



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 3

**Trabajo de Diploma para optar por el Título de
Ingeniero en Ciencias Informáticas**

Título:

**Desarrollo de una Base de Datos para la Gestión Documental de las sentencias penales en e
Centro Nacional de Documentación e Información Judicial.**

Autora: Yasmín Peña Jimenez

Tutor: Ing. Martín Villalón Cruzata

Co-Tutor: Ing. Yordanis García Leiva

Ciudad de La Habana, junio del 2012

“Año 54 de la Revolución”

DECLARACIÓN DE AUTORÍA

Declaro ser la autora de la tesis titulada **Desarrollo de una Base de Datos para la Gestión Documental de las sentencias penales en el Centro Nacional de Documentación e Información Judicial** y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste se firma la presente a los _____ días del mes de junio del año 2012.

Yasmín Peña Jimenez.

Firma de la Autora

Ing. Martín Villalón Cruzata

Firma del Tutor

Ing. Yordanis García Leiva

Firma del Co-Tutor

Datos de Contacto

Tutor: Ing. Martín Villalón Cruzata

Ingeniero en Ciencias Informáticas. Graduado en la Universidad de las Ciencias Informáticas en el año 2009.

Correo: mvillalon@uci.cu

Teléfono: 8372269

Co-tutor: Ing. Yordanis García Leiva

Ingeniero en Ciencias Informáticas. Graduado en la Universidad de las Ciencias Informáticas en el año 2011.

Correo: ygleiva@uci.cu

Teléfono: 8373378

Agradecimientos

Son muchas las personas que merecen mi eterno agradecimiento por su apoyo incondicional en esta misión emprendida para convertirme en ingeniera y tal vez el espacio disponible para esto sea insuficiente, aún así existen los que no pueden dejar de mencionarse:

Mi querido Comandante Fidel, por su sistema educacional completo, puedo decir que he tenido la dicha de estudiar en escuelas geniales gracias a la Revolución Cubana. Le agradezco mucho por crear una universidad de excelencia donde personas de todas partes de Cuba pueden acudir y formarse como profesionales.

A mi papito querido, que siempre ha sido de carácter fuerte pero a su vez el hombre más fiel e importante en mi vida, dándome todo el apoyo y las fuerzas necesarias para seguir adelante, enseñándome que el trabajo ennoblece y el deber se cumple.

A mi mamita hermosa, por sus consejos y comprensión, por su lucha constante conmigo, guiándome siempre por el mejor camino y haciéndome buscar el lado positivo de los sucesos de la vida.

A mi hermanita linda, por sus correos constantes, haciéndome reír, llorar, reflexionar, aprender y entender. Por ser un ejemplo de cumplimiento laboral, y por estar en las buenas y en las malas con mi mamá y mi papá cuando yo no puedo debido a la distancia. Por entenderme y ayudarme, gracias Tutty Frutty.

A Hussein, mi Tito, mi sol, mi santito, mi cielito trelladito, mi amigo y mi amor, el hombre con el que espero compartir cada día de mi vida. Te amo mucho y les agradezco por todo, a ti y a tu familia.

A las hermanas que la vida me regaló como amigas y a la familia de cada una de ellas que son casi como mi familia también. Las amigas viejas que nunca han fallado: Yarima, Yoa. Las grandes amigas que encontré en la universidad: Nely, Ricnelys (la mita), gracias por sus consejos y regaños oportunos, por escuchar, por confiar, por el tiempo, las lágrimas y las risas, muchas gracias. Las quiero mucho.

A mi familia, mis abuelos que a pesar de que no les di el orgullo de tener una nieta bailarina, nunca perdieron la fe en mí, y me siguieron apoyando en todas mis decisiones.

A mis tíos por estar cerca siempre que los necesito, siempre atentos, preocupados.

A mis primos, la Nena, Yai, Alfredito, David, espero que se sientan orgullosos de tener una prima Ing.

A mi tutor Martín, por su amistad, su tiempo, por sus consejos y a mi co-tutor por su paciencia y dedicación.

Al tribunal por cada sugerencia y señalamiento en pos de mejorar la calidad de este trabajo.

A todas las personas que pusieron su fe en mí, que cuando me caí me levantaron, y me enseñaron a mantenerme en pie. Los que me ayudaron a finalizar mis estudios en la UCI, Eduardo, Raúl, Carlos Rafael, Manuel, Yvonne (la decana), Yalice, Yoan, Pascual.

A Heiler, Yonelbys, Berenice por todo y a los profesores que ayudaron en mi formación como profesional.

Los compañeros de cada grupo de los que formé parte, de cada apartamento, de cada ensayo, de cada seminario, de algún momento de mi vida en la UCI: Arislay, Annier, Abelito, Yoan, Elizabeth, Yojaimer, Ariannne, Alberto, Yoslenys, Deysi, Lisandra (triple L), Yadiani, Aimé, Dayana, Paul, Pedro, Adrián, Yiliam, El Rega, Espacio Abierto (grupo danzario).

A todos los mencionados y los que no es por falta de espacio. Muchas gracias por todo, forman parte de mi vida para siempre, los quiero mucho.

Dedicatoria

A mi papá, por aceptar mis decisiones en el ámbito profesional desde pequeña y por no perder la confianza en mí cuando no logré ser bailarina. Espero que te sientas orgulloso de mí, hoy soy Ing. en Ciencias Informáticas. Este título es para ti Lore, de tu hija que te quiere con la vida.

A mi mamá, por estar siempre para mí, por soportar tantos años separada de mí mientras yo encontraba mi camino. Este título te lo dedico a ti también mamita, desde lo más profundo de mí.

A mi hermanita, espero que te inspires a superarte algún día, al final recoges lo que siembras, si pasas trabajo cosechando algo, más delicioso será su sabor al final, esta es mi experiencia propia y te la transmito a través de este logro.

A Hussein, por la confianza, el ánimo y todo su amor. Para ti también es este título mi angelito.

A mi primo Alfreto, que está en medio de su camino a convertirse en ingeniero, no dejes que nada te impida cumplir tu propósito, tú puedes lograrlo, a ti también te dedico esta tesis.

A mis vecinos queridos, Delia, Madelaine, Elida, Rubén, Milagro, Marlenis, Belén, Mimita, Annia, Yoendris, Albertico. Se hizo ingeniera la muchachita del 1er piso y les quiere dedicar su tesis a ustedes también.

A mis abuelos por mantener su confianza en mí y apoyarme siempre.

A mis abuelitas por su dulzura y comprensión en todo momento.

A mis tíos, mis tías, mis primos, mis primas.

A toda mi familia, por no perder la confianza ni las esperanzas de que lograra ser profesional un día, espero que se sientan orgullosos de mí.

A mis amigas: Yoa, Ric, Neki y Yarima, por todo lo vivido, para ustedes también va mi título.

A Isabel, Made, Nadeen, Mariela, Idolidia, por traer al mundo a esas personitas maravillosas y hacerme parte de sus familias también. Ya tienen otra Ing. en la familia.

A todas las personas que siempre me han deseado lo mejor, ayudándome y guiándome, a todo ellos les dedico el fruto de estos años de esfuerzo, sacrificio y aprendizaje.

RESUMEN

El presente trabajo de diploma tiene como objetivo desarrollar una base de datos documental que almacene las sentencias penales del Centro Nacional de Documentación e Información Judicial (CENDIJ) para acceder al contenido íntegro de la información persistente en las mismas. Para el desarrollo de la investigación se realizó un amplio estudio sobre los temas relacionados con las bases de datos documentales (BDD), Informática Jurídica y Lenguajes Documentales, lo que permitió identificar a la Informática Jurídica Documental como la ciencia que permite resolver el problema central de la presente investigación.

En el proceso investigativo se realizaron entrevistas al personal que labora en el CENDIJ y profesionales de la Universidad de las Ciencias Informáticas (UCI), especializados en el manejo de información jurídica, con el objetivo de conocer cómo funciona el CENDIJ y analizar la importancia del uso de una BDD en una entidad como esta.

En la propuesta de solución se hace uso de la metodología para el desarrollo de BDD definida por el Dr. Lluís Codina. Las herramientas utilizadas para la realización de la base de datos son Visual Paradigm for UML en su versión 8.0, PostgreSQL versión 9.1, PGAdmin III versión 1.14.1. Durante la validación del trabajo se emplean los software Data Generator for PostgreSQL versión 2.2.0.1 y Apache JMeter en su versión 2.3.1. Utilizando estas herramientas se verificó la integridad de los datos, la capacidad de almacenamiento que tiene la BDD creada, así como su robustez cuando la carga es extrema.

Palabras claves: base de datos documental, informática jurídica, lenguaje documental, sentencias penales.

ÍNDICE

RESUMEN	IV
INTRODUCCIÓN	1
Capítulo 1: Fundamentación Teórica.....	6
1.1 Introducción.....	6
1.2 Bases de Datos Documentales.....	6
1.2.1 Clasificación de las Bases de Datos Documentales.....	7
1.2.2 Tipología de las Bases de datos Documentales.....	9
1.2.3 Sistemas de recuperación de información.	10
1.2.4 Recuperación de información en BDD.	12
1.3 Informática Jurídica Documental.....	13
1.3.1 Problemas de la informática jurídica documental.	15
1.4 Lenguajes Documentales.	16
1.4.1 Objetivos y Funciones de los Lenguajes Documentales.	16
1.4.2 Tipologías de los Lenguajes Documentales.....	17
1.5 Metodologías para el diseño de bases de datos.	19
1.5.1 Metodología de creación de bases de datos documentales.	19
1.6 Herramientas de Modelado.....	21
1.6.1 Rational Rose.	21
1.6.3 Visual Paradigm.....	22
1.7 Patrones en el desarrollo de Base de Datos.....	23
1.8 Gestores de Base de Datos.....	26
1.8.1 Herramientas de Administración de base de datos.	29
1.8.2 Herramientas para la validación de BD.....	31
1.8.3 Sistemas Gestores de Base de Datos Documentales (SGBDD).	32
1.9 Índices en BD	36
1.10 Conclusiones parciales.....	37
Capítulo 2: Diseño e implementación de la Base de Datos.....	38
2.1 Introducción.....	38
2.2 Sistema de Actividades Humanas en el CENDIJ.....	38
2.3 Sistema de Conocimientos en el CENDIJ.....	39

2.4 Requisitos funcionales.	39
2.5 Requisitos no funcionales.	42
2.6 Modelo conceptual.	43
2.7 Modelo Entidad Relación.	44
2.7.1 Estándares de nomenclatura en la base de datos.	44
2.7.2 Patrones de diseño usados.	45
2.7.3 Esquema Relacional.	45
2.8 Diccionario de Datos.	47
2.9 Optimización de base de datos.	49
2.9.1 Optimización de consultas mediante el uso de índices.	49
2.10 Pasos a seguir para diseñar la BD que permita la gestión documental de todos los documentos persistentes en el CENDIJ.	51
2.11 Conclusiones parciales.	54
CAPÍTULO 3. Validación de la base de datos.	55
3.1 Introducción.	55
3.2 Validación teórica.	55
3.2.1 Integridad de la información.	55
3.2.2 Normalización de la base de datos.	57
3.2.3 Redundancia de los datos.	58
3.3 Seguridad de la base de datos.	59
3.4 Validación Funcional.	59
3.4.1 Pruebas de Volumen.	60
3.4.2 Prueba de Carga.	61
3.5 Conclusiones parciales.	63
Conclusiones.	64
Recomendaciones.	65
Bibliografía Referenciada.	66
Bibliografía Consultada.	68
Anexos.	70

INTRODUCCIÓN

Durante siglos la gestión documental se ha realizado manualmente a través del personal que labora en bibliotecas y archivos, haciendo uso de herramientas como libros de registros, carpetas, archivadores, cajas o estanterías que guardan toda la documentación almacenada en estos locales. Con el desarrollo de las Tecnologías de la Informática y las Comunicaciones (TIC) surge la digitalización de toda la documentación que hasta el momento se almacenaba en formato duro.

En la década de los años 60 del siglo XX se crean las primeras bases de datos (BD), estos medios se convierten rápidamente en la principal fuente para el almacenamiento de la información digital. Con el uso de las computadoras y el diseño de bases de datos, el proceso de gestión de documentos en formato digital fue adquiriendo grandes avances en cuanto al almacenamiento y rápida recuperación de información.

En la actualidad existen diversos sistemas informáticos destinados a la gestión documental que permiten realizar búsquedas sofisticadas y una rápida recuperación de información. Entre los sistemas de mayor impacto se encuentran los que hacen uso de las Bases de Datos Documentales (BDD). Estas se ocupan de la gestión de documentos, optimizando el almacenaje y facilitando su recuperación y acceso a la información que estos contienen, por el contrario, otras bases de datos solo cubren un primer paso en la búsqueda de información, ya que el usuario solo obtiene la noticia de la existencia o no de un documento. Existen disímiles y casi innumerables áreas del conocimiento, una de estas es el Derecho y la gestión documental no deja de estar presente en ella, por lo que el uso de la informática se hace indispensable en el perfeccionamiento de este proceso.

Una de las ciencias dedicadas a la aplicación del avance de las TIC en el sistema jurídico es la Informática Jurídica, la cual se divide en tres ramas fundamentales, la Informática Jurídica de Gestión, la Informática Jurídica Decisional y la Informática Jurídica Documental. Actualmente el desarrollo alcanzado en esta última rama ha estado avalado principalmente por el uso de sistemas informáticos que utilizan BDD auxiliadas por herramientas lingüísticas como los lenguajes documentales. Entre las naciones de mayor avance en el tema se pueden citar en el hemisferio europeo países como España e Italia y en el continente americano los más destacados son Colombia y Chile respectivamente.

En Cuba, el mayor desarrollo en relación a la gestión de documentación jurídica tiene lugar en el Centro Nacional de Documentación e Información Judicial (CENDIJ), este es una biblioteca

especializada cuyo objetivo principal es la prestación de servicios a todos los usuarios, independientemente de su cargo u ocupación en la sociedad. Tiene como misión dotar a jueces, abogados, fiscales y profesionales en general, de información en correspondencia con sus intereses, para contribuir de manera ética y eficaz a su preparación para la toma de decisiones en la misión de administrar justicia. Dentro de los servicios que presta se destacan la búsqueda de documentos digitales y referencias a libros u otro tipo de información en formato duro, además de los servicios de búsqueda especializada en internet.

Este centro permite el estudio y análisis de información jurídica, a través de la vinculación de medios informáticos con los lenguajes documentales. Entre la información jurídica analizada en el CENDIJ se encuentran las disposiciones del consejo del gobierno, legislaciones, doctrinas, jurisprudencias¹ y sentencias almacenadas en la base de datos del sistema informático allí presente.

Dentro de esta BD se almacenan sentencias de diferentes calificaciones, estas pueden pertenecer a la materia Laboral, Económico, Administrativo, Civil y Penal. Las sentencias penales pueden ser de diferentes tipos: sentencia de Privación de libertad, sentencia de Tratamiento Correccional Con Internamiento (TCCI), sentencia de Tratamiento Correccional Sin Internamiento (TCSI), sentencia de Limitación de libertad, sentencia de multa y sentencia por retirada. Las sentencias llegan por un canal legal y oficial al CENDIJ, luego las bibliotecólogas inician el proceso de indización² y posteriormente pasan a guardar la sentencia en la BD.

El sistema informático vigente en el CENDIJ goza de herramientas que permiten gestionar la información pero se encuentran aisladas. Posee una BD referencial en la que se hace difícil la consulta de la información, puesto que los usuarios no tienen manera de acceder al contenido de los documentos almacenados, solo pueden conocer si existen o no. Cuenta con otra BD que presenta como principal problema la cantidad de información que se recupera al realizar la búsqueda, pues dependiendo del tipo de búsqueda, se puede obtener un gran número de documentos o uno bien reducido. Este fenómeno recibe el nombre de ruido y silencio documental respectivamente. Teniendo en cuenta que esta BD no permite realizar búsquedas en el contenido de los documentos guardados,

¹ **Jurisprudencia:** resultado de un juicio, incluyendo toda la información generada en el proceso jurídico del mismo.

² **Indización:** técnica del análisis documental para representar y describir el contenido de los documentos, mediante conceptos principales contenidos en ellos (palabras clave) o vocabularios controlados (descriptores, términos o encabezamientos de materia), con el fin de guiar al usuario en la recuperación de los documentos que necesita.

para encontrar cierta información es necesario revisar cada documento jurídico, en este caso, cada una de las sentencias penales.

La **problemática** antes descrita ha generado la necesidad de desarrollar una investigación que dé respuesta al siguiente **problema a resolver**:

¿Cómo almacenar las sentencias penales en el CENDIJ de manera que permita acceder al contenido íntegro de la información persistente en las mismas?

Para dar solución a la problemática planteada se define como **objeto de estudio** la gestión documental en la informática jurídica.

Se ha determinado como **objetivo general** de la investigación: desarrollar una base de datos documental que almacene las sentencias penales en el CENDIJ de manera que permita acceder al contenido íntegro de la información persistente en las mismas.

El **campo de acción** se centra en el proceso de gestión documental de las sentencias penales en el CENDIJ.

Se tiene como **idea a defender** la siguiente propuesta: la implementación de una base de datos documental que almacene las sentencias penales en el CENDIJ permitirá acceder al contenido íntegro de la información persistente en las mismas.

Objetivos Específicos:

1. Definir el marco teórico-referencial de la investigación.
2. Identificar el modelo conceptual que represente al dominio de conocimientos que contienen las sentencias penales.
3. Diseñar el modelo de datos.
4. Implementar el modelo de datos.
5. Validar los resultados obtenidos.

