducir o limitar la capacidad del mismo en cuanto al tratamiento de la documentación: búsquedas, ordenación, etc.
- ?? **Actualidad.** La jurisprudencia la información existente con la mayor rapidez posible. Esto exige podrá estar accesible prácticamente a medida que se disponga de ella en el servicio de documentación.
- ?? **Disponibilidad.** La utilidad del servicio depende también de su grado de disponibilidad de forma espacial. Para facilitar su utilización se pretende que no existan restricciones de disponibilidad, es decir, que pueda accederse a la información desde cualquier lugar y a cualquier hora del día ininterrumpidamente durante todo el año.



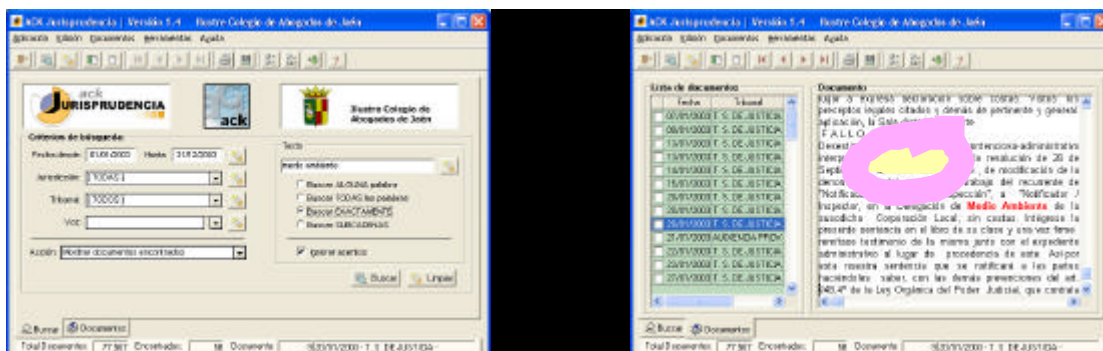
- ?? Exactitud. La documentación ofrecida por el servicio de documentación de jurisprudencia debe ser fiel a las sentencias emitidas ya que cualquier cambio en su formato o redacción puede alterar el espíritu de la misma. Sin embargo, podrán proporcionarse opcionalmente versiones corregidas de las sentencias en las cuales se habrán subsanado las erratas tipográficas triviales.
- ?? Autenticidad. El sistema debe facilitar la incorporación de mecanismos que aseguren la autenticidad de la información mostrada. Además, en todo momento, debe preservarse el origen de la documentación sea cual sea el mecanismo de obtención de la información (visualización en pantalla, impresión, etc.)
- ?? Protección. Se deben articular diferentes niveles de acceso que condicionen el grado del servicio. De esta forma la documentación original puede presentarse con campos ocultos, de forma abreviada e incluso ofrecer limitaciones de acceso o de extracción de la información. También se deben proveer los mecanismos oportunos para preservar la información de alteraciones maliciosas tanto en la base de datos original como durante el recorrido hasta la presentación al usuario.
- ?? Flexibilidad. El sistema debe considerar posibilidades de futuro y resultar lo más flexible posible ante modificaciones en el servicio y en la documentación tratada como por ejemplo la incorporación de documentación judicial en varios idiomas o cambios en el formato de las sentencias.