Para dar cumplimiento a los objetivos anteriores se plantean las siguientes **tareas de la investigación**:

1. Estudio del estado del arte del tema de la investigación.
2. Caracterización de la estructura documental de las sentencias penales en el CENDIJ
3. Definición de un modelo conceptual que represente el dominio de conocimiento que contienen las sentencias penales en el CENDIJ.
4. Selección de los patrones de diseño y estándares de nomenclatura para el modelado de los datos.
5. Estudio de los métodos para optimizar las búsquedas documentales.
6. Diseño del modelo de datos.
7. Implementación del modelo de datos.
8. Realización de las pruebas teóricas a la base de datos.
9. Realización de las pruebas funcionales a la base de datos.

Durante la realización de este trabajo se utilizan diferentes **métodos científicos** para un mejor desarrollo de la investigación:

Métodos teóricos:

Analítico-Sintético: para realizar el estudio del estado del arte es necesaria la investigación de la mayor cantidad de información posible y a través del análisis de la misma lograr la extracción de los elementos más importantes que se relacionan con el objeto de estudio.

Inductivo-Deductivo: a partir del conocimiento general adquirido a través de los estudios realizados se deduce, por medio de un razonamiento lógico que la mejor forma de resolver el problema de esta investigación es realizando la implementación de una BDD.

Análisis Histórico-Lógico: para la realización de este trabajo, se hace un estudio previo del desarrollo de las bases de datos documentales (BDD) a nivel internacional. Además se realiza un análisis al desarrollo y evolución de las BD presentes en el CENDIJ, para tener una visión tanto de los principales problemas como de los avances obtenidos dentro de este centro.

Métodos empíricos:

La entrevista: Permitted realizar entrevistas a especialista del CENDIJ y profesionales especializados en el tema de la informática jurídica, así como trabajadores del centro CEGEL, con el objetivo de fundamentar las ideas que permitieron dar cumplimiento al objetivo general de la investigación.

El presente documento está estructurado en tres capítulos, que incluyen todo lo relacionado con el trabajo investigativo, así como el diseño e implementación y validación de la BD:

CAPÍTULO 1. Fundamentación teórica: contiene los aspectos esenciales para entender el entorno del problema a resolver. Se describen los conceptos fundamentales asociados al dominio del problema, se realiza una investigación sobre los sistemas informáticos existentes en el ámbito nacional e internacional afines con las bases de datos documentales que gestionan la información relacionada con las sentencias de la materia penal, principalmente en el CENDIJ. Se muestra además un estudio de las herramientas y tecnologías que se utilizan para desarrollar la BDD que ocupa el eje central de esta investigación.

CAPÍTULO 2. Diseño e implementación del sistema: se describen los conceptos fundamentales relacionados con la metodología a utilizar. Se realiza el modelo conceptual de la BD, se ejecuta la especificación y descripción de los requisitos capturados, se obtiene el Esquema Relacional, el Diccionario de Datos y finalmente se realiza la optimización de la base de datos a través de la utilización de índices.

CAPÍTULO 3. Validación de la base de datos: incluye la realización de las pruebas que permiten la validación teórica y funcional de la BD a obtener.

Además contiene las conclusiones de la investigación, las recomendaciones, las referencias bibliográficas y la bibliografía consultada para dar solución al problema planteado en la presente investigación.

Capítulo 1: Fundamentación Teórica.

1.1 Introducción.

La gestión documental data de hace miles de años, pues ya desde esa época existían personas dedicadas única y exclusivamente a recolectar, conservar, preservar y tener disponible la información existente. Actualmente se han logrado desarrollar técnicas impresionantes para este fin, dentro de las que se encuentran las Bases de Datos Documentales (BDD); es por esta razón que en este capítulo se exponen importantes conceptos sobre estas y otros aspectos importantes dentro de la gestión documental en la informática jurídica, además se caracterizan las herramientas que se utilizarán para la posterior implementación de un modelo de datos basado en las mismas.

1.2 Bases de Datos Documentales.

El señor C.J. Date es autor, conferencista, investigador y consultor independiente, especializado en la tecnología de bases de datos y plantea que: Una Base de Datos (BD) es un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa dada.

Es posible considerar a la propia base de datos como una especie de armario electrónico para archivar; es decir, es un depósito o contenedor de una colección de archivos de datos computarizados. Los usuarios del sistema pueden realizar una variedad de operaciones sobre dichos archivos, como:

- ✓ Agregar nuevos archivos vacíos a la base de datos.
- ✓ Insertar datos dentro de los archivos existentes.
- ✓ Recuperar datos de los archivos existentes.
- ✓ Modificar datos en archivos existentes.
- ✓ Eliminar datos de los archivos existentes.
- ✓ Eliminar archivos existentes de la base de datos. (Date, 2001)

Existen tres grandes modelos de bases de datos según el contenido que almacenan en sus registros. Estas pueden ser:

Bases de datos con información factual: recogen informaciones muy concretas y actuales, generalmente de carácter numérico: datos estadísticos, series históricas, resultados de encuestas, convocatorias de becas o subvenciones, convocatorias de eventos, ofertas de empleo y toda información de esta índole. (Yunta, 2001)

Directorios: almacenan datos sobre personas o instituciones especializadas en una actividad o materia concreta. Hay directorios de profesionales, investigadores, centros de investigación,

bibliotecas, de revistas científicas, empresas, editoriales o cualquier otra información similar. (Yunta, 2001)

Bases de datos documentales: cada registro se corresponde con un documento, sea este de cualquier tipo: una publicación impresa, un documento audiovisual, gráfico o sonoro, un documento de archivo, un documento electrónico o libros que contengan la información solicitada. Este tipo de bases de datos, es adecuado para almacenar información no estructurada, es decir, no dividida en campos. (Yunta, 2001)

1.2.1 Clasificación de las Bases de Datos Documentales.

Según el contenido completo o no de los documentos que describen sus registros: (Yunta, 2001)

Bases de datos de texto completo: son aquellas que estén constituidas por los propios documentos en formato electrónico, por un volcado completo de su texto. Pueden incorporar además campos en los que se contiene la información fundamental para facilitar su descripción y recuperación. En estos sistemas la operación de búsqueda (que puede abarcar la totalidad del texto) y la consulta del documento se producen sin salir del propio sistema de información.

Archivos electrónicos de imágenes: están constituidos por referencias que permiten un enlace directo con la imagen del documento original, sea éste un documento iconográfico (fotografías, imágenes de televisión) o un documento impreso digitalizado en formato de imagen. En estas bases de datos normalmente la búsqueda está limitada a los campos de la referencia bibliográfica y no se pueden localizar otros términos presentes en el texto completo del documento original.

Bases de datos referenciales: sus registros no contienen el texto original sino tan sólo la información fundamental para describir y permitir la localización de documentos impresos, sonoros, iconográficos, audiovisuales o electrónicos. En estos sistemas de información sólo se puede obtener referencias sobre documentos que habrá que localizar posteriormente en otro servicio (archivo, biblioteca, fototeca, fonoteca) o solicitar a un servicio de suministro de documentos. Sin embargo, una base de datos referencial puede incluir campos que faciliten la localización del documento (bibliotecas, signaturas, direcciones en Internet) o incluso enlaces directos para obtener directamente el original a través de otro programa (tratamiento de texto, navegador de Internet).

Según la cobertura temática se pueden clasificar en: (Yunta, 2001)

Bases de datos científico-tecnológicas: contienen información destinada a los investigadores de cualquier ámbito científico o técnico. A su vez, este grupo puede dividirse en:

- ✓ *Bases de datos multidisciplinarias:* abarcan varias disciplinas científicas o técnicas.
- ✓ *Bases de datos especializadas:* recopilan y analizan documentos pertinentes para una disciplina o subdisciplina concreta: investigación biomédica, farmacéutica, química, agroalimentaria, social, humanística, jurídica y todas las demás existentes.

Bases de datos económico-empresariales: contienen información de interés para empresas, entidades financieras o corporaciones.

Bases de datos de medios de comunicación: contienen información de interés para los profesionales de medios de comunicación de masas: prensa, radio y televisión.

Bases de datos del ámbito político-administrativo y jurídico: contienen información de interés para los organismos de la administración y los profesionales del Derecho: legislación, doctrina y jurisprudencia.

Bases de datos del ámbito sanitario: además de las propias del primer grupo especializadas en ciencias de la salud, existen otros sistemas con información de interés sanitario: historiales médicos, archivos hospitalarios

Bases de datos para el gran público: contienen información destinada a cubrir necesidades de información general, de interés para un gran número de usuarios.

Según el modo de acceso la clasificación quedaría: (Yunta, 2001)

Bases de datos de acceso local: para consultarlas es necesario acudir al organismo productor, a su biblioteca o centro de documentación. Pueden ser consultables en una sola computadora o en varios puntos de una red local.

Bases de datos en CD-ROM: pueden adquirirse por compra o suscripción bien directamente por un particular o por una biblioteca o centro de documentación que permita su consulta a sus usuarios.

Bases de datos en línea: pueden consultarse desde cualquier ordenador conectado a través de un modem a la línea telefónica o a través de un servidor a Internet. La consulta puede ser libre (gratuita) o exigir la solicitud previa de un clave personal de entrada (denominada comúnmente con el término inglés password). Para obtener un password o contraseña, puede exigirse la firma de un contrato.

Es importante aclarar en este punto que la BD presente actualmente en el CENDIJ es de acceso local, aunque también posee ediciones en CD-ROM que son enviadas a las entidades inferiores dependientes de la misma, como son los tribunales provinciales y los tribunales municipales.

1.2.2 Tipología de las Bases de datos Documentales.

Según su cobertura documental

Bases de datos centradas en un único tipo de documento: su objetivo es recopilar y permitir la localización de un tipo documental muy concreto. Hay bases especializadas en patentes, tesis doctorales, informes, artículos de revista y libros. (Yunta, 2001)

Bases de datos que reúnen varios tipos de documentos: su objetivo es dar información sobre una disciplina, incorporando para ello diferentes tipologías documentales. (Yunta, 2001)

Según el modelo de tratamiento documental.

Bases de datos de sumarios o sin análisis de contenido: se componen de referencias bibliográficas sencillas, en las cuáles el productor se limita a grabar los datos de la propia fuente y no realiza ningún análisis del contenido. Incorporan solamente los datos descriptivos fundamentales para localizar el documento, presentes en el sumario: autor, título y datos de la fuente. En ocasiones pueden incorporar el resumen editado con el propio documento original, generalmente elaborado por el propio autor. En la mayoría de estas bases de datos la búsqueda por materias sólo puede realizarse a través de las palabras del título del artículo. En muchas áreas temáticas los títulos pueden resultar poco significativos. Los resultados obtenidos en una búsqueda tienen cierta pertinencia, pero la escasez de puntos de acceso para la búsqueda por materias no garantiza la exhaustividad. (Yunta, 2001)

Catálogos de bibliotecas: son bases de datos que responden estrictamente a los fondos contenidos en una biblioteca o en una red de bibliotecas (catálogos colectivos). Tienen una alta homogeneidad, gracias a la aplicación de normas internacionales de Catalogación. En estos sistemas se puede garantizar la búsqueda por ejemplar conocido. Por el contrario, la búsqueda por materias es en ocasiones dificultosa y con algunas variaciones entre diferentes catálogos. Puede realizarse a través de la clasificación o de los encabezamientos de materias. (Yunta, 2001)

Bases de datos con análisis documental más completo: son sistemas de información que incorporan un mayor número de puntos de acceso para facilitar la localización por materias. Cada registro bibliográfico incluye bien un resumen del contenido del documento original y/o un conjunto de

conceptos o términos representativos de los temas tratados en el mismo. Se denomina indización al proceso de asignación a un documento de estos términos que describen su contenido y que constituyen entradas en un índice de materias (palabras clave o descriptores).

Estas BD pueden ser independientes de un fondo documental concreto, ya que su objetivo no es describir una colección sino proporcionar el mayor número de referencias potencialmente relevantes para cubrir unas necesidades de información. (Yunta, 2001)

Dentro de este grupo de bases de datos no hay un modelo único, sino una gran variedad de posibilidades, según los criterios de análisis documentales adoptados por el productor del sistema de información: (Yunta, 2001)

- *Bases de datos con clasificación y resúmenes*: en estos sistemas de información la búsqueda por materias se realiza generalmente a través de las palabras del título y del resumen. Tienen una alta cantidad de puntos de acceso, pero en ocasiones la búsqueda puede ser muy dificultosa.

- *Bases de datos con clasificación e indización por descriptores o palabras*: se caracterizan por permitir una mayor pertinencia de la recuperación. Sin embargo para sacarles un óptimo rendimiento es conveniente que el usuario conozca el listado de términos utilizados para la descripción del contenido de los documentos.

- *Bases de datos con clasificación, indización y resúmenes*: teóricamente reúnen las ventajas de los dos grupos anteriores. Dentro de este modelo conviven bases de datos con diferentes criterios de indización (genérica o específica) y diferentes modelos de resúmenes (informativos o indicativos).

Índices de citas: son sistemas de información en los que, además de extraer los datos de descripción de los documentos, se vacían sistemáticamente las referencias bibliográficas citadas en los artículos de las revistas científicas. Su mayor utilidad radica en servir de apoyo para la elaboración de estudios bibliométricos y de productividad científica: autores, instituciones y publicaciones científicas más citadas; pervivencia de la literatura científica. (Luis, 2001)

1.2.3 Sistemas de recuperación de información.

La recuperación de información es el siguiente nivel a la determinación de las necesidades de almacenar la misma. La recuperación se hace mediante consultas a la base de datos donde se guarda la información organizada mediante un lenguaje de consulta adecuado. Es necesario tener en cuenta los elementos claves que permiten realizar la búsqueda, proporcionando un alto grado de precisión (índices, palabras claves y fenómenos que pueden surgir en el proceso de búsqueda como ruido o silencio documental). (Ernest Abadal, 2005)

El objetivo primordial de la recuperación de información es desarrollar, mediante el estudio, métodos que, ya sean algorítmicos o intelectuales, faciliten el siguiente grupo de operaciones: (Herrera, 2006)

1- **Indización:** es una operación realizada de forma intelectual que a su vez se divide en análisis, que es la identificación de los temas y conceptos más importantes del documento y normalización, cuyo principal objetivo es transformar los conceptos que expresan el contenido del documento en sus descriptores.

2- **Selección:** es la identificación de un conjunto de documentos relevantes dentro de un grupo mayor para resolver una necesidad de información dada. También es denominada recuperación, la cual, siendo la parte más significativa del proceso, a veces sirve para dar nombre a todo.

3- **Ordenación:** determinación del orden más adecuado de presentación al usuario del grupo de documentos recuperados o seleccionados. La idea es ofrecerle los documentos en un orden decreciente por relevancia debido a la probabilidad de satisfacer la necesidad de información. A este proceso se le denomina ranking.

4- **Interconexión:** es el establecimiento de un conjunto de relaciones hipertextuales, caminos, y en general, estructuras de navegación en secciones del mismo o de distintos documentos.

5- **Categorización:** Es la asignación de cada documento a un grupo, clase o subclase de un cuadro de clasificación, taxonomía u ontología.

6- **Abstracción:** es la producción de resúmenes de los documentos los cuales dependiendo de las circunstancias podrían sustituir la lectura del documento completo.

7- **Visualización:** es la representación de forma gráfica de un conjunto de informaciones, conceptos o procesos no necesariamente representativos.

Los SRI siguen como pauta un modelo conceptual definido que representa los procesos y entidades que interactúan en el proceso de recuperación de información.

Modelos de un Sistemas de recuperación de información.

Un Sistemas de recuperación de información (SRI) se puede construir utilizando varios modelos de búsqueda y su nombre va dado por la forma en la que realizan dicha búsqueda:

Modelo Booleano Puro: el resultado de una ecuación de búsqueda booleana es un conjunto de documentos relevantes en el caso de que los haya. Estos son seleccionados siguiendo la lógica booleana la cual plantea que el documento es verdadero cuando contiene al menos un término de la pregunta si el operador es un OR, cuando contiene todos los términos de la pregunta si el operador es un AND o cuando no contiene ninguno de los términos de la pregunta (cuando es un NOT). Plantea la

lógica booleana que solo existen verdaderos y falsos así que en este caso solo habrán dos conjuntos de documentos: relevantes y no relevantes lo cual no permite establecer grados de importancia en estos. Otro de los problemas a los que se enfrenta este tipo de modelo es que son anti-intuitivos pues a un usuario poco experimentado le son confundibles los términos booleanos. (Ernest Abadal, 2005)

Modelo Vectorial: plantea que en un espacio vectorial, donde la cantidad de vectores está dada por los términos de indización, y que pueden tomar valores de ceros y unos, todos los documentos situados cercanos a la “pregunta” en el espacio vectorial serán semejantes a esta, siendo este umbral de semejanza el que define los documentos relevantes o no. Este modelo permite situar a los documentos por orden de relevancia poniendo en primer lugar a los que cumplen todos los términos. Este modelo sin embargo es, en la práctica, poco implementado, pues requiere de muchos recursos computacionales para ejecutarse y presenta problemas de recálculo cada vez que se añaden nuevos documentos. (Ernest Abadal, 2005)

Modelo Booleano-Vectorial: en este modelo al igual que el anterior, los documentos y las preguntas se presentan como vectores pero en lugar de calcular su similitud sobre la base de clústeres y espacios vectoriales lo hace contando los elementos en común de los vectores respectivos. La selección del documento se realiza mediante la lógica booleana pero se organiza mediante el método vectorial. (Herrera, 2006)

1.2.4 Recuperación de información en BDD.

Un mismo sistema de información puede estar accesible desde distintos soportes y generalmente todos los SRI permiten realizar diferentes modalidades de búsqueda. Dentro de las que se encuentran:

Búsqueda directa: se teclea directamente una o varias palabras en el espacio reservado para ello por el sistema de interrogación en la base de datos. Se distinguen dos tipos dentro de esta modalidad: (Ernest Abadal, 2005)

- ✓ **Interrogación en texto libre:** el usuario realiza su consulta sin tener en cuenta la estructura de campos de los registros de la base de datos.
Generalmente, el sistema busca el término o los términos solicitados en la totalidad de los campos de la base. En algunas BD esta búsqueda se realiza solamente en determinados campos, aquellos que el administrador de la misma ha establecido.
- ✓ **Interrogación en campos individuales:** el usuario realiza su consulta directamente sobre el campo o campos que ha seleccionado previamente: autor, revista, descriptores y todos los demás que estime convenientes.

Búsqueda a través de índices: en vez de teclear un término, el usuario visualiza un diccionario o índice alfabético de las entradas de todos los campos o de uno concreto (índice de autores, descriptores) y selecciona los más adecuados a su búsqueda. El proceso puede parecer más lento pero es más seguro, ya que la búsqueda se adapta mejor al contenido real de los registros de una base de datos. (Ernest Abadal, 2005)

En esta posibilidad hay que tener en cuenta si el sistema trabaja con uno o varios índices y cómo se realizan los mismos: (Ernest Abadal, 2005)

- ✓ **Índices por palabras:** incluyen solamente unitérminos (conceptos expresados en una sola palabra). Se elaboran, por ejemplo, con todas las palabras extraídas del título o del resumen de los documentos. La mayoría de los productores de bases de datos establecen una lista de “palabras vacías”, preposiciones y/o artículos para reducir el tamaño de estos índices. Los términos presentes en esta lista de palabras vacías no serán consultables.
- ✓ **Índices por frase:** incluyen entradas compuestas por varias palabras. Por ejemplo: apellidos y nombre del autor, descriptores o encabezamientos de materias.

Búsqueda jerarquizada: la interrogación se realiza a través de una estructura jerárquica. Es decir, a partir de un concepto genérico se pueden localizar, no sólo los registros en los que aparece dicho término, sino además todos aquellos en los que figure algún concepto más específico de su campo semántico. (Ernest Abadal, 2005)

Esta opción de búsqueda sólo será posible en aquellos sistemas de información que cuenten con campos codificados o que utilicen lenguajes estructurados jerárquicamente en la clasificación o en los descriptores.

Búsqueda a través de códigos: en determinados campos de algunas bases de datos la interrogación no se realiza a través de un texto sino a través de códigos numéricos o alfanuméricos. Pueden ser campos codificados la clasificación, el idioma, así como la tipología documental. (Ernest Abadal, 2005)

1.3 Informática Jurídica Documental.

La informática Jurídica es el conjunto de aplicaciones de la informática en el ámbito del Derecho. En sus primeros años, la Informática Jurídica se presentó en los términos de una Informática Documentaria de carácter jurídico, o sea, creación y recuperación de información que contenía datos jurídicos (leyes, jurisprudencias, doctrinas) o al menos de interés jurídico. (Pérez-Luño, 2006)

La informática jurídica, vista como una forma de análisis u ordenación de la información jurídica, está dividida en 3 ramas fundamentales: (Valdes, 2005)

1. Informática jurídica documental.
2. Informática jurídica de control y/o gestión y/o administración.
3. Informática jurídica de ayuda a la decisión y/o metadocumental y/o metadecisional.

Enfocando el desarrollo de esta investigación específicamente a la gestión documental, se centra la atención en la *informática jurídica documental*, esta tiene por objeto la automatización de los sistemas de información relativos a las fuentes de conocimiento jurídico: legislación, jurisprudencia y doctrina. En las circunstancias actuales el acceso de los profesionales del derecho a sistemas adecuados de documentación jurídica informatizada resulta del todo imprescindible.

De la Informática Jurídica Documental se desprende el análisis de la información contenida en documentos jurídicos para la formación de bancos de datos documentales. Como norma de la informática jurídica existe la aplicación de técnicas informáticas a la documentación jurídica en los aspectos sobre el análisis, archivo, y recuperación de información contenida en la legislación, jurisprudencia, doctrina o cualquier otro documento con contenido jurídico relevante.

Además de otras áreas comprende la aplicación de técnicas documentales, entendida la documentación como el acto de reunir datos sobre un tema dado y el tratamiento de estos en vistas a su definición. También comprende el análisis documental, el cual puede ser definido como un conjunto de ejecuciones realizadas para representar el contenido de un documento de forma distinta a la original, con el fin de facilitar la consulta o la búsqueda futura de información en estos, permitiendo las operaciones de recuperación y localización del contenido informativo. (Valdes, 2005)

Aspectos relevantes para el desarrollo de la informática jurídica documental

1.- La aplicación técnico-jurídica: se ajusta a la metodología especial de análisis de unidades de información de acuerdo con el sistema adoptado previamente; es decir, habla de sistemas de tratamiento y recuperación de la información y en ellos se aplican comúnmente los siguientes sistemas: (Pérez-Luño, 2006)

- a) Indexación:** consiste en la elaboración de una lista rígida de descriptores a través de la calificación de la información contenida en un documento fuente, mediante el descriptor o descriptores que se consideran apropiados, se identifica la información por medio de la designación de una o varias palabras claves o locuciones claves (descriptores), tomados de una lista previamente elaborada de acuerdo al tipo de información de que se trate.

- b) **Full text o Texto completo:** consiste en el almacenamiento del texto integral en la computadora con el fin de recuperar la información contenida en él por cualquiera de las materias a que hace referencia.
- c) **Abstract o Resumen:** es el acto cuya información obtenida en un acta fuente, es organizada en forma lógica a través del empleo de restrictores de distancia con el fin de lograr su recuperación, así como su presentación sintética.

2.- La formación de Bancos de Datos: cuyo punto de partida pueden ser archivos mensuales sistematizados, ya sea sectorizados e integrales.

3.- La utilización de lenguajes (vocabularios) o mecanismos de recuperación de información con apoyos de instrumentos lingüísticos: esta rama de la informática jurídica está constituida por bases de datos de información jurídica.

Para hablar de informática jurídica, es necesario identificar de la información, que en este caso sería la jurídica, toda aquella pertinente al caso concreto; teniendo esta, es determinante tratarla por medio de la estructuración con la aplicación de la lógica, o la argumentación para posteriormente mediante los instrumentos lingüísticos apropiados, incorporarla a la computadora. No será informática jurídica documental la simple incorporación de textos jurídicos a una computadora.

1.3.1 Problemas de la informática jurídica documental.

El tratamiento automatizado de la documentación jurídica solamente puede lograr sus objetivos de celeridad, exactitud y exhaustividad, si resuelve los problemas que plantea su implantación. Dichos problemas revisten distinto alcance y significado: unos son de naturaleza predominantemente técnica, esto se refiere a las exigencias de adecuación del *hardware* y el *software*, y a las metas propuestas a los sistemas de documentación jurídica automatizada, para garantizar su operatividad; otros poseen una dimensión ideológica o política y tienen como objeto la formulación de una auténtica política de la informática jurídica documental, a la delimitación de la estructura de funcionamiento de los sistemas, así como establecer su finalidad. Esta distinción, en cierto modo, permite situar la problemática técnica en el ámbito de la informática jurídica, mientras que muchas de las cuestiones relacionadas con la política de los sistemas informatizados de documentación jurídica se insertarán en la sede teórica del Derecho Informático. (Carrasco, 2006)

1.4 Lenguajes Documentales.

Un lenguaje documental es un lenguaje no natural, aunque utiliza los signos de este. Esos signos adquieren valor semántico por medio de su normalización y de las reglas morfosintácticas³ que lo articulan. Son precisamente esas reglas las que le dan categoría de lenguaje. (Gavilán, 2009)

La descripción antes mencionada del término lenguaje documental, se formaliza con las definiciones dadas por los siguientes autores:

[C. Guinchat y Menou, 1984] en el contexto del uso de herramientas terminológicas para el manejo de información, *son lenguajes convencionales utilizados por una unidad de información para describir el contenido de los documentos para almacenarlos y recuperar las informaciones.*

[J. Rowley, 1987] en su libro Organizing knowledge, *“una lista de términos o notaciones que pueden ser utilizados como punto de acceso en un índice”, o “un conjunto de términos (el vocabulario) y las técnicas para utilizar las relaciones entre ellos en un sistema para dar descripciones de índice”.*

[Gil B., 1996] en el contexto de la gestión del conocimiento plantea, *todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios.*

1.4.1 Objetivos y Funciones de los Lenguajes Documentales.

El lenguaje documental es un lenguaje intermediario que se utiliza de puente entre las informaciones contenidas en los documentos y las informaciones solicitadas por los usuarios.

Sirve fundamentalmente para normalizar la indización, que es un proceso doble. Se indizan los documentos en la fase de entrada en el sistema y se indizan las preguntas de los usuarios en la etapa de salida o de recuperación. Los lenguajes documentales dotan de coherencia al proceso de indización. Los mismos tienen la capacidad de representar sin ambigüedad los contenidos documentales que le permiten cumplir dos objetivos fundamentales, el de inducción; y el de representación. (Mañá, 2006)

El primer objetivo se cumple cuando el usuario tiene un instrumento de consulta que lo guía a utilizar determinados términos para el concepto requerido, proporcionándole además otros términos que pueden también interesarle para su búsqueda.

³**Morfosintácticas:** Proviene del término morfosintaxis que se refiere al conjunto de elementos y reglas que permiten construir oraciones con sentido y carentes de ambigüedad mediante el marcaje de relaciones gramaticales, concordancias, indexaciones y estructura jerárquica de constituyentes sintácticos.

El segundo objetivo se cumple controlando la ambigüedad de los términos y reduciendo el volumen del vocabulario de la lengua natural. El procedimiento consiste en normalizar y seleccionar los sustantivos o sintagmas nominales.

1.4.2 Tipologías de los Lenguajes Documentales

Existen diferentes tipologías de lenguajes documentales: (Gavilán, 2009)

- ✓ Listas de términos.
- ✓ Tesoros.
- ✓ Ontologías.

Listas de términos.

Las listas de términos constituyen una expresión lingüística (palabra o conjunto de palabras) que representa el contenido temático de un documento. Se utilizan para hacer búsquedas en un catálogo, bibliografía o índice. Son lenguajes controlados pre-coordinados, formados por un conjunto de términos extraídos del lenguaje natural. Hacen uso preferentemente de las mismas bibliotecas y centros de documentación. Estas pueden ser adaptadas a cualquier dominio del conocimiento. (Gavilán, 2009)

Tesoros.

Un tesoro es un vocabulario controlado y estructurado formalmente, formado por términos que guardan entre sí relaciones semánticas y genéricas: de equivalencia, jerárquicas y asociativas. Se trata de un instrumento de control terminológico que permite convertir el lenguaje natural de los documentos en un lenguaje controlado, ya que representa, de manera unívoca, el contenido de estos, con el fin de servir tanto para la indización, como para la recuperación de los documentos. (Constructor de Tesoros CAT. Confección Automática de Tesoros., 2007)

Los tesoros pueden ser de diferentes tipos, entre los que cabe destacar los siguientes: (Morato, 2009)

- ✓ **Tesoro facetado:** se trata de un tesoro que combina una clasificación facetada sistemática, con un tesoro alfabético que reemplaza al índice de la clasificación.
- ✓ **Tesoro monolingüe:** es el que contiene descriptores en una única lengua.
- ✓ **Tesoro multilingüe:** contiene descriptores en más de una lengua.
- ✓ **Macrot tesoro:** se trata de un tesoro que incluye algunos descriptores básicos comunes y los propios de su campo o especialidad.

En cuanto a su estructura, se pueden establecer 3 tipos de tesoros: (Lapiente., 2011)

- ✓ **Tesoro lineal:** presenta los descriptores de forma simple, sin conexiones.
- ✓ **Tesoro en forma de árbol:** se construye siguiendo una jerarquía en forma de árbol. Cada descriptor posee un descriptor genérico y varios descriptores con los que se relaciona de forma de forma ascendente o descendente.
- ✓ **Tesoro reticular:** tesoro construido en forma de red en el cual los descriptores se entrecruzan. Cada descriptor puede tener varios descriptores genéricos y varios descriptores específicos.

En cuando a su presentación, un tesoro puede ser: (Lapiente., 2011)

- ✓ **Alfabético:** los descriptores y no descriptores se agrupan en una única secuencia alfabética junto con sus relaciones.
- ✓ **Sistemático:** estructurado en 2 parte. La primera de ellas o parte principal, contiene las categorías o jerarquías, y la segunda parte o parte auxiliar, consta de un índice alfabético que conduce a los usuarios a la sección semántica correspondiente a la que pertenece el término.
- ✓ **Gráfico:** los términos se presentan en forma de figura gráfica donde se asocian los términos relacionados. Esta representación gráfica suele representarse en forma de árbol, o utilizar flechas.

Ontologías.

Constituyen otro tipo de lenguaje documental además de los tesoros y las listas de términos. Es la expresión más alta de lenguaje documental que existe. El término proviene de la filosofía, una ontología es una teoría que trata la naturaleza y organización de la realidad, es decir, de lo que existe. El significado de la palabra es, conocimiento del ser (del griego onto: ser y logos: conocimiento). (Cruzata., 2011)

En el mundo jurídico las ontologías son denominadas: *una conceptualización del dominio jurídico que posibilita la construcción de prototipos de asistencia, búsqueda y organización documental de la información almacenada en los bancos de datos jurídicos.* (Valente A., 2005).

Generalizando una ontología es una especificación de una conceptualización, que posibilita la rápida búsqueda y recuperación de información en bases de datos documentales. Las mismas definen los términos y las relaciones básicas para la comprensión de un área del conocimiento.

1.5 Metodologías para el diseño de bases de datos.

Metodología es un vocablo generado a partir de tres palabras de origen griego: *meta* (“más allá”), *odòs* (“camino”) y *logos* (“estudio”). El concepto hace referencia al plan de investigación que permite cumplir ciertos objetivos en el marco de una ciencia. (Definición.DE, 2008-2012)

Se puede definir entonces una metodología como: un conjunto o grupo de procedimientos o bien técnicas y ayudas para el desarrollo de un software o de una BD. Traza las actividades a seguir en el desarrollo de principio a fin de la BD y qué es lo que debe realizar en cada actividad, señalando la entrada y el producto de salida, sin olvidar quién está involucrado.

1.5.1 Metodología de creación de bases de datos documentales.

Lluís Codina, es profesor titular de la Universidad Pompeu Fabra (UPF), en su tesis doctoral en 1994 desarrolló la “Metodología de análisis de sistemas de información y diseño de bases de datos documentales: aspectos lógicos y funcionales.” Posteriormente continuó perfeccionándola y la presentó por vez primera en un Congreso sobre Documentación organizado por el Departamento de Documentación de la Universidad de Zaragoza en el año 1996. Un año después, en 1997 la metodología salió nuevamente a la luz a través de la revista *Information World en Español*. A su vez, los fundamentos científicos de esta metodología fueron publicados en la *Revista Española de Documentación Científica* (1994b). Finalmente, en febrero de 1998 se presenta una versión revisada y actualizada que desde entonces, viene siendo testada y puesta a prueba en proyectos reales de diseño y creación de bases de datos documentales.

La metodología explica claramente que un primer punto de partida muy útil en el diseño de todo sistema de información y, por tanto, también en el diseño de una BDD, consiste en definir un **sistema de información** como algo que mantiene registros sobre otro conjunto del mundo real denominado **sistema objeto**.

El segundo punto de partida es considerar que, desde el punto de vista de los intereses de la documentación, todo sistema objeto se compone de dos subsistemas:

1. Sistema de actividades humanas (SAH).
2. Sistema de conocimiento (SCO).

El SAH es el sistema social, o sea formado por personas y cosas, que justifica la existencia del sistema de información, porque en él desarrollan sus actividades los futuros usuarios que necesitarán que exista un sistema de información. (Codina, 1998)

Por su parte, el SCO está formado por los documentos o las entidades sobre los cuales el sistema de información debe mantener algún tipo de registros. (Codina, 1998)

Ciclo de vida

El principio general de diseño de sistemas de información indica que el proceso comienza siempre por un diseño lógico o conceptual y que, una vez aprobado éste, se procede a su implantación, en un proceso que es más circular que lineal, ya que la implantación puede obligar a repensar aspectos del diseño.

El proceso de diseño de un sistema de información debe ajustarse siempre al siguiente ciclo de vida:

1. Análisis
2. Diseño
3. Implantación.

En el caso de una base de datos documental, estas fases quedarían divididas de la siguiente forma:

1. Análisis

1 a. Análisis del sistema de actividades humanas (SAH)

1 b. Análisis del sistema de conocimiento (SCO). (Se considera necesario aclarar que en las versiones de 1994 se definía con las siglas SIC)

2. Diseño

- Diseño del modelo conceptual
- Modelo entidad-relación.
- Diccionario de datos.

2b. Determinación de los procedimientos de tratamiento documental (descripción, análisis e indexación documental, etc.) si es el caso.

3. Implantación

Una vez aprobado el modelo conceptual de la base de datos, puede procederse a su implantación, la cual suele seguir el siguiente proceso:

1. Selección del sistema informático (software + hardware) que pueda satisfacer los requerimientos del modelo conceptual y del modelo de normativa de indexación. Primera instalación y nombramiento de un administrador de la base de datos que, a partir de ahora, será su máximo responsable.
2. Pruebas con una colección-test de documentos.
3. Cambios o ajustes necesarios.

Cada una de las fases precedentes (Análisis, Diseño, Implantación) tiene unos objetivos, debe producir unos resultados concretos y utilizar unas herramientas determinadas.