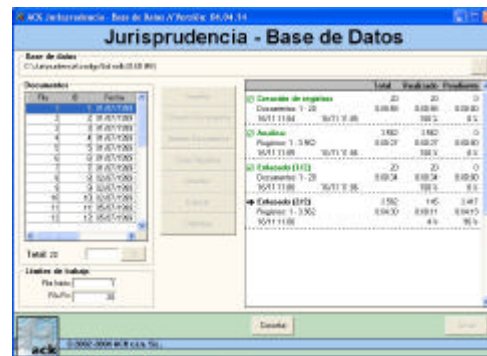
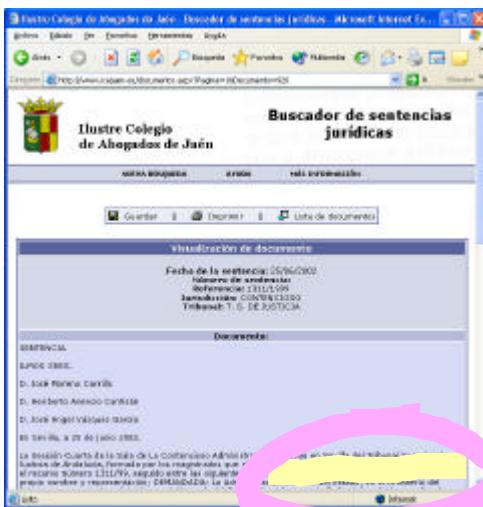
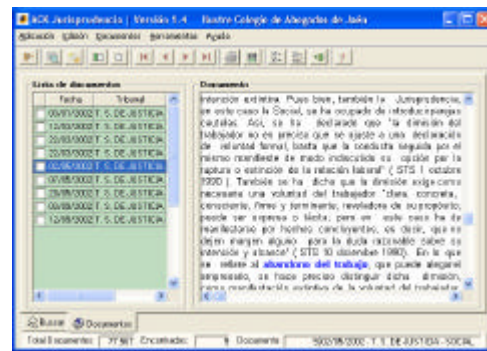
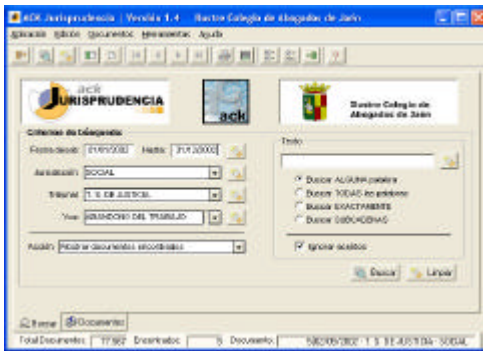
Tomando como base esas características se pueden proponer los siguientes requerimientos concretos para un sistema de documentación de jurisprudencia:

- ?? Almacenamiento de la información. La documentación relativa a sentencias se almacenará de forma íntegra, sin ninguna modificación para conservar la exactitud. La encriptación necesaria para preservar la información sensible (p.e. nombres de intervinientes) se hará, en la medida de lo posible, en tiempo real ante peticiones concretas de documentos, conservándose en la base de datos original la información sin codificar.
- ?? Búsquedas de documentación. Las búsquedas se facilitarán mediante la incorporación de campos generales que puedan servir de guía al usuario: fecha de sentencia, provincia, partido judicial, jurisdicción, tribunal.
- ?? También se incluirá un sistema de búsqueda más completo que permita al usuario la realización de búsquedas de una forma flexible y personalizada: búsqueda de documentos según términos basados en lemas, sinónimos, subcadenas. De igual forma se propondrá un conjunto de palabras relacionadas con el término propuesto por el usuario para guiarle en la búsqueda. Además se ofrecerá la opción de limitar la búsqueda a partes del documento como los antecedentes, exposición o fallo.
- ?? Una vez finalizada la búsqueda por el sistema se listarán los documentos encontrados y se facilitará la opción de seleccionar varios para proceder a su lectura, almacenamiento o impresión. Se incorporará una opción que permita limitar el número de documentos listados.
- ?? Se ofrecerá un servicio de recomendación basado en el historial de búsquedas de manera que ante una petición de búsqueda y tras ser elegidos varios documentos