El valor de esta metodología radica, en que ayuda a que el producto final sea resultado del diseño consciente más que de las fuerzas ciegas del azar y/o del ensayo y error, aunque se entiende que su utilidad aumenta conforme se aplica a situaciones poco canónicas o a situaciones atípicas, como las que el entorno cambiante del desarrollo tecnológico introduce en cada momento, tal como el horizonte de las autopistas de la información y del mundo digital que existe actualmente. (Codina, 1998)

1.6 Herramientas de Modelado.

Las herramientas de modelado de procesos posibilitan la estandarización, mejoran la calidad y el rigor del diseño del proceso, son el punto de partida para el desarrollo de sistemas de software o sistemas de flujos de trabajo (workflow) y facilita la retroalimentación desde los usuarios finales.

Una herramienta, en términos informáticos, no es más que un sistema de software que intenta proporcionar ayuda automatizada a las actividades a realizar durante el proceso de desarrollo de software que permitirá aumentar la velocidad de desarrollo de los sistemas. En el presente estudio se realiza un análisis las herramientas: Rational Rose, Visual Paradigm y el Enterprise Architect que son algunas de las más usadas a nivel internacional.

1.6.1 Rational Rose.

Rational Rose Enterprise es el producto más completo de la familia Rational Rose. Todos los productos Rational Rose incluyen soporte UML. Posee soporte para análisis de patrones y un control por separado de componentes modelo que permite una administración más granular y el uso de modelos. Brinda soporte de ingeniería Forward y/o reversa para algunos de los conceptos más comunes de Java 1.5 y genera código Ada, ANSI C ++, C++, CORBA, Java y Visual Basic, con capacidad de sincronización modelo- código configurables. (Innova, 2008)

Esta herramienta propone la utilización de cuatro tipos de modelo para realizar un diseño del sistema, utilizando una vista estática y otra dinámica de los modelos del sistema, uno lógico y otro físico. Permite crear y refinar estas vistas creando de esta forma un modelo completo que representa el dominio del problema y el sistema de software.

Dentro de sus principales características se encuentran las siguientes: admite Especificar, Analizar, Diseñar el sistema antes de Codificarlo. Mantiene la consistencia de los modelos del sistema software.

Chequeo de la sintaxis UML. Generación de documentación automáticamente. Generación de código a partir de los modelos. Ingeniería Inversa. (Tobalina)

1.6.2 Enterprise Architect.

El Enterprise Architect es un ambiente de modelado para trabajo en equipo, potenciado por UML 2.3, para la realización de un diseño gráfico que ayude a que el equipo de desarrollo construya y mantenga un robusto sistema. Enterprise Architect tiene incorporada la capacidad de manejar de forma clara los requerimientos en un sistema o negocio, esto ayuda al arquitecto de empresa a desarrollar especificaciones con un alto nivel de análisis, diseño, implementación, prueba y mantenimiento usando UML, SysML (conocido por sus siglas en inglés Systems Modeling Language, que se traduce al español como Lenguaje de Modelado de Sistema), BPMN y otras normas de modelado. Es una herramienta que abarca el ciclo de vida completo del desarrollo de software, con herramientas que pueden proveerle una estructura competitiva en modelado de negocio, diseño de software, ingeniería de sistemas, arquitectura de empresas, gestión de requisitos, pruebas y mucho más. (Sparx Systems, 2010)

Es una herramienta multiusuario (Multiusuario: Propiedad que permite proveer servicio y procesamiento a múltiples usuarios simultáneamente), con seguridad y administración de permisos incorporada. Soporta diferentes repositorios basados en DBMS (Sistemas Manejadores de Base de Datos), incluyendo Oracle, SQL Server, My SQL, PostgreSQL. Provee trazabilidad completa desde el análisis de requerimientos hasta los artefactos de análisis y diseño, a través de la implementación y el despliegue (System, 2008)

1.6.2 Visual Paradigm.

Visual Paradigm es una herramienta de lenguaje de modelado unificado (UML por sus siglas en inglés Unified Modeling Language) profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. En la actualidad está consolidado como el lenguaje estándar en el análisis y diseño de sistemas de cómputo. El software de modelado UML ayuda a una construcción de aplicaciones de calidad rápida, mejor y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. (Danysoft, 2006)

Presenta grandes ventajas y funcionalidades, dentro de las que se pueden citar las siguientes:

- ✓ Soporte para UML.

- ✓ Generador de informes para generación de documentación.
- ✓ Distribución automática de diagramas - Reorganización de las figuras y conectores de los diagramas UML.
- ✓ Editor de figuras.
- ✓ Disponibilidad en múltiples plataformas (Windows, Linux, etc.)
- ✓ A partir del Diagrama Entidad Relación (DER) permite la sincronización con el diagrama de clases y viceversa.
- ✓ A partir del diagrama de clases sincronizado con el DER permite generar la BD.

Por las razones antes expuestas se selecciona Visual Paradigm como herramienta de modelado para la realización de este trabajo. Es importante destacar que otra de las principales razones por la que se escoge Visual Paradigm como herramienta de modelado es por ser una herramienta CASE de software propietario que brinda la venta de licencias universitarias a un costo menor y la Universidad de las Ciencias Informáticas posee la licencia para utilizar la misma.

1.7 Patrones en el desarrollo de Base de Datos.

Los patrones constituyen una solución estándar para un problema común, en este caso para el diseño de Base de datos (BD). El diseño y construcción de una BD requiere del mayor esfuerzo y análisis posible, ya que a partir de este diseño es que esta se crea y la calidad con que se obtenga determinará su comportamiento futuro. En estos tiempos las BD suelen ser muy grandes y a veces el trabajo con los patrones de diseño hacen que el diseño sea más fácil y se asegure un resultado satisfactorio. (Blaha, 2010)

A continuación se describen algunos de los patrones existentes y su notación y condiciones en que deben ser usados cada uno.

Árboles.

Un árbol es un término de la teoría de grafos. Es un conjunto de nodos conectados en la estructura de hijo a padre y que se asemeja a la forma de un árbol. Los árboles son sumamente utilizados en los diseños actuales, por lo que existen varios patrones de diseño de BD asociados a los mismos: (Blaha, 2010)

Patrón 1. Árboles fuertemente codificados (Hardcoded tree): En este caso a cada nivel del árbol se le asocia una entidad. Normalmente constituyen relaciones de 1 a muchos (n). Usualmente utilizado para

representar jerarquías donde es bien conocida la estructura y es importante representar la correspondencia.

Patrón 2. Árboles simples: Patrón normalmente utilizado cuando el árbol es la representación de una estructura de datos. Los elementos a almacenar son del mismo tipo, es decir, pueden ser almacenados en la misma entidad. No pueden existir ciclos.

Patrón 3. Árbol estructurado: Este modelo es usado cuando se necesitan diferenciar los nodos hojas (leaf), de aquellos que generan una nueva rama (branch), porque ambos tipos de nodos tienen diferentes atributos, relaciones y/o semántica. No pueden existir ciclos. La generalización tiene cubrimiento total y exclusivo, cada elemento de la entidad Node, debe tener su correspondiente elemento en la entidad Leaf o en la entidad Branch.

Grafos.

Los grafos dirigidos son otro elemento de la teoría de grafos. Un grafo dirigido es un conjunto de nodos y un conjunto de caminos dirigidos entre los nodos. Cada camino puede representar un cambio de estado o una conexión entre nodos donde cada uno representa un rol determinado, ya sea el origen o el destino o ambos (en el caso de los ciclos o lazos). Desde un mismo nodo pueden salir un sinnúmero de caminos. (Blaha, 2010)

Patrón 4. Grafo dirigido simple: Se utiliza cuando todos los nodos contienen el mismo tipo de datos. Es similar al árbol simple, la diferencia es que en este caso la relación recursiva sobre Node tiene cardinalidad de muchos a muchos y por tanto se genera una nueva entidad.

Patrón 5. Grafo dirigido estructurado: Este modelo es usado cuando se necesita diferenciar los nodos hojas (leaf), de aquellos que generan una nueva rama (branch), porque ambos tipos de nodos tienen diferentes atributos, relaciones y/o semántica. Es similar al árbol estructurado, la diferencia es que en este caso la relación recursiva sobre Node tiene cardinalidad de muchos a muchos.

Patrón 6. Patrón de llaves subrogadas.

Este patrón es muy utilizado pues facilita la interacción con la BD en un futuro. El mismo plantea que se genere una llave primara única para cada entidad, en vez de usar un atributo identificador en el contexto dado.

Normalmente se usan enteros en columnas identity o GUID⁴ que están demostradas que no se repiten o con una probabilidad extremadamente baja. Esto permite que las tablas sean más fáciles de

⁴ **GUID:** Globally Unique Identifier o identificador único global.

consultar a partir del identificador, pues todos tienen el mismo tipo en cada una de las tablas. (Blaha, 2010)

Patrones para flujos de trabajo.

En la vida práctica a diario surge la necesidad de representar y hacer persistente un flujo de trabajo. Si bien UML tiene su notación para la representación de un flujo de trabajo, en el modelado de la base de datos es recurrente el tema de cómo representar esto. Existen dos patrones definidos para diseñar un flujo de trabajo: máquinas de estado para un tipo de entidad y máquinas de estado para escenarios. (Blaha, 2010)

Patrón 7. Máquinas de estado para un tipo de entidad.

El patrón máquina de estado para un tipo de entidad representa los posibles cambios de estado por los que puede atravesar un tipo de entidad.

Este modelo es descriptivo, por lo que no refleja la ocurrencia, para eso es necesario incluir un conjunto de valores adicionales como el tiempo y el escenario donde ocurren las transiciones. (Blaha, 2010)

Patrón 8. Máquina de estado para escenarios (control de flujo).

Este modelo representa la ocurrencia del cambio de estado en un escenario de una entidad dada, por lo tanto considera el tiempo y la persistencia del mismo en las tablas resultantes. También representa la ocurrencia de un estímulo en una fecha y los estados por los que ha pasado, caracterizados por la fecha de inicio y la fecha fin.

La diferencia con respecto al patrón anterior es que este sí permite guardar la información específica de un flujo en un momento determinado.

Es común encontrar ambos modelos juntos y en muy rara ocasión la configuración o diseño de una máquina de estado para una entidad está separado de las ocurrencias. (Blaha, 2010)

Patrón 9. Modelo Entidad-Atributo-Valor.

El patrón entidad-atributo-valor es la representación de un modelo flexible donde se pueden representar objetos con sus atributos. Es un acercamiento al modelo orientado a objeto representado en el modelo relacional, donde la entidad Class representa las clases, la entidad Attribute representa los atributos de las clases, por su parte la entidad Object representa las instancias de las clases, mientras que la entidad Value representa los valores de cada atributo para cada objeto dado.

Un ejemplo del modelo entidad-atributo-valor se puede presentar en el almacenamiento de documentos captados a través de un sistema, que como pueden ser disímiles y surgen nuevos a cada momento, deben almacenarse en una estructura flexible. (Blaha, 2010)

Patrón 10. Patrón una tabla por dominio: Este patrón plantea el uso de una tabla nomencladora para cada dominio lo que trae como ventaja que es fácil de entender cuando se escriben consultas, posee un mejor rendimiento en las consultas, dada la menor cardinalidad de los datos asociados y además permite una fácil implementación de interfaces de usuario dado un mapeo casi exacto de las tablas. (Blaha, 2010)

1.8 Gestores de Base de Datos.

Un Sistema Gestor de base de datos (SGBD) es un conjunto de programas que permiten crear y mantener una base de datos, asegurando su integridad, confidencialidad y seguridad, es decir cumpliendo con sus normas de seguridad informática. Por tanto debe permitir: (Date, 2001)

- ✓ Definir una base de datos: especificar tipos, estructuras y restricciones de datos.
- ✓ Construir la base de datos: guardar los datos en algún medio controlado por el mismo SGBD
- ✓ Manipular la base de datos: realizar consultas, actualizarla, generar informes.

Características deseables en un SGBD:

- ✓ *Control de la redundancia:* la redundancia de datos tiene varios efectos negativos (duplicar el trabajo al actualizar, desperdicia espacio en disco, puede provocar inconsistencia de datos) aunque a veces es deseable por cuestiones de rendimiento.
- ✓ *Restricción de los accesos no autorizados:* cada usuario ha de tener unos permisos de acceso y autorización.
- ✓ *Cumplimiento de las restricciones de integridad:* el SGBD ha de ofrecer recursos para definir y garantizar el cumplimiento de las restricciones de integridad. (Collector, 2004)

Existen tres grandes agrupaciones de SGBD: Los considerados productos libres, constituyendo los más relevantes PostgreSQL y MySQL. La segunda agrupación son el conjunto de los gestores no libres donde resaltan como principales ORACLE y Microsoft SQLServer. La tercera agrupación y más pequeña, contiene los gestores considerados productos no libres y que poseen comunidades de desarrollo gratuitas como: Microsoft SQLServer Compact Edition. (Oracle, 2011)

Oracle.

Es un SGBD relacional, desarrollado por Oracle Corporation surge a finales de los años 70 bajo el nombre de Relational Software a partir de un estudio sobre SGBD de George Koch. Es básicamente una herramienta cliente/servidor para la gestión de bases de datos. Es un producto vendido a nivel mundial. Presenta las siguientes características: es una herramienta de administración gráfica muy intuitiva y cómoda de utilizar. Ayuda a analizar datos, efectuar recomendaciones concernientes a mejorar el rendimiento y la eficiencia en el manejo de los datos almacenados. Apoya el diseño y optimización de modelos de datos. Asiste a los desarrolladores con sus conocimientos de SQL, de construcción de procedimientos almacenados y triggers. Apoya la definición de estándares de diseño y nomenclatura de objetos. Documenta y mantiene un registro periódico de las actualizaciones de hardware y software, cambios en las aplicaciones y, en general, todos aquellos eventos relacionados con cambios en el entorno de utilización de una base de datos. (Oracle, 2011)

Microsoft SQL Server.

Es un SGBD relacional basado en el lenguaje de consulta estructurado⁵, capaz de poner a disposición de muchos usuarios grandes cantidades de datos de manera simultánea. El mismo se encuentra entre los software que están bajo licenciamiento propietario. Dentro de sus principales características se pueden citar: (Server, 2012)

- ✓ Soporte de transacciones.
- ✓ Gran estabilidad, seguridad, escalabilidad.
- ✓ Soporta procedimientos almacenados.
- ✓ Incluye también un potente entorno gráfico de administración que permite el uso de comandos DDL⁶ y DML⁷ gráficamente.
- ✓ Permite trabajar en el modo de cliente – servidor donde la información y datos se alojan en el servidor y las terminales o clientes de la red solo accedan a la información.
- ✓ Permite administrar información de otros servidores de datos.

MySQL.

Es un SGBD relacional, multihilo y multiusuario, con más de seis millones de instalaciones. Es desde enero de 2008 una subsidiaria de Sun Microsystems y esta a su vez de Oracle Corporation desde abril

⁵ **Lenguaje de consulta estructurado:** Structured Query Language. SQL por sus siglas en inglés.

⁶ **DDL:** Data Definition Language.

⁷ **DML:** Data Manipulation Language.

de 2009 que desarrolla MySQL como software libre en un esquema de licenciamiento dual. Por un lado se ofrece bajo la GNU GPL para cualquier uso compatible con esta licencia, pero para aquellas empresas que quieran incorporarlo en productos privativos deben comprar a la empresa una licencia específica que les permita este uso. Entre sus principales características se encuentran: gran rapidez y facilidad de uso. Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo. Soporta gran cantidad de tipos de datos para las columnas. Gran portabilidad entre sistemas. Soporta hasta 32 índices por tabla. Gestión de usuarios y passwords, manteniendo un muy buen nivel de seguridad en los datos. Fácil instalación y configuración. (MySQL, 2011)

PostgreSQL.

PostgreSQL es un sistema gestor de base de dato objeto-relacional, distribuido bajo licencia BSD⁸ y con su código fuente disponible libremente. PostgreSQL está ampliamente considerado como el sistema de bases de datos de código abierto más avanzado del mundo. Posee muchas características que tradicionalmente sólo se podían ver en productos comerciales de alto calibre. (González, 2012).

Su licencia de tipo BSD, permite manejar libremente el código fuente, mejorando u optimizando su código. Incluso se permite redistribuirlo como producto comercial y combinarlo con herramientas de licencia propietaria. Es multiplataforma, pues puede operar sobre distintas plataformas incluyendo Linux, Unix, MacOSX, Solaris y Windows. Tiene una buena escalabilidad ya que es capaz de ajustarse al número de CPUs y a la cantidad de memoria que posee el sistema de forma óptima, soportando una mayor cantidad de peticiones simultáneas a la base de datos de manera correcta; se dice que “PostgreSQL ha llegado a soportar el doble de carga de lo que soporta MySQL”

La última serie de producción es la 9.1. Sus características técnicas la hacen una de las bases de datos más robustas y potentes del mercado. Su desarrollo comenzó hace más de 16 años, y durante este tiempo, *estabilidad, potencia, robustez, facilidad de administración e implementación de estándares* han sido las características que más se han tenido en cuenta durante su desarrollo. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema. (Simon Riggs, 2010)

⁸ **BSD:** Licencia de Distribución de Software de Berkeley (BSD por sus siglas en inglés: Berkeley Software Distribution). No impone ninguna restricción a los desarrolladores de software en lo referente a la utilización de código en derivados y licencias de estos programas.

La versión 9.1 ofrece muchas características que los usuarios han estado solicitando por años, retirando obstáculos para el despliegue de aplicaciones nuevas o migradas en PostgreSQL. Estas incluyen: (Group, 2012)

- ✓ **Replicación Sincrónica:** permitiendo alta disponibilidad con consistencia sobre múltiples servidores.
- ✓ **Regionalización por columna:** soportando correctamente el ordenamiento por lenguaje en las bases de datos, tablas o columnas.
- ✓ **Tablas unlogged:** importante incremento del rendimiento para datos efímeros.
- ✓ **Indexamiento de los K vecinos más próximos (K-Nearest-Neighbor):** índices sobre "distancia" para consultas rápidas de ubicación y búsquedas de texto.
- ✓ **Nivel de Aislamiento Serializable a través de "Snapshots":** mantiene consistentes múltiples transacciones concurrentes sin el uso de bloqueos, usando "verdadera serialización"
- ✓ **Writeable Common Table Expressions:** ejecuta actualizaciones multi-fases complejas en una simple consulta
- ✓ **Security-Enhanced Postgres:** despliega seguridad de nivel militar y control de acceso mandatorio.

La extensibilidad de PostgreSQL permite a los usuarios añadir nuevas funcionalidades en bases de datos que corren en producción, y usar estas para tareas que no puede realizar ningún otro sistema de base de datos. La versión 9.1 añade nuevas herramientas de extensibilidad, incluyendo: (Group, 2012)

- ✓ **Conectores de datos foráneos:** añada y consulte fuentes de datos externas desde PostgreSQL.
- ✓ **Extensiones:** cree, cargue, y administre fácilmente nuevas características de la base de datos.

PostgreSQL es el líder en sistemas de bases de datos de código abierto, con una comunidad mundial de miles de usuarios y contribuyentes y docenas de empresas y organizaciones. El proyecto PostgreSQL se basa en 25 años de ingeniería, a partir de la Universidad de California, Berkeley, y tiene un ritmo sin precedentes de desarrollo en la actualidad. El conjunto de características maduras que tiene PostgreSQL no sólo rivaliza con sistemas de bases de datos propietarias, sino que los supera en características avanzadas, extensibilidad, seguridad y estabilidad. (Group, 2012)

1.8.1 Herramientas de Administración de base de datos.

Las herramientas de administración de datos son las encargadas de administrar una base de datos determinada dentro de un gestor ya instalado, en este caso el Postgres SQL. Estas herramientas

permiten realizar múltiples tareas pues facilitan la interacción directa del programador con la base de datos.

Navicat.

Este potente administrador de bases de datos relacionales incluye un amplio abanico de herramientas para gestionar, crear, y sincronizar bases de datos ya sea servidores locales o remotos. Además integra una amplia gama de características y funcionalidades que facilitan ampliamente toda la gestión de BD en Mysql, Postgres, Oracle, por solo mencionar algunos. Facilita la migración de ficheros y posee una interfaz simple, de fácil adaptación para los nuevos usuarios. (PremiumSoft, 1999-2012)

EMS⁹ Postgres SQL Manager.

Es una útil aplicación que soporta todas las características y gestiona la información que recopila de la planilla, tomando todos o determinados campos sincronizando todos los servidores compatibles con esta base de datos permitiendo visualizarlo desde cualquier PC. Sus principales características son el espacio entre tablas, la posibilidad de renombrar argumentos, y una fácil administración de todos los objetos Postgres así como una herramienta para la interpretación de las consultas entre otras. Su interfaz es muy sencilla aunque poderosa y optimiza ágilmente la navegación entre los datos. (SQLManager.net, 1999-2012)

PGAdmin III.

Es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia Open Source. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de la PostgreSQL 7.3 ejecutándose en cualquier plataforma, así como versiones comerciales de PostgreSQL como Pervasive Postgres, EnterpriseDB, Mammoth Replicator y SRA PowerGres. (Ubuntu, 2008)

Está diseñado para responder a las necesidades de los usuarios, desde escribir consultas simples hasta el desarrollo de las bases de datos más complejas. El interfaz grafico soporta todas las funciones de Postgres y facilita enormemente la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor, un agente para lanzar scripts programados, soporte para el motor de replicación Slony-I y mucho más. La conexión al

⁹EMS: Enterprise Manager System.

servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets, y puede ser encriptado mediante SSL¹⁰ para mayor seguridad.

Debido a que Navicat, el Microsoft SQL Server y EMS Postgres SQL Manager son privativas y que la política informática tanto de la Universidad como del país es enfocada hacia el software libre, para la administración de Postgres se utilizará la herramienta PGAdmin III.

1.8.2 Herramientas para la validación de BD.

Data Generator 2005 for PostgreSQL.

Data Generator para PostgreSQL es un poderoso utilitario para generar datos de prueba a tablas de bases de datos PostgreSQL de una vez. La aplicación asistente permite definir tablas y campos para generar datos, configurar valores de rangos, generar campos de tipo carácter por máscara, cargar valores desde archivos para campos de tipo objetos binarios (BLOB), obtener listas de valores desde consultas SQL y otras características para generar datos de prueba de manera simple y directa. También provee una aplicación consola, que permite generar datos de un "clic" usando plantillas de generación.

Posee una amigable interfaz asistente de usuario. Puede generar datos a varias tablas desde diferentes bases de datos en un solo ordenador anfitrión (host). Brinda soporte para todos los tipos de datos de PostgreSQL. Posee diferentes tipos de generación por cada campo, incluyendo lista, azar (random), generación incremental de datos y más. Tiene capacidad para usar resultados de consultas SQL como lista de valores para la generación de datos. Además de un control automático sobre integridad referencial para la generación de datos a tablas vinculadas, amplia variedad de generación de parámetros para cada tipo de campo, capacidad para configurar valores Null (nulos) para ciertos casos, posibilidad de salvar todos los parámetros de generación, para comenzar la sesión del asistente actual y es utilitario de líneas de comandos para generar datos usando el archivo que contiene la plantilla. (Manager, 2012)

Apache JMeter 2.3.

Apache JMeter es software de código abierto, un 100% Java, pura aplicación de escritorio diseñada para carga funcional, ensayo de comportamiento y medir los resultados. Originalmente se diseñó para probar aplicaciones web, pero desde entonces se ha expandido a otras funciones de prueba. Puede ser utilizado para probar el rendimiento tanto en estática y dinámica de los recursos (archivos, scripts

¹⁰ **SSL:** son las siglas en inglés de Secure Socket Layer (capa de conexión segura).

de Perl, Objetos Java, bases de datos y consultas, servidores FTP y mucho más). Para simular una carga pesada en un servidor, la red o un objeto para poner a prueba su resistencia o para analizar el rendimiento global por diferentes tipos de carga. Para hacer un análisis gráfico de rendimiento o para probar el servidor, script y comportamiento del objeto con cargas pesadas concurrentes. (Foundation 2009)

El Apache JMeter incluye una interfaz gráfica de usuario que facilita el diseño de las pruebas. Esta interfaz además de aportar un cómodo entorno de trabajo, también permite guardar y alterar tanto los test desarrollados como los componentes que lo integran. Gracias a esto se pueden reutilizar las pruebas o módulos de las mismas en el desarrollo de nuevas pruebas.

1.8.3 Sistemas Gestores de Base de Datos Documentales (SGBDD).

Un Sistema de Gestión de Bases de Datos Documentales (SGBDD) es un software que dispone de todas las funciones necesarias para el almacenamiento, indexación, consulta y modificación de la información textual en forma estructurada, variada y de gran longitud, que aplica de manera limitada las técnicas de la informática documental y posee capacidad para: recuperar con rapidez, en base a la generación de índices, registros que responden a un criterio de búsqueda; realizar búsquedas multicriterio sobre ficheros inversos utilizando la lógica booleana y finalmente su capacidad para administrar tesauros y diccionarios terminológicos. A diferencia de cualquier otro SGBD, un SGBDD no realiza ningún tratamiento sobre la información. Simplemente la almacena y posibilita su recuperación. (Yunta, 2001)

Existen varios SGBDD, estos están en constante perfeccionamiento, a continuación se mencionan algunos, sus características y factibilidad de uso.

ProCite®.

Es un gestor de bases de datos documental producido por ISI ResearchSoft! que se adapta al trabajo más exigente del investigador y del documentalista. El programa dispone de una amplia variedad de formatos de entrada y salida, esta característica, junto a la facilidad de uso, hace del mismo uno de los programas de gestión de bases de datos más populares.

Este gestor posee la posibilidad de utilizar y definir cientos de formatos diferentes: Monografías, artículos de revista, fotografías, mapas, cartas, manuscritos, películas, patentes, informes y demás formatos.

ProCite posibilita trabajar con licencia multiusuario en una red distribuida, lo que posibilita que varios operarios trabajen de manera simultánea con la misma base de datos.

Permite importar datos de programas similares como EndNote o Reference Manager, para ello lleva el programa BiblioLink II. Éste es un programa que permite crear o modificar plantillas para la importación o exportación de archivos de otras bases de datos. Las bases de datos generadas con ProCite llevan la extensión PDT y los índices de las mismas PDX. (ProCite, 2005)

Knosys Blue.

Knosys Blue es una potente y veloz herramienta que permite organizar, gestionar, publicar y compartir información digital. Las bases de datos de *Knosys Blue* permiten importar información de documentos Word y PDF entre otros, imágenes, correos electrónicos y mucho más. Además *Knosys Blue* ofrece la posibilidad de enlazar unas bases de datos con otras y lanzar consultas sobre múltiples bases de datos. Asimismo, el nuevo interfaz de la herramienta permite un acceso inmediato a las funciones adecuadas en cada momento gracias a su cualidad contextual. (S.A., 2011)

En noviembre del 2011 Micronet, empresa desarrolladora, editora y distribuidora de software de consumo para PC y consolas, anunció el lanzamiento las dos primeras ediciones de su renovada solución de Enterprise Content Management:

Knosys Blue Home Edition es un gestor de bases de datos documentales que permite organizar, consultar y compartir la información de bases de datos de gastos, facturas, contratos, escrituras, películas, fotografías, vídeos, música, libros y cualquier archivo de documentos. Además ofrece la posibilidad de importar la información de otros formatos de archivo y trabaja con los tipos de archivo más frecuentes, desde Microsoft Office a HTML, PDF y correo electrónico.

Knosys Blue Standalone Professional es un gestor de bases de datos documentales con características muy similares a su hermano de edición que además ofrece la posibilidad de importar la información desde otros formatos de archivo.

Greenstone

Greenstone es un conjunto de programas y aplicaciones de software especialmente diseñados para la creación y difusión de colecciones documentales digitales, el cual le ofrece una nueva forma de organizar la información y publicarla en Internet o CD-ROM. Greenstone ha sido elaborado como parte del proyecto de Biblioteca Digital de Nueva Zelanda de la Universidad de Waikato y actualmente es desarrollado y distribuido en colaboración con la UNESCO y la ONG Human Info. Es un software de

código abierto disponible en <http://greenstone.org> bajo los términos y condiciones de la Licencia Pública General de GNU.

Las colecciones Greenstone pueden contener texto, imágenes, sonidos y videoclips. La mayor parte del material no textual está enlazado a los documentos textuales o va acompañado de descripciones textuales para permitir la búsqueda y consulta en texto completo. Sin embargo, la arquitectura del programa permite la utilización de plugins y clasificadores aún para datos no textuales. Funciona con los sistemas Windows (3.1/3.11, 95/98/Me, NT/2000) y Unix (Linux y SunOS). Cualquiera de estos sistemas puede utilizarse como servidor Web. (Ian H. Witten, 2010)

INMAGIC DB/TextWorks®

Es un programa propietario que combina la creación de bases de datos y la facilidad de búsqueda y recuperación textual. Permite trabajar con bases de datos en un entorno de red o bien en un entorno aislado. DB/TextWorks® gestiona diversos tipos de información y de materiales, como por ejemplo, documentos textuales, imágenes, vídeos u otros.

Proporciona a su organización las herramientas necesarias para gestionar su documentación, de forma que puede convertirse en la herramienta clave para gestionar su repositorio de conocimiento. DB/TextWorks® combina un gestor de bases de datos documentales y un sistema ágil y rápido de búsqueda y recuperación textual. Esta combinación proporciona un sistema potente y flexible y permite compartir el conocimiento de toda su organización, lo que facilita que se incremente la eficiencia y eficacia de su organización. (DB/TextWorks, 2010)

CDS/ISIS

Programa creado por la UNESCO en 1985 con el fin de promover la normalización y el intercambio de Información bibliográfica entre organismos. El sistema CDS/ISIS (Computerized Documentation System - Integrated Set for Information System) permite construir y administrar bases de datos estructuradas no numéricas, es decir, bases de datos constituidas principalmente por textos.

Aunque CDS/ISIS maneja con textos y palabras, y por lo tanto ofrece muchas de las características encontradas comúnmente en un procesador de palabras, hace mucho más que un simple proceso de textos. Esto se debe a que los textos que procesa CDS/ISIS están estructurados en datos elementales que define el usuario. ((CSIC), 2012)

Las características específicas más notorias que convierten al MicroISIS en un software indicado para bases de datos documentales son:

- ✓ Puede contener un máximo de 16 millones de registros de hasta 8.000 caracteres desagregables en un máximo de 200 campos diferentes, a su vez, divisibles en subcampos.
- ✓ Campos de longitud variable. No es necesario predefinir una longitud del campo, ésta es variable de registro en registro, pudiendo ser nula o extenderse hasta completar la longitud máxima del registro.
- ✓ Potentes recursos de recuperación (búsqueda) de registros seleccionados de acuerdo a su contenido en determinados campos con uso de operadores booleanos y otras expresiones condicionales. Las búsquedas se pueden efectuar por el contenido completo de un campo, por palabras aisladas o secuencia de caracteres.
- ✓ Amplias posibilidades de diseño de formularios de ingreso de datos y de formatos de salida impresos, ambos adaptados a información de tipo textual. (atomica, 2012)

Después de analizar detenidamente los SGBDD, sus principales características y su uso a nivel mundial, se puede decir que son sistemas integrados, capaces de crear y administrar BDD, que poseen un grupo de funcionalidades definidas y que en su mayoría son privativos.

La BDD que se desea implementar podrá ser utilizada en el desarrollo de sistemas de información tales como los Sistemas de Información Basados en Ontologías (SIBO). Los cuales permiten una eficiente gestión del conocimiento jurídico.

En el trabajo de diploma *“Procedimiento para el diseño de un Sistema de Información Basado en Ontologías Jurídicas”* de los autores Yordanis García Leiva y Héctor L. Duran Cruzata (Cruzata., 2011), se define que uno de los principales componentes dentro de un SIBO lo constituye la BDD, la cual permitirá recoger toda la información referente a los documentos jurídicos que se deseen analizar y establecer una relación con otra base de datos encargada de almacenar una ontología jurídica con el objetivo de brindarle un análisis semántico a la información procesada.

Teniendo en cuenta los posibles usos de la BDD a desarrollar en el presente trabajo, se decide utilizar PostgreSQL como SGBD. Pues los SGBDD no permiten la creación de una BD que pueda ser integrada a otros sistemas, debido a que ellos crean su propio sistema integrado para la gestión de los documentos que almacenan. Con la utilización del PostgreSQL se puede obtener una BD con las características de las bases de dato documental y fácil de adaptar a sistemas como el resultante de la implementación del procedimiento propuesto en la tesis antes mencionada. Además la decisión de utilizar este gestor también se hace por el motivo de ser una herramienta de código abierto a diferencia

de los SGBDD que en su mayoría son de licencias privativas, lo cual atenta contra las políticas de soberanía tecnológicas implementadas en el país.

1.9 Índices en BD

Un índice es esencialmente un subconjunto ordenado de las columnas de una tabla, con una entrada que apunta a la tupla correspondiente. Trabajando con un conjunto indexado se puede realizar un procesamiento más rápido para cada solicitud de consulta, ya que se elimina la necesidad de buscar en toda la tabla, optando en su lugar por concentrarse en sólo una porción relativamente pequeña de esta. (Group, 2012)

Un índice se define sobre las columnas de las tablas que se utilicen repetidamente en consultas, sobre aquellos campos por los cuales se realizan operaciones de búsqueda con mayor frecuencia. Una tabla puede ser indexada por campos de tipo numérico, de tipo texto o de tipo varchar, también se puede indexar por un campo que contenga valores NULL, excepto las llaves primarias. Su uso puede mejorar el rendimiento de las base de datos, permitiendo recuperar las filas específicas mucho más rápido de lo que podía hacerse sin él.

Balanceo de la BD

Balancear el costo de mantenimiento de los índices: para esto se considera el impacto de cada uno de los índices candidatos en las actualizaciones de la carga de trabajo. Es necesario valorar si mantener un índice ralentiza las operaciones de actualización frecuente, caso en el que es necesario recrearlo; o bien si el índice puede acelerar una actualización dada. (Group, 1996-2012)

Tipos de índices.

BTREE: Gestiona las consultas usando operadores de igualdad o de rango de datos. Su estructura tiene la forma de un árbol invertido donde las estructuras superiores son llamadas ramas y las inferiores hojas. Generalmente los índices en BTREE tienen uno o más niveles de ramas donde son estas ramas las que contienen la columna índice (claves) y la dirección de otro bloque y los nodos hojas poseen la clave de cada fila de la columna. Estos índices se utilizan además para evitar las grandes operaciones de ordenación y obtienen su mejor resultado cuando son aplicados sobre columnas de alta cardinalidad, o sea, que contengan muchos valores diferentes.

HASH: Este tipo de índices solo pueden manejar simples comparaciones de igualdad. El planificador de consulta considera su uso cada vez que una columna indexada está involucrada en una comparación con el operador igual. No se admiten búsquedas usando el operador IS NULL.

GIST: Estos no son solo un tipo de índice sino una infraestructura sobre las que se pueden implementar muchas estrategias de indexación, los operadores usados varían según estas estrategias. GIST es sinónimo de búsqueda generalizada sobre un árbol. Una de sus ventajas es que permite desarrollar tipos de datos personalizados con sus adecuados métodos de acceso. Gist sustituyó los antiguos r-trees, es usado para tipos complicados como tipos geométricos, búsquedas en texto, por solo mencionar algunos.