- entre los encontrados, el sistema buscará otros documentos que puedan presentar similitudes con los escogidos y no hayan sido obtenidos en la búsqueda previa.
- ?? Ordenación. El usuario podrá ordenar los resultados obtenidos de la forma que considere más oportuna según sus necesidades y en función las opciones proporcionadas para tal efecto (ordenación por fechas, jurisdicción, provincia, tribunal y combinación de estos campos).
 - ?? El sistema ofrecerá la posibilidad de imprimir tanto documentos completos como partes de los mismos así como la copia en fichero de los textos seleccionados.
 - ?? Se incorporará un resumen de estadísticas que incluyan porcentajes de accesos a sentencias por jurisdicción, provincia, partido judicial, palabras clave.
 - ?? Para proporcionar el adecuado nivel de seguridad en el acceso de los documentos se solicitará al iniciar la sesión un nombre de usuario y la contraseña correspondiente y se almacenará información anónima de las acciones que realice para incorporarla a las estadísticas. Al menos debe haber dos niveles de acceso: administrador y usuario.
 - ?? Para mantener la integridad del sistema se debe incorporar un módulo de administración que permita la gestión de los documentos: añadir, eliminar, modificar. Otra opción consistirá en mantener los documentos en la base de datos pero impidiendo su visualización.
 - ?? El gestor de base de datos que se utilice será un elemento primordial ya que debe proporcionar soporte para acceso multiusuario, copias de seguridad, acceso a través de red.

Las imágenes que siguen corresponden a algunas ventanas representativas del prototipo. Se han sobre-rayado algunas ocurrencias de ocultación de nombres de intervinientes. A continuación, se adjuntan notas sobre el generador de la base de datos y sobre el entorno de consulta.





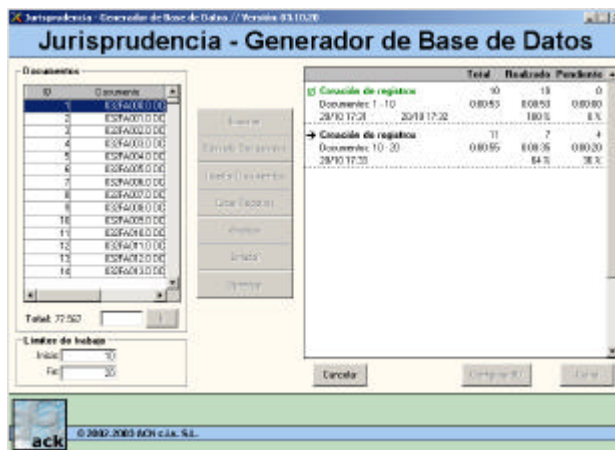
GENERADOR DE BASE DE DATOS

Información general

La creación de la base de datos de jurisprudencia se realiza mediante la ejecución secuencial de un conjunto de procesos. Se recomienda que entre cada uno de los procesos se realice una compactación de la BD y una copia de seguridad de la misma para poder recuperarla en caso de error.

Algunos de los procesos tienen un alto coste temporal, por ello se recomienda fraccionarlo en varias tandas de un número limitado de sentencias cada una. Para ello utilizaremos la zona marcada como 'Límites de trabajo' de la aplicación que podrá ser indicada para los siguientes procesos:

- ?? Convertir documentos
- ?? Insertar documentos
- ?? Crear registros
- ?? Enlazar



Preparación de la base de datos

Primeramente tendremos que preparar una base de datos vacía para comenzar a trabajar con el generador. Tanto el generador como el programa de búsqueda que se distribuye con el CD trabajan con el mismo formato de BD, de tal manera que el resultado final de los procesos del generador es la BD que se distribuye en el CD.

La base de datos es Access 97 aunque su extensión ha sido renombrada a .JUR para evitar su apertura por los usuarios. Contiene la clave de acceso '2k2_ilaci' como una primera medida de seguridad.

El programa por defecto trabaja con una base de datos situada en la carpeta de instalación de la aplicación y con el nombre 'BASE.JUR'. Es aquí donde deberemos

dejar una base de datos vacía al empezar. Si se desea trabajar con una base de datos situada en otro lugar o con otro nombre usaremos el botón 'Configurar BD' para seleccionarla.

La base de datos viene con una serie de datos ya introducidos que pueden ser modificados antes de la generación para personalizarla:

?? Voces: Listado de voces por las que se podrá buscar en el programa (Tabla voces)

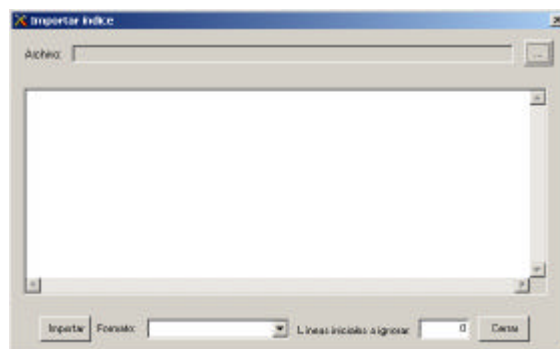
?? Jurisdicciones (Tabla: mantenimiento_jurisdiccion)

?? Tribunales (Tabla: tribunal)

Importar documentos

Como primer paso a la hora de generar la base de datos tendremos que introducir las sentencias de trabajo en ésta. Para ello nos valdremos de los índices que envía el CENDOJ junto con los documentos de las sentencias.

Para importar estos índices pulsaremos el botón 'Importar' con lo que se abrirá la siguiente ventana:



Mediante el botón '...' seleccionaremos el índice que deseamos importar, el cual será mostrado en la zona central de la ventana.

Seguidamente elegiremos el formato del índice y, en caso de que tenga, el número de líneas de cabecera que hay que ignorar.

Pulsaremos el botón importar y comprobaremos como el número total de documentos se incrementa.

Es posible que durante el proceso de importación el programa nos pregunte el valor que tenemos que reasignar cuando tiene dudas sobre algún campo.

Realizaremos este proceso con todos los índices de los documentos que deseemos incorporar a la base de datos.

El resultado debe ser un conjunto de N documentos numerados de 1 a N.

Convertir Documentos

Los documentos que nos envía el CENDOJ están normalmente en un formato de tipo DOC o RTF. El siguiente paso que tenemos que realizar es transformarlos a un formato de texto plano (TXT) para facilitar su tratamiento.

Situaremos todos los documentos que deseamos transformar (los .DOC o .RTF) en una carpeta y crearemos otra carpeta donde se situarán los documentos de texto.

Pulsamos el botón 'Convertir Documentos' y le indicamos la localización de las carpetas del fuente y del destino:



La aplicación creará un fichero de texto por cada una de las sentencias de la base de datos.

Insertar Documentos

En el siguiente paso se incorporan los documentos en formato TXT creados a la base de datos.

Para ello hay que pulsar el botón 'Insertar Documentos' e indicar la localización de los documentos TXT:



En este proceso también se comprime y encripta las sentencias conforme se van incluyendo en la base de datos.

Crear Registros

Las futuras búsquedas en la base de datos están optimizadas mediante el uso de registros indexados. En este proceso se crean estos registros.



Para realizarlo sólo hay que pulsar el botón 'Crear Registros'.

Recomendamos realizar este proceso en bloques de como mucho 15.000 sentencias, ya que de otra manera el proceso se ralentizaría mucho.

Analizar

En este proceso se analiza la información procesada en el paso anterior.

Solamente hay que pulsar el botón 'Analizar'.

Enlazar

En este proceso se determinan y ocultan los nombres propios de los inculpados, despersonalizándose las sentencias.

Hay que pulsar el botón 'Enlazar' para lanzar las dos fases que componen este proceso.

Recomendamos realizar este proceso en bloque de como mucho 15.000 sentencias, ya que de otra manera el proceso se ralentizaría mucho.

Optimizar

Por último se realiza un proceso de optimización que agiliza las posteriores búsquedas.

Solamente hay que pulsar el botón 'Optimizar'.

Una vez finalizado este proceso ya tendremos generada la base de datos que se podrá utilizar con el programa de búsqueda.

No olvidar comprimir la base de datos también al final para que el tamaño de la misma sea el mínimo.