GIN: Este índice además conocido como índice invertido puede manejar valores que contienen más de una clave como los arreglos. Al igual que GIST puede apoyar diferentes estrategias de indexación que son definidas por el usuario. Es usado para buscar en texto, tarea en la cual llega a ser más rápido que Gist pero más lento para añadir nuevos valores, por tanto se opta por este método para las búsquedas en texto. (Group, 1996-2012)

1.10 Conclusiones parciales.

Al finalizar el capítulo se concluye lo siguiente:

- A través del estudio del estado del arte realizado, se logró conocer las características de las BDD, la metodología para la creación de estas y las herramientas que se utilizarán para su desarrollo.
- La fundamentación teórica desarrollada en este capítulo permitió sentar las bases necesarias para dar cumplimiento a los objetivos propuestos.

Capítulo 2: Diseño e implementación de la Base de Datos.

2.1 Introducción.

Tomando como base la investigación realizada para confeccionar este y el anterior capítulo se diseñará una BDD según plantea la metodología del Dr. en ciencias de la información Lluís Codina. Por lo que se define el SAH, el SCO y los requisitos correspondientes a la investigación. Así como la definición del Modelo conceptual, el Esquema Relacional y el diccionario de datos definido para la implementación de la base de datos donde se almacenarán las sentencias penales del CENDIJ.

2.2 Sistema de Actividades Humanas en el CENDIJ.

Teniendo en cuenta la definición de SAH explicada en el capítulo anterior, el SAH vendría siendo el CENDIJ, pues si se considera la base de datos que automatiza el centro de documentación del mismo, este centro de documentación es el SAH respecto a la base de datos.

Actividades.

Como el entorno de un sistema siempre influye en él de alguna forma, los diseñadores de la base de datos, aunque deberán concentrarse en modelar las características del centro de documentación, también deberán conocer las características de su entorno, o sea del CENDIJ. Esto se logra después de la investigación realizada en el capítulo 1, a partir de la cual se definen las siguientes actividades que conforman el SAH:

1. **Recepción de la sentencia:** es el proceso de recepción de todas las sentencias que llegan al CENDIJ. Estas sentencias llegan al centro de forma digital y en formato duro, o sea en papel.
2. **Procesamiento de la sentencia:** una vez que las sentencias son recepcionadas, comienza un proceso de ordenamiento, catalogación, clasificación y almacenamiento de las mismas.
3. **Almacenamiento de la sentencia:** el proceso de almacenar las sentencias, es un actividad compleja, que se divide en 2 actividades fundamentales:
 - a. Las sentencias digitales, una vez procesadas, son insertadas en el sistema actual, de forma manual.
 - b. Las sentencias en formato duro, son almacenadas en estantes ubicados en habitaciones específicas, son el objetivo de mantener constancia y base de resguardo a la información digitalizada. Esta tarea se desarrolla almacenando las sentencias por criterios propios de la entidad, según su clasificación y fecha.

4. **Búsquedas de sentencias:** esta actividad se desarrolla principalmente cuando se presta alguno de los servicios que presta el CENDIJ. Estas búsquedas se desarrollan sobre el sistema actual, y una vez encontradas la o las sentencias deseadas, se identifica sus ubicaciones para poder prestar el servicio.
5. **Recuperación de sentencias:** es un proceso que se desarrolla en conjunto con las búsquedas de sentencias, para optimizar las búsquedas. Es el proceso de recuperar las sentencias desde la base de datos, esta recuperación no se desarrolla de forma íntegra en la actualidad.
6. **Diseminación de sentencias:** esta es otra de las actividades que desarrolla el centro orientada a los servicios del centro. Es la preparación y entrega sistemática de información de documentos nuevos sobre un tema seleccionado por usuarios suscritos.

Nota 1: El SAH correspondiente a la presente investigación se encuentra bien argumentado en el documento adjunto titulado “SAH Tesis BD sentencias penales del CENDIJ 2012”.

2.3 Sistema de Conocimientos en el CENDIJ.

Un ejemplo de la definición de SCO planteada en el capítulo 1 es: muchos profesionales de distintas ramas de la actividad jurídica y social necesitan explotar el conocimiento contenido en los documentos jurídicos que se procesan diariamente en todo el país para desempeñar su trabajo. Por esta razón el CENDIJ, entre otras labores, realiza una descripción de sus fondos documentales y una labor de representación del conocimiento que contienen esos fondos que, en cada caso serán distintos. En este trabajo el enfoque está dirigido específicamente a las sentencias de la materia penal.

El resultado del SCO es la identificación clara y sin ambigüedades de los documentos o las cosas sobre las cuales la base de datos mantendrá información, así como poner de manifiesto las propiedades más relevantes de esas entidades. (Codina, 1998)

Nota 2: El SCO correspondiente a la presente investigación se encuentra bien explicado en el documento adjunto titulado “SCO Tesis BD sentencias penales del CENDIJ 2012”.

2.4 Requisitos funcionales.

Los requisitos funcionales son un conjunto de capacidades o condiciones que el sistema debe cumplir, estableciendo de esta forma el cómo se debe comportar el mismo.

Siguiendo el ciclo de vida de la metodología de Lluís Codina, en la primera fase se realiza el Análisis de requisitos y entrevistas a expertos y usuarios. A partir del análisis de documentos y las entrevistas

realizadas al personal que labora en el CENDIJ y profesionales de la UCI, especializados en el manejo de información jurídica, se definen los requisitos funcionales de la presente investigación, la misma está relacionada directamente con cada una de las partes que conforman el documento sentencia penal independientemente de su tipo. Razón por la cual los requisitos funcionales son:

RF1: Introducir tipo tribunal: la base de datos debe permitir persistir el tipo de tribunal que emitió la sentencia, este tribunal solo puede ser de tres tipos: supremo, provincial o municipal.

RF2: Insertar número de sentencia: cada sentencia tiene un número que del cual la base de datos debe permitir la persistencia.

RF3: Insertar Presidente: la base de datos debe permitir la persistencia del presidente del juicio en el cual se emite la sentencia que se está tratando.

RF4: Insertar jueces: la base de datos debe permitir la persistencia de todos los jueces del juicio en el cual se emite la sentencia que se está tratando.

RF5: Introducir localización en tiempo y espacio: la base de datos debe permitir la persistencia del lugar y la fecha en la que se realiza el juicio en el cual se emite la sentencia que se está tratando.

RF6: Insertar Vista: la base de datos debe permitir la persistencia de todos los datos primarios del juicio a ejecutarse. La Vista es donde el jurado conoce la causa y el delito cometido por el acusado, quién es el acusado y todos sus datos; además de quién está defendiendo al acusado y la medida a cautelar a la que está sujeto el acusado.

RF7: Insertar nombre Ponente: la base de datos debe permitir la persistencia del nombre del ponente del juicio que debe ser un juez o una jueza.

RF8: Insertar resultando: toda sentencia de la materia Penal, independientemente de su tipo, tiene resultando. Si es una sentencia por retirada la que se emite entonces tiene un Único Resultando. Pero si es de otro tipo entonces tiene cuatro Resultandos.

1. **Insertar PRIMER RESULTANDO:** la base de datos debe permitir la persistencia de lo que se ha probado durante la realización del juicio.

2. **Insertar SEGUNDO RESULTANDO:** la base de datos debe permitir guardar el segundo resultando que se deriva de lo que se ha probado en el primero.
3. **Insertar TERCER RESULTANDO:** la base de datos debe permitir guardar la opinión emitida por el ministerio fiscal en este punto del juicio.
4. **Insertar CUARTO RESULTANDO:** la base de datos debe permitir guardar la opinión del letrado (a) de la defensa.

RF9: Insertar considerando: toda sentencia de la materia Penal, independientemente de su tipo, tiene considerando. Si es una sentencia por retirada la que se emite entonces tiene un Único Considerando. Pero si es de otro tipo entonces tiene cinco Considerandos.

1. **Insertar PRIMER CONSIDERANDO:** la base de datos debe permitir la persistencia del primer considerando teniendo en cuenta el delito al que tipifican los hechos declarados y probados y además el artículo del Código Penal en el que son previstos y sancionados.
2. **Insertar SEGUNDO CONSIDERANDO:** la base de datos debe permitir la persistencia del Segundo considerando del juicio donde se responsabiliza al acusado del delito cometido.
3. **Insertar TERCER CONSIDERANDO:** la base de datos debe permitir guardar lo que concurre en la comisión de los hechos realizados por el acusado.
4. **Insertar CUARTO CONSIDERANDO:** la base de datos debe permitir guardar la indemnización por los daños y perjuicios causados por el acusado en sus actos ilícitos.
5. **Insertar QUINTO CONSIDERANDO:** la base de datos debe permitir guardar la acción a realizar para adecuar la medida de la sanción.

RF10: Insertar fallo: la base de datos debe permitir la persistencia de la sanción que se le aplica al acusado debido a su autoría en el delito registrado en la Vista.

RF11: Insertar sanción: la base de datos debe permitir guardar la sanción aplicada al acusado como autor del delito almacenado en la Vista. Se define aquí cuál será y dónde cumplirá la sanción.

RF12: Insertar notificación: la base de datos debe permitir guardar la notificación de la sentencia emitida en la que se le permite al acusado realizar la apelación del juicio a un nivel superior.

2.5 Requisitos no funcionales.

Los Requisitos no funcionales especifican las propiedades o cualidades que el sistema debe cumplir; tienen que ver con las características que de una u otra forma puedan limitar el sistema.

Teniendo en cuenta lo antes planteado los requisitos no funcionales capturados en la presente investigación se dividen en Seguridad, Soporte, Hardware y Software, cada uno de ellos se explica a continuación.

Seguridad:

Se realizarán copias de seguridad de forma tal que los datos puedan ser recuperados en su mayoría o totalmente en caso de fallo. Para dar cumplimiento a este requisito se establece una política de copias de respaldo y recuperación que indica, que se debe realizar una copia normal todos los últimos días de cada mes a las 7:00pm y una copia incremental diaria, excepto el último día de cada mes, a las 6:00pm. Toda esta información se debe guardar en dispositivos masivos de almacenamientos, en el caso de las salvadas mensuales, se harán sobre CD-ROM los cuales son trasladados fuera de la institución para poder ser salvados en caso de derrumbe, incendios o inundaciones en la misma. Se debe mantener también una copia de los datos dentro de la institución, en una computadora habilitada para esto. Además el sistema gestor seleccionado debe presentar facilidades para administrar los roles de usuarios logrando restringir de esta forma el acceso a los datos.

Soporte:

Debido a la cantidad de documentación que se desea gestionar, se requiere de un servidor para la base de datos, que soporte un gran volumen de información.

Requisitos de Software:

Tener instalado sistema operativo Windows XP Service Pack 2 o una versión superior. Tener instalado en el servidor de BD el SGBD PostgreSQL, en una versión igual o superior a la 8.3. Se instala además PGAdmin III a partir de su versión 1.8.4 en correspondencia con la versión del SGBD instalado.

Requisitos de Hardware:

Se requiere de un servidor para la BD con las siguientes características:

1. Una tarjeta de red.
2. Un microprocesador Dual Core o superior, a 2.5 GHZ mínimo.
3. Una memoria RAM DDR2 o superior, a 1Gb mínimo (2 Gb recomendado).
4. Un Disco Duro con 1 Tb de capacidad mínima.
5. Un UPS para salvaguardar el sistema ante altos voltajes, descargas eléctricas y apagones.

Las máquinas clientes deben poseer las siguientes características:

1. Una tarjeta de red para la conexión con el servidor de BD.
2. Un microprocesador a 2.0 GHZ o superior.
3. 512 Mb de RAM como mínimo (1 Gb recomendado o superior).

Debe existir además un maquina independiente, fuera del local del servidor, para almacenar las salvas de la BD, cumpliendo con los requisitos de seguridad. Esta máquina debe poseer las mismas características de almacenamiento que el servidor de BD.

2.6 Modelo conceptual.

La fase de diseño de la metodología de Lluís Codina para la realización de BDD tiene como propósito obtener el modelo conceptual y una propuesta de tratamiento documental.

El modelo conceptual, también conocido como diseño conceptual, consiste en obtener una representación de los recursos de información, con independencia de usuarios, aplicaciones en particular y sin considerar aspectos como eficiencia de la computadora. Esta primera fase consta de dos momentos: análisis de requisitos, donde se centra el trabajo en definir qué es lo que se va a representar, y la conceptualización, donde se piensa en cómo se va a proceder para representar lo antes definido. (Arias, 2007)

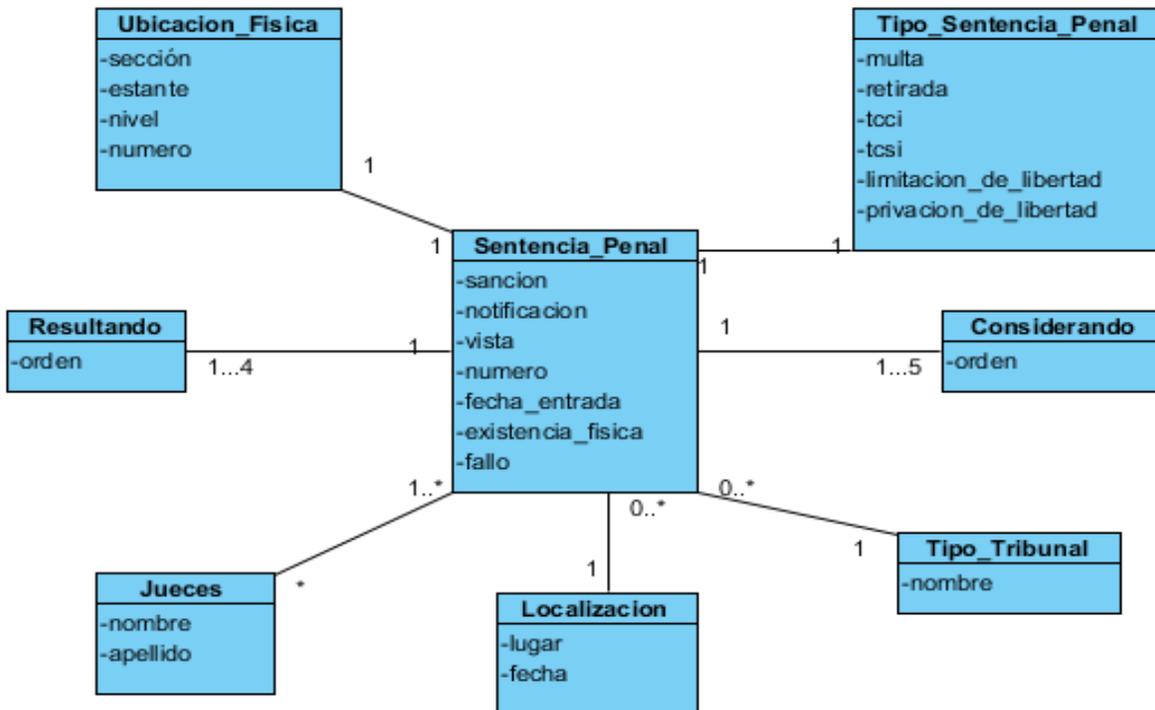


Figura 2. Modelo conceptual. Fuente: Elaboración propia.

Es completamente independiente de los aspectos de implementación, como puede ser el SGBD. El modelo conceptual permite ver con claridad la información utilizada en algún problema o negocio. El objetivo principal de la BD propuesta es permitir la gestión de las sentencias de la materia penal que llegan al CENDIJ, o sea almacenar y recuperar la información contenida en estas, por esta razón el modelo conceptual queda como se muestra en la figura anterior.

2.7 Modelo Entidad Relación.

Un Diagrama o Modelo Entidad Relación¹¹, es una herramienta para el modelado de datos de un sistema de información. Estos modelos expresan entidades relevantes para un sistema de información así como sus interrelaciones y propiedades. (eva.uci.cu)

“... el modelo ER puede ser usado como una base para una vista unificada de los datos, adoptando el enfoque más natural del mundo real que consiste en entidades e interrelaciones...” Peter P. Chen 1976. (eva.uci.cu)

2.7.1 Estándares de nomenclatura en la base de datos.

Nomenclatura de tablas:

El nombre de cada tabla debe ser descriptivo y no se utilizarán tildes.

Las Tablas identifican una entidad de la BD con un nombre completo. Asimismo el prefijo a utilizar para cada tabla será: “T_”.

Una tabla hija debe llevar el nombre de la tabla padre. Si hay espacios entre palabras, deberán separarse por “_”.

Nomenclatura de columnas:

Los nombres de las columnas deben describir el dato que van a almacenar de manera entendible.

Nomenclatura Llave Primaria (Primary Key): PK_*NombreTabla* Ejemplo: PK_Sentencia

Nomenclatura Llave Foránea (Foreign Key): FK_*NombreTablaOrigen_NombreTablaReferenciada*

Ejemplo: FK_T_Vista_T_Sentencia

Nomenclatura de índices: La nomenclatura utilizada en los nombre de los índices consta de la palabra identificadora ‘*indice*’ seguido del signo ‘_’ y seguido a continuación del nombre simple de la tabla a la que se le aplica este índice. Para el caso de la tabla **T_Jueces**, el nombre del índice sería: ‘*indice_jueces*’.

¹¹ **Modelo entidad relación:** en ocasiones denominado por sus siglas, E-R “Entity Relationship”, o “DER” Diagrama Entidad Relación.

2.7.2 Patrones de diseño usados.

Los patrones de diseño de una Base de Datos permiten al usuario crear una BD más fortalecida ya que constituyen una guía que especifica cómo debe ser la BD. Para la realización de la BD que ocupa la solución de la problemática de esta investigación se utilizan dos patrones de diseño de BD: **Patrón de llaves subrogadas** y **Patrón una tabla por dominio**. (Véase epígrafe 1.8).