SISTEMA DE CONSULTA

Jurisprudencia Local

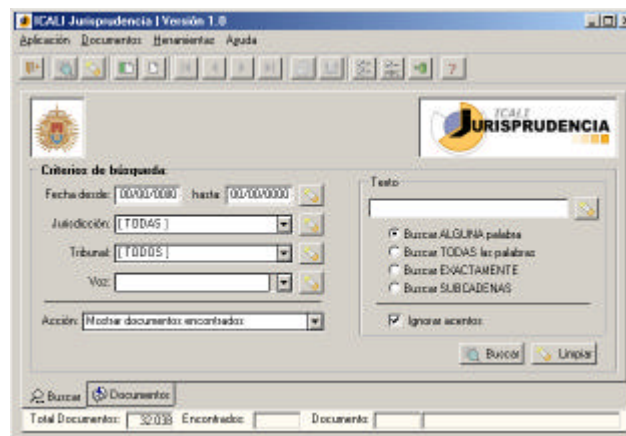
Descripción

El CD de Jurisprudencia Local es una base de datos con sentencias de recursos provenientes de diversos órganos jurisdiccionales.

El programa se distribuye en formato CD-ROM y consta de una base de datos con las sentencias incluidas y la instalación del software necesario para la búsqueda y tratamiento de las sentencias.

Software de búsqueda

El programa de búsqueda de sentencias está estructurado de la siguiente manera:



?? Barra de título: Donde se informa del nombre y versión de la aplicación.

?? Menú: Zona de acceso a todas los módulos de la aplicación.

Aplicación Documentos Herramientas Ayuda

?? Barra de botones: Lista de botones de acceso a las principales áreas de la aplicación.



?? Zona de búsqueda / Visualización de documentos: Mediante un sistema de pestañas incluido en la parte inferior de esta zona se puede seleccionar uno de estos dos componentes:

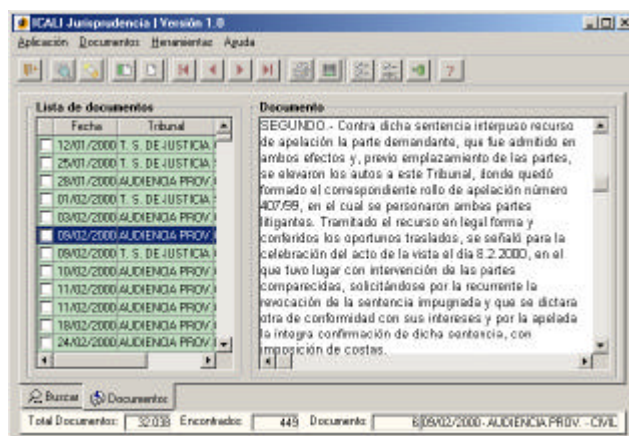
- ?? **Buscar:** Desde aquí se introducen los criterios de búsqueda necesarios para localizar un conjunto de sentencias.

Fecha desde: 00/00/0000 hasta: 00/00/0000
Jurisdicción: [TODAS]
Tribunal: [TODOS]
Voz: []
Texto:

 Buscar ALGUNA palabra
 Buscar TODAS las palabras
 Buscar EXACTAMENTE
 Buscar SUBCADENAS
 Ignorar acentos

Los diferentes criterios son:

- ?? Fecha. Es el rango de fechas a las que pertenece la sentencia. Se pueden indicar rangos de fecha abiertos omitiendo el *desde* o el *hasta*.
- ?? Jurisdicción. Se puede indicar la jurisdicción de las sentencias a buscar o bien buscar las sentencias con cualquier jurisdicción.
- ?? Tribunal. Se indicaría el tribunal de las sentencias a buscar.
- ?? Voz. Seleccionaríamos de una lista la voz de las sentencias a buscar. También es posible introducir una voz que no esté reflejada en la lista.
- ?? Texto. Indicaríamos una o más palabras que deseamos que aparezcan en las sentencias encontradas. Se debe especificar si se desea buscar alguna de las palabras introducidas, todas, fragmentos de las introducidas o la frase tal cual se escribió, es decir todas las palabras y en el mismo orden. También se pueden ignorar los acentos.
- ?? **Documentos:** En la pestaña de documentos se muestra una lista de las sentencias encontradas y el contenido de aquella que está seleccionada.



Es posible maximizar la zona de visualización del documento para su mejor lectura.

?? **Barra informativa:** En ella se muestra información sobre el total de las sentencias de la base de datos, el total encontrado y el documento actualmente seleccionado.

Total Documentos: 32.038 Encontrados: 449 Documento: 6/09/02/2000 - AUDIENCIA PROV. - CIVIL

Trabajo con los documentos

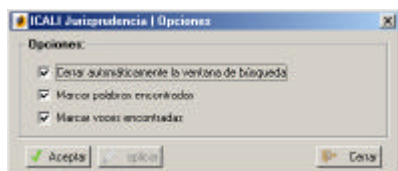
Existen una serie de utilidades para el manejo de los documentos:

- ?? **Navegación:** Mediante la lista de documentos o los botones de *Siguiente*, *Anterior*, *Primero* y *Último* se puede seleccionar fácilmente la sentencia a visualizar.
- ?? **Impresión:** Se pueden imprimir las sentencias encontradas. La impresión se puede realizar del documento seleccionado, del total de sentencias encontradas o de un conjunto de ellas.
- ?? **Exportación:** De igual forma el documento puede ser exportado a un fichero para su posterior edición.
- ?? **Marcado:** Tanto en la impresión como en la visualización en pantalla de las sentencias se marcan con colores configurables las palabras o voces encontradas.

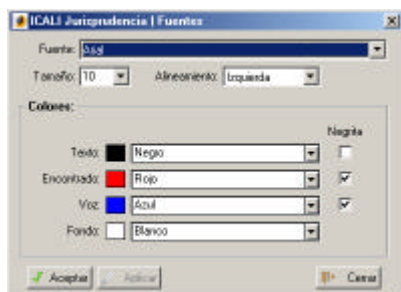
Utilidades

El programa aporta también una serie de utilidades como:

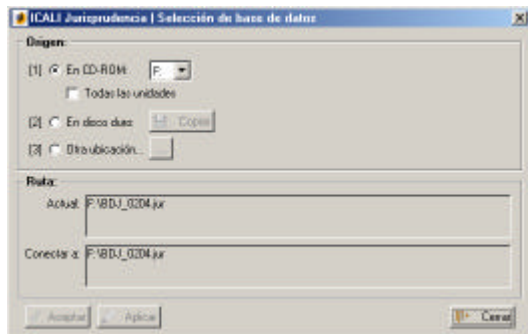
?? **Opciones:** Mediante las que podemos configurar la aplicación



?? **Configuración de fuentes:** Para indicar el aspecto, color y tamaño de las fuentes con las que se muestran las sentencias.



?? **Configuración de la base de datos:** Que nos permite usar la base de datos desde el CD original, instalarlo en una unidad de disco o instalarlo en un servidor de nuestra red local.



?? **Ayuda:** La aplicación cuenta con un sistema de ayuda que nos da información sobre las diferentes partes del programa y su uso.



Conclusiones

En términos generales, se concluye que la tecnología informática se encuentra actualmente en condiciones de aportar un importante grado de automatización de las actividades de consulta jurídica, sin que ello signifique que está en condiciones de resolver todos los requerimientos del sector. De hecho, progresivamente se irán demandando nuevas prestaciones a medida que vayan satisfaciéndose las más inmediatas. Técnicamente, se trata de un trabajo de minería de datos que requiere, para su mejor realización emplear técnicas acordes con las peculiaridades propias de la documentación jurídica porque de esa manera es como se podrán optimizar las características arquitecturales de los sistemas que se desarrollen.

Aunque es evidente que la solución debe proporcionarla la tecnología informática, la demanda de prestaciones debe surgir de la dirección de los organismos que tienen encomendada la custodia de la documentación. Es fundamental para el éxito de los resultados que la actitud con que se formule dicha demanda sea de satisfacción de necesidades, sin asumir las limitaciones que la tecnología informática pueda plantear en un momento dado, respecto de algún aspecto puntual. De hecho, la automatización de la consulta jurídica plantea problemas que requieren de un apreciable esfuerzo de investigación.