La utilización de estos patrones se evidencia en el siguiente fragmento del esquema relacional:

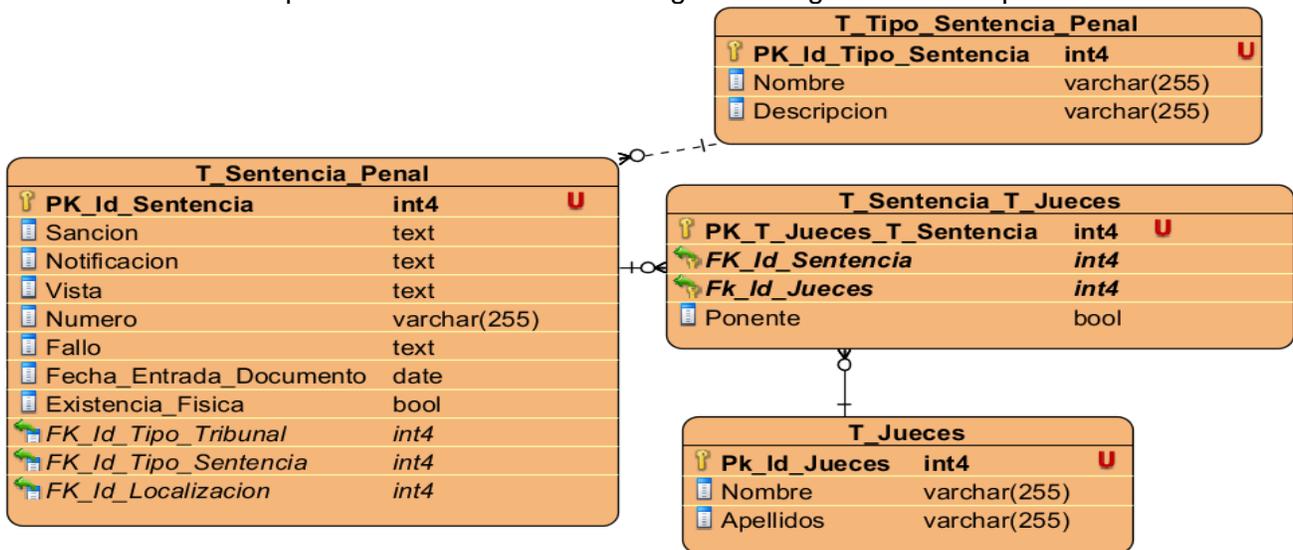


Figura 3. Ejemplo del uso de los patrones de diseño en la BD: Llaves subrogadas y Patrón una tabla por dominio.

La utilización del Patrón llaves subrogadas queda evidenciada en cada una de las tablas mostradas, pues cuentan con una llave primaria única e irrepetible, el tipo de datos de estas es entero. En las tablas de la figura las llaves son: de **T_Sentencia_Penal** es **PK_Id_Sentencia**, de **T_Tipo_Sentencia_Penal** es **PK_Id_Tipo_Sentencia**, de **T_Jueces** es **PK_Id_Jueces**. La tabla **T_Sentencia_T_Jueces**, se genera automáticamente debido a la relación de mucho a mucho existente entre las entidades **T_Sentencia_Penal** y **T_Jueces**, a esta también se le crea su llave primaria quedando: **PK_T_Jueces_T_Sentencia**, siguiendo los estándares de nomenclatura definidos. Los jueces relacionados con una sentencia pertenecen a un dominio, utilizando el patrón de diseño de BD *una tabla por dominio* se crea la tabla nomencladora **T_Jueces**, que contiene los atributos propios de los jueces en cuestión. El tipo de sentencia penal es también un dominio y se le realizó la tabla nomencladora **T_Tipo_Sentencia_Penal** con sus atributos correspondientes.

2.7.3 Esquema Relacional.

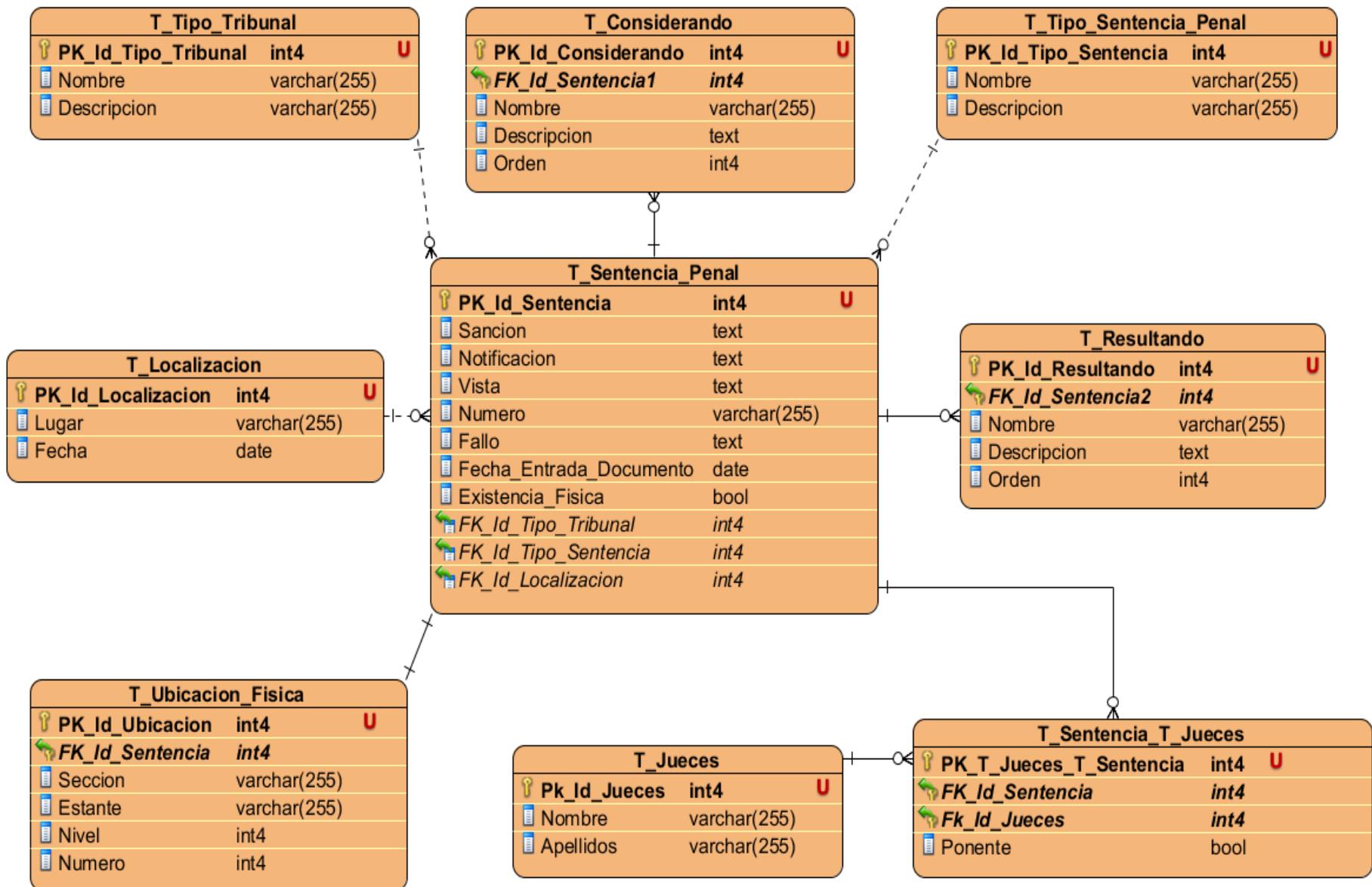


Figura 4. Esquema Relacional.

El esquema relacional es la representación lógica del diagrama entidad relación. Este es el modelo de bases de datos más utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente. (Javier Alcolea Velázquez, 2007)

El Modelo Relacional provee una estructura cercana a la manera física en que los datos de una base de datos están almacenados, las relaciones. Una relación es como una tabla en la que cada fila es una tupla de información y cada columna un atributo que toma valores sobre un dominio específico.

El esquema relacional correspondiente a la investigación se representa en la figura 4.

2.8 Diccionario de Datos.

El diccionario de datos es un documento en el que se especifica la naturaleza y descripción de toda la información persistente en el sistema. Es decir, toda la información que deberá ser almacenada en la base de datos. Viene a la par del esquema relacional, pues en el diccionario de datos es donde se define lo más detallado posible la estructura de las tablas que lo conforman. (Codina, 1998)

Tabla T_Sentencia_Penal:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
PK_Id_Sentencia	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
Sancion	Sanción impuesta al acusado, reflejada en la sentencia.	Texto	Si	Lenguaje libre	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Notificacion	Notificación adjunta en la sentencia.	Texto	No	Lenguaje libre	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Vista	Vista de la	Texto	No	Lenguaje	No puede quedar como

	sentencia, introducción de la misma.			libre	campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Numero	Número identificativo de la sentencia.	Alfanumérico	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Fallo	Disposición de la sentencia, sanción impuesta al acusado	Texto	No	Lenguaje libre	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Fecha_Entrada_Documento	Es la fecha en que fue recepcionado el documento en la entidad.	Date	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Existencia_Fisica	Dato específico de si existe o no la sentencia en formato duro.	Bool	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Id_Tipo_Tribunal	Identificativo de la tabla T_Tipo_Tribun	Integer	No		Se obtiene de la tabla T_Tipo_Tribunal .

	al como llave foránea en esta tabla.				
Id_Tipo_Sentencia_Penal	Identificativo de la tabla T_Tipo_Sentencia_Penal como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Tipo_Sentencia_Penal .
Id_Localización	Identificativo de la tabla T_Localizacion como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Localizacion .

Nota 3: El diccionario de datos correspondiente a la presente investigación se encuentra bien argumentado en el documento adjunto titulado “Diccionario de Datos Tesis BD sentencias penales del CENDIJ 2012”.

Nota 4: Las restantes tablas del diccionario de datos se encuentran anexadas al final del presente documento.

2.9 Optimización de base de datos.

Una BD solo alcanzará su nivel de máximo rendimiento cuando esté optimizada, pues este proceso prevé que realice su esfuerzo total con menor cantidad de recursos, logrando la eficacia. Para optimizar una base de datos, en este caso en PostgreSQL, es necesario aplicar una serie de técnicas que garanticen que continuará realizando el mismo trabajo que se desea pero consumiendo menos recursos y tiempo.

El **optimizador** es el componente del sistema que determina cómo implementar las peticiones del usuario (las cuales se ocupan del *qué*, no del *cómo*). Por lo tanto, ya que los sistemas relacionales asumen la responsabilidad de navegar alrededor de la base de datos almacenada para localizar los datos deseados, en ocasiones se les describe como sistemas de navegación automática. La optimización y la navegación automática son prerrequisitos para la independencia física de los datos.

2.9.1 Optimización de consultas mediante el uso de índices.

Índices utilizados en la Base de Datos

En la BD se identificaron los campos que son objeto de búsquedas con mayor frecuencia. Utilizando el método de indexación BTREE, se definió un índice por tabla, implementándolos sobre los campos identificados en el diccionario de datos. Existen tablas que no poseen índices implementados dado que sus campos no son objeto de búsquedas frecuentemente.

Para los campos de tipo texto, se implementaron métodos de normalización de los datos insertados en dichos campos, con el objetivo de optimizar las búsquedas mediante índices. O sea, como trabajar acentos, eñes, minúsculas y mayúsculas para que el usuario encuentre lo que busca, ej. Para que una búsqueda 'viola', encuentre ocurrencias como 'VIOLACIÓN', 'Violadores' o 'violando la ley'. Para ello se guardarán los datos insertados tal como vienen, respetando el tipo de lenguaje documental expuesto en el diccionario de datos, y creará un índice normalizado, usando una función PL/PGSQL de tipo IMMUTABLE. Para el índice '*índice_considerando*', de la tabla T_Considerando, se implementa la siguiente función:

```
CREATE OR REPLACE FUNCTION str_normalize (value TEXT) RETURNS text AS $$
BEGIN
RETURN lower(translate(value,
'áàèèíóòóúüäëïöüÁÀÉÈÏÓÒÓÚÙÄËÏÖÜñÑçÇ"?¿¡[]` } ,:;=&%$#!\!^oa<>',
'aaeeiiioouuaeiouAAEEIIIOOUUAEIOUnNcC '));
END;
$$ LANGUAGE plpgsql IMMUTABLE;
```

Una vez implementada la función que permita desarrollar esta normalización, el índice se crearía de la siguiente manera:

```
CREATE INDEX indice_considerando ON t_considerando USING btree
(str_normalize(descripcion) text_pattern_ops);
```

De esta forma normalizamos los datos del campo descripción dentro de la tabla T_Considerando, y a la vez se crea el índice '*índice_considerando*' optimizado para la búsqueda seleccionada. De igual manera se implementó el índice utilizado para la tabla T_Resultando:

```
CREATE INDEX indice_resultando ON t_resultando USING btree
(str_normalize(descripcion) text_pattern_ops);
```

Para el caso donde el campo indizado sea de tipo *varchar* y no de tipo *texto*, como el utilizado en la tabla T_Jueces, el índice quedó implementado como se muestra a continuación:

```
CREATE INDEX indice_jueces ON t_jueces USING btree (str_normalize(nombre)
varchar_pattern_ops);
```

El usar índices evita la sobrecarga del CPU y permite mayor rapidez en la ejecución de las consultas, también es una ventaja en los campos que no contengan datos duplicados.

2.10 Pasos a seguir para diseñar la BD que permita la gestión documental de todos los documentos persistentes en el CENDIJ.

El presente epígrafe constituye una guía a seguir para realizar la ampliación de la BD obtenida. Entiéndase por ampliación, modelar y generar una BD que almacene todos los documentos que se encuentran en el CENDIJ, permitiendo la gestión de la información contenida en los mismos, siguiendo la metodología y los estándares establecidos para la realización de la BD que almacena las sentencias penales y permite el acceso al contenido íntegro de la información persistente en las mismas, desarrollada en la presente investigación.

Un aspecto muy importante a tener en cuenta, son los cambios que pueden ocurrir en el modelo conceptual y en el Esquema Relacional correspondiente a la BD desarrollada, ya que en estos se tendrán en cuenta los documentos restantes, como son: disposiciones del Consejo de Gobierno, legislaciones, doctrinas, jurisprudencias y las sentencias correspondientes a las materias: económica, civil, laboral y administrativa.

Tomando como punto de partida la BD implementada, se recomienda hacer uso de las mismas herramientas y seguir y respetar la metodología utilizada; para esto último se debe cumplir el ciclo de vida definido para la misma. Por esta razón se explica **paso por paso** el procedimiento a seguir para la implementación de la BDD a texto completo que almacene todos los documentos presentes en el CENDIJ y permita el acceso al contenido íntegro de la información persistente en ellos.

1. **Definir el SAH:** teniendo en cuenta el concepto emitido en la "Metodología de análisis de sistemas de información y diseño de bases de datos documentales: aspectos lógicos y funcionales." Ya se conoce que el SAH es el *CENDIJ*, por lo que ya está definido.
2. **Definir el SCO:** aumenta en gran medida el volumen del SCO ya que ahora debe almacenar todos los documentos existentes en el CENDIJ: *disposiciones del Consejo de Gobierno, legislaciones, doctrinas, jurisprudencias y las sentencias correspondientes a las materias: económica, civil, laboral y administrativa.*
3. **Definir Requisitos funcionales:** se deben realizar las entrevistas y averiguaciones pertinentes, obtener un modelo de cada tipo de documento y analizar la estructura de los

mismos para elegir por cada uno de ellos las acciones que el sistema debe cumplir para lograr su almacenamiento total y no perder ningún aspecto o punto importante que defina la integridad de los datos que debe contener un documento específico.

4. **Definir Requisitos no funcionales:** los requisitos no funcionales definidos para la BD desarrollada pueden soportar la ampliación de la misma, por lo que estos deben ser comunes para toda la BD, se recomienda utilizar los que ya se encuentran definidos, a menos que se considere indispensable cambiarlos para un correcto o mejor funcionamiento de la nueva BD.
5. **Definir Modelo Conceptual:** basándose en las entrevistas y el estudio de los documentos que se realice pertinentemente, se deben identificar los principales conceptos que forman parte de este nuevo negocio.

La siguiente figura muestra una propuesta inicial de cómo quedaría el modelo conceptual para esta BD ampliada, este modelo constituye un punto de partida para realizar el modelo conceptual correspondiente a la nueva BD.

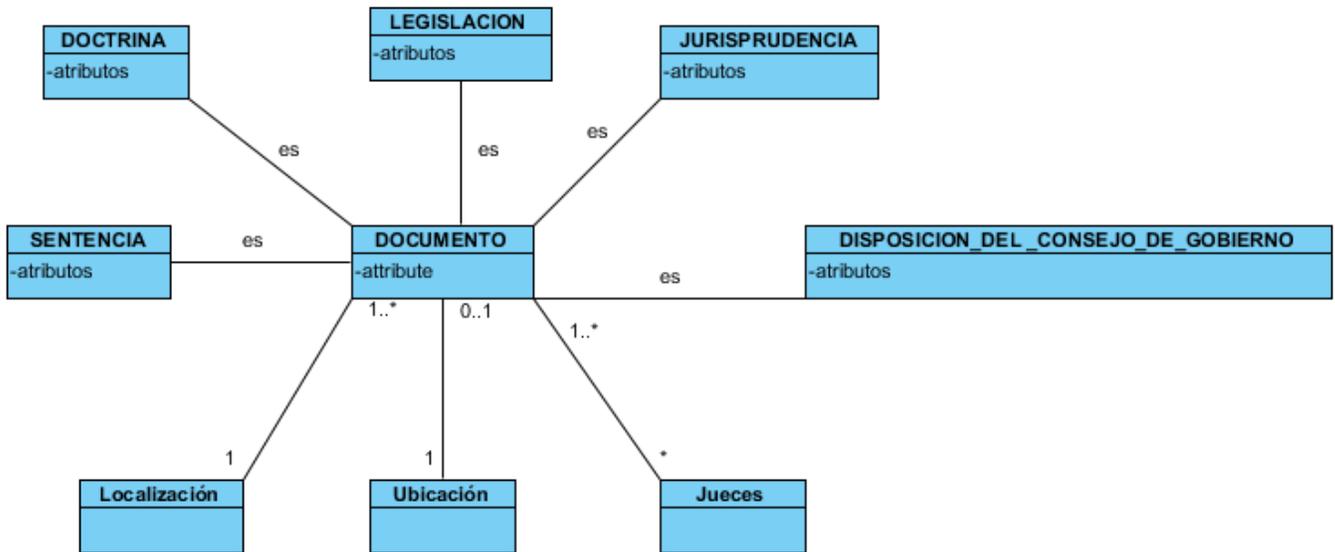


Figura 5. Propuesta inicial para el modelo conceptual de la BD ampliada. Fuente: Elaboración Propia

6. **Definir Esquema Relacional:** El esquema sufrirá cambios que no estarán enfocados solamente al tamaño del mismo, pues existen tablas que son comunes para todos los documentos que almacena el CENDIJ, tales como: *T_Ubicación* (todos los documentos en formato duro poseen una ubicación física y esta es almacenada en esta tabla); *T_Localización* (se guarda el lugar y la fecha en el que fue emitido un documento independientemente de su tipo); *T_Jueces* (contiene los datos de los jueces involucrados en la emisión de algún

documento jurídico, independientemente de su tipo). Guiándose por el modelo conceptual resulta menos compleja la realización del esquema relacional.