A modo de conclusiones concretas, se propone el esquema de los pasos a seguir para elaborar una propuesta de desarrollo de un sistema inteligente de ayuda a la consulta de jurisprudencia:

1. Establecer un compromiso contractual entre los organismos e instituciones participantes en el proyecto.
2. Crear un grupo de trabajo multidisciplinar (juristas, documentalistas e informáticos) para establecer cuáles son los servicios y utilidades que debe satisfacer el sistema de consulta. Así mismo, gestionar la consecución de ayudas económicas para la financiación del proyecto.
3. Especificar y diseñar el sistema informático; esto es, concretar hasta el último detalle las prestaciones del sistema informático, la interfaz de uso, el hardware de soporte, el software y elaborar el proyecto de realización en el cuál se establezca, además, el plan temporal de realización y la valoración económica. Ese paso es complejo y laborioso, al tiempo que previo a cualquier desarrollo informático de envergadura y que tenga que ser exitoso; es decir, es el paso que garantiza la calidad de los resultados. Consiste en:
 - ✍ Un exhaustivo estudio de requerimientos en el que se recogerá con detalle las características de la documentación de carácter judicial y los procesos de recogida y tratamiento de esta información.
 - ✍ Se realizará un análisis para precisar los servicios que se desea ofrecer relacionados con el tratamiento de la información al igual que las tecnologías necesarias para ofrecer dichos servicios. Para el estudio de estos



requerimientos se tendrá en cuenta el ámbito nacional, o incluso internacional, de operación del sistema.

- ✍ A continuación, tomando como base la especificación realizada se procederá al diseño del sistema. El diseño incluirá:
 - ✍ La interfaz gráfica y el estudio de los algoritmos necesarios para el tratamiento de la información de forma segura y eficaz.
 - ✍ También se incluirán con detalle los procedimientos y funciones implementados según un lenguaje de modelado y diseño (p.e. UML).
 - ✍ Se realizará un estudio del hardware necesario tanto a nivel de red como a nivel de almacenamiento y computación para que el sistema funcione correctamente.
 - ✍ De igual forma se ofrecerá un análisis a nivel de software incluyendo el sistema operativo, el gestor de bases de datos y los lenguajes de programación adecuados para la implementación del sistema. El trabajo de análisis y diseño se organizará en diversos documentos y prototipos software que recogerán de forma completa toda la arquitectura del sistema.



Referencias

- [REV,98] Reviriego, J. y Maciá, M.
Introducción general a la documentación jurídica
Manual de documentación jurídica (Maciá ED.)
Madrid: Síntesis, 1998
- [ALV,04] Alvite Díez, M.L.
Evolución de las bases de datos jurídicas en España
Anales de documentación, nº 7. pg: 7-27, 2004
- [PAE,94] Páez Mañá, J.
Bases de datos jurídicas: características, contenido, desarrollo y marco
legal
CINDOC, Madrid, 1994
- [PAE,95] Páez Mañá, J.
Comentarios sobre algunas particularidades de las bases de datos jurídicas
Actualidad informática Aranzadi, nº 16, pg: 4-10, 1995
- [LOP,84] López-Muñiz Goñi, M.
Informática jurídica documental
Díaz de Santos, Madrid, 1984
- [BIN,91] Bing, J.
Legal text retrieval and information services
Encyclopedia of Library and Information Science. New York, 1991
- [MAC,96] Maciá, M. La documentación de la Unión europea
Síntesis. Madrid, 1996
- [CAR,93] Caridad Sebastián, M.
Fundamentos teóricos en documentación automatizada: programa y
bibliografía
PPU. Barcelona, 1993
- [GAR,99] García Moreno, M.A.
De la telecomunicación a internet: la industria española de las bases de
datos
Fragua. Madrid, 1999