7. **Definir Estándares de Nomenclatura para la BD:** para lograr uniformidad en el diseño de la BD se hace necesario seguir el estándar de nomenclatura ya definido en el epígrafe 2.7.1.
8. **Definir Patrones de Diseño de BD a utilizar:** existen 2 patrones de diseño de BD definidos y deben ser utilizados en el diseño del esquema relacional de esta nueva BD y se deja a consideración del diseñador valorar la utilización de otro(s) en caso de ser necesario.
9. **Definir Diccionario de Datos:** es necesario confeccionar el diccionario de datos correspondiente al ER relacional obtenido, siguiendo los parámetros establecidos para cada atributo. Estos son:
 - Atributo o Etiqueta:* nombre del atributo
 - Domino:* lo que representa en esa tabla, por ejemplo el ID.
 - Tipos de datos:* tipo de dato del atributo. Puede ser integer, varchar, bool, etc.
 - Indexación:* se define si el atributo para por el proceso de indexación o no.
 - Tratamiento Documental:* se determina el tipo de lenguaje del atributo. Ej.: lenguaje libre.
 - Controles de Validación:* se explica las pautas que debe cumplir un atributo para poder ser almacenado.
10. **Implementación de la BD:** Una vez realizado el esquema relacional en el Visual Paradigm, se sincroniza al diagrama de clases y a través del mapeo relacional de objetos se genera la BD, haciendo uso del driver *postgresql-8.1-404.jdbc3.jar* y contando con una BD vacía en el SGBD.
11. **Optimización de la BD:** se recomienda realizar la optimización de la BD a través de índices. Utilizar la función *str_normalize()* para el tratamiento de los textos y para la creación de los nuevos índices.
12. **Validación de la BD:** Desarrollar pruebas teóricas y funcionales a la BD que se obtenga y si se considera necesario realizar los cambios o ajustes pertinentes cumpliendo con el último punto de la etapa implantación según plantea la metodología utilizada.

Este procedimiento servirá de guía para la confección de una BDD que almacene los documentos del CENDIJ, permitiendo su gestión, optimizando su almacenaje y facilitando el acceso y recuperación a la información que estos contienen. La BD que se obtenga, debe poder ser integrado a un SIBO.

2.11 Conclusiones parciales.

Al finalizar el capítulo se concluye lo siguiente:

- El análisis realizado a las sentencias y su estructura permitió realizar el modelo conceptual correspondiente para dar solución a la investigación.
- La realización del esquema relacional permitió la correcta implementación de la BD haciendo uso de la metodología seleccionada.
- La BD obtenida posee las características de una BDD, dando solución al objetivo propuesto en la presente investigación.

CAPÍTULO 3. Validación de la base de datos.

3.1 Introducción.

Dos formas de validación conforman el contenido de este capítulo. En primer lugar se tiene la validación teórica que está conformada por el análisis de la integridad de la información, normalización de la BD y redundancia de los datos. En segundo lugar se encuentra la validación funcional conformada por las pruebas de volumen y las pruebas de carga realizadas a la BD. Además de la propuesta de un mecanismo que garantice la seguridad de la BD.

3.2 Validación teórica.

La etapa de diseño constituye un punto crucial en el proceso de desarrollo de la BD que se desarrolla en la presente investigación; razón por la cual se deben tener presente algunos aspectos elementales que garantizarán el buen diseño de dicha BD. Tales como la normalización del diseño, la integridad de los datos, el análisis de redundancia de la información y la seguridad de los datos. Garantizando, con este último, el control de acceso a los datos por el personal autorizado y la protección de estos datos para que no se dañen debido a acciones que no se hayan controlado anteriormente.

3.2.1 Integridad de la información.

La integridad de la información en una BD se refiere a la corrección y exactitud de toda la información almacenada. Estas sin importar su tipo pueden estar sometidas a un grupo de restricciones de integridad con complejidad diversa, asegurando que los datos no se vuelvan corruptos.

Integridad de los datos.

La integridad de los datos se refiere a la validez, corrección, consistencia y completitud de los datos almacenados en la BD. Su principal función se basa en proteger la base de datos contra operaciones que introduzcan inconsistencia en los datos. La integridad se expresa normalmente a través de restricciones o reglas, las cuales la base de datos no puede quebrantar para certificar que los registros de tablas relacionadas son válidos y que no se eliminen o modifiquen accidentalmente datos relacionados ya que pueden producirse errores de integridad. La integridad de los datos puede dividirse en varias categorías tales como:

- Integridad de entidad:

Define una fila como entidad única para una tabla determinada. La integridad de entidad exige la integridad de las columnas de los identificadores o la clave principal de una tabla, mediante índices y restricciones UNIQUE, o restricciones PRIMARY KEY (llave primaria). (Rodríguez., 2009)

Este tipo de integridad se cumple en la base de datos propuesta ya que cada tabla tiene definida su llave primaria, la cual no está nula ni repetida. Es de vital importancia especificar la llave principal para cada registro, ya que esta representa la herramienta fundamental que utiliza el servidor de base de datos para seleccionar la información que se necesite, evitando así la duplicidad de los registros.

➤ Integridad de dominio:

Este tipo de integridad viene dada por la validez de las entradas para una columna determinada. Puede exigir la integridad de dominio para restringir el tipo mediante el tipo de datos, además del formato y el intervalo de valores posibles mediante restricciones FOREIGN KEY (llaves foráneas), restricciones CHECK, definiciones DEFAULT, definiciones NOT NULL y reglas. (Microsoft, 2010)

Una definición adecuada de las restricciones de los dominios no solo permite verificar los valores introducidos en la base de datos, sino también examinar las consultas para asegurarse de que tengan sentido las comparaciones que hagan.

Para lograr la integridad de dominio en la BD resultante de la presente investigación se hizo uso de los tipos de datos predefinidos por la herramienta de modelado Visual Paradigm, específicamente integer, varchar, date, text y bool. Además se revisan los valores que no pueden clasificarse de nulos haciendo uso de la restricción NOT NULL.

➤ Integridad referencial:

La integridad referencial es un concepto de base de datos que garantiza que las relaciones entre las tablas siguen siendo coherentes. Cuando una tabla tiene una clave externa a otra tabla, el concepto de integridad referencial establece que no se puede agregar un registro a la tabla que contiene la clave externa a menos que exista un registro correspondiente en la tabla vinculada. También incluye las técnicas conocidas como actualización en cascada y eliminar en cascada, que aseguran que los cambios realizados a la tabla vinculada se reflejan en la tabla principal. (Chapple, 2010)

La integridad referencial implica que en todo momento todos los datos involucrados en la BD sean correctos y sin repeticiones innecesarias. Todas las bases de datos relacionales presentan esta propiedad pues el software gestor es responsable de su cumplimiento.

3.2.2 Normalización de la base de datos.

La normalización es la expresión formal del modo de realizar un buen diseño. Provee los medios necesarios para describir la estructura lógica de los datos en un sistema de información. El proceso de normalización permite evitar algunos elementos que pueden ser no deseados en las bases de datos como:

- Redundancia o repetición de los datos en el sistema.
- Inconsistencias de actualización de datos como resultado de las actualizaciones parciales y la redundancia de la información.
- Anomalías de borrado o pérdidas no intencionadas de datos.
- Anomalías de inserción o imposibilidad de adicionar datos en la base de datos debido a la ausencia de otros datos.

El proceso de normalización consta de diferentes formas normales, esto no quiere decir que para que una base de datos esté normalizada tiene que haber pasado por todas estas; incluso, no siempre es recomendable tener una base de datos que se encuentre en el nivel más alto de normalización, puede llevar a un nivel de complejidad no deseado por aquellos usuarios que interactúen con la base de datos, que pudiera ser evitado si estuviera en un nivel más bajo de normalización.

Decir que una base de datos está en la forma normal N es decir que todas sus tablas se encuentran en la forma normal N. Habitualmente, las tres primeras formas normales son suficientes para cubrir las necesidades de la mayoría de las bases de datos.

Primera Forma Normal: una tabla estaría en primera forma normal (1FN) si y solo si todos los atributos son atómicos. Un atributo es atómico si los elementos del dominio son indivisibles, o sea, mínimos, por lo que se deben eliminar todos los atributos compuestos y multivaluados. Otra característica que define una tabla en 1FN es que cada uno de los campos contiene un único valor para un registro determinado.

Segunda Forma Normal: Una BD se encuentra en segunda forma normal (2FN) si está en 1FN y los atributos que no forman parte de ninguna clave dependen de completamente de la llave primaria (Primary Key), o sea, que no existan dependencias parciales. Se podría decir que la segunda forma normal está basada en el concepto de dependencias completamente funcionales.

Tercera Forma Normal: Para llevar una tabla a tercera forma normal (3FN) es necesario que ella esté en 2FN y no debe existir ninguna dependencia funcional transitiva entre los atributos que no son clave. Entiéndase por transitiva aquella en la cual las columnas que no son llave son dependientes de otras

columnas que tampoco son llave. En otras palabras los campos de las tablas no dependen unos de otros.

La BD que se está analizando se encuentra en 3FN puesto que cumple con las especificaciones de las tres formas expuestas anteriormente. Desde un principio se tuvo en cuenta la creación de atributos atómicos, que los atributos que no son llave primaria o foránea, dependieran completamente de la llave primaria y los campos de las tablas no dependen unos de otros.

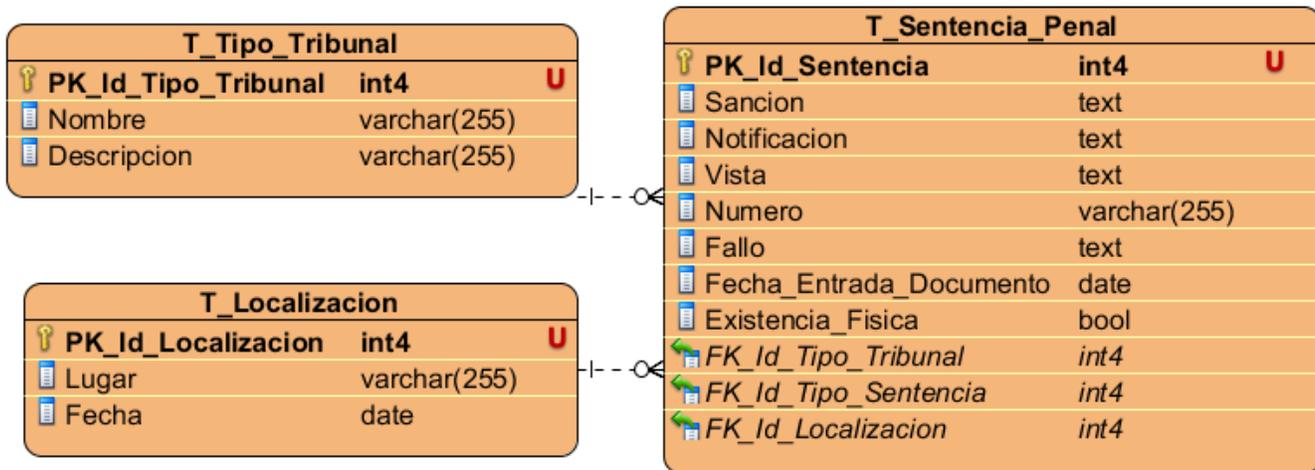


Figura 6. Ejemplo de tablas en 3ra Forma Normal.

La forma normal en la que se encuentra la BD es la más usada en la totalidad de los productos que utilizan bases de datos porque garantiza eliminar la redundancia de información y permite soportar anomalías frente a las actualizaciones o modificaciones que se le realicen a la misma.

Las diversas formas normales constituyen un ordenamiento total, en el sentido de que toda BD en un determinado nivel de normalización está también automáticamente en todos los niveles inferiores; aunque lo contrario no es cierto, ya que existen BD en cada nivel que no se encuentran necesariamente en ningún nivel superior.

3.2.3 Redundancia de los datos.

La redundancia de datos es aquella información duplicada o almacenada varias veces en la misma base de datos. Esto dificulta la tarea de modificación de datos y es el motivo más frecuente de su inconsistencia. Además representa un mayor espacio de almacenamiento, que influye negativamente en el tiempo de acceso a los datos. Con un buen diseño de una base de datos se logrará evitar la aparición de información repetida o redundante en el sistema. (CORONADO, 2004)

Un modelo de datos normalizado reduce en gran medida la redundancia. La presencia de redundancia implicaría la necesidad de utilizar triggers para asegurar que se mantenga la consistencia cuando los datos son insertados, modificados o eliminados. En la BD propuesta como solución al problema de la presente investigación no existe redundancia pues la misma se encuentra en 3FN y además cuenta con tablas nomencladoras para almacenar datos comunes que puedan existir dentro de la misma. Ej.: T_Jueces.

3.3 Seguridad de la base de datos.

La seguridad de la BD se basa en las acciones a tomar para garantizar que la información almacenada sea consultada por los usuarios autorizados. Los datos que guarda una BD tienen el riesgo de sufrir ataques que puedan provocar la pérdida o modificación no deseada de la información, por lo cual es sumamente importante establecer los mecanismos necesarios para asegurar la integridad de los mismos, dejando claro los privilegios de acceso de los usuarios. Es importante además garantizar la recuperación de la información en caso de que ocurra alguna anomalía en la cual se produzca algún fallo con la BD que provoque la pérdida de información.

Para garantizar la seguridad de la BD propuesta se define la implementación de roles y usuarios para el acceso a la BD. Se define el rol *Administradores* para el o los administradores del sistema, los cuales tendrán acceso pleno sobre toda la BD. Se define además el rol *Cendij* para los trabajadores del centro con acceso limitado a las funciones que realizan de insertar y actualizar datos sobre las sentencias penales. Por último se define el rol *Usuarios* para todas las personas que acceden al CENDIJ en busca de información, este rol tendrá privilegios de solo lectura sobre la base de datos, pues su función principal es permitir el acceso a los datos sin modificarlos. Las contraseñas de todos los usuarios con acceso al sistema estarán encriptadas.

El cumplimiento de los requisitos no funcionales de seguridad garantizará salvaguardar los datos de la BD en caso de desastres.

3.4 Validación Funcional.

Una vez analizada la seguridad, redundancia e integridad de la base de datos se procede a la validación funcional de la misma. Este es un punto muy importante, pues se comprueba que la base de datos propuesta, cumple con los requisitos funcionales que fueron definidos en el capítulo anterior. Para ello se realizan un conjunto de pruebas que tienen como objetivo simular una carga de

producción real y posteriormente observar el comportamiento que tiene la base de datos bajo dicha carga. Esto posibilitará darle solución a los problemas de rendimientos existentes en la base de datos, antes de que la misma vaya a ser utilizada por los usuarios establecidos. Las herramientas seleccionadas para dichas pruebas fueron: Data Generator 2005 for PostgreSQL y Apache JMeter 2.3.

3.4.1 Pruebas de Volumen.

Las pruebas de volumen son realizadas con el fin de conocer si la BD soporta la cantidad de datos para la que fue diseñada o si esta falla ante cierta cantidad de datos, pues analiza el comportamiento de esta a través de la generación de datos aleatorios en la BD, determinando el tiempo que demora en generarlos y dando a conocer errores en caso de que se produzcan. Los datos generados para llenar la BD serán útiles para realizar las pruebas de carga.

Se configuró el Data Generator para que llenara la BD con un rango de datos equivalentes al doble del que tendrá en un período no menor de un año, teniendo en cuenta que en la BD diseñada cada registro corresponde con un documento es necesario realizar las pruebas de volumen a la BD en general y no se recomienda realizar las pruebas en cada una de las tablas por separado puesto que esto puede generar valores NULL en atributos que tienen la restricción NOT NULL activa provocando errores que no permiten el correcto llenado de la BD.

Se pueden definir diferentes valores a generar para cada tabla, por ejemplo en la tabla T_Tipo_Sentencia_Penal no es necesario generar más de diez datos pues se conoce que los tipos de sentencia penales son: Limitación de libertad, Privación de libertad, Multa, Retirada, TCCI y TCSI lo que nos da un total de seis tipos pero se generan diez previendo que se definan nuevos tipos de sentencias en el futuro. Otra tabla en la que se aplica este ejemplo es en la tabla T_Tipo_Tribunal, pues solo existen tres tipos de tribunales: Tribunal Municipal, Tribunal Provincial y Tribunal Supremo, pero se generan cinco tipos de tribunales en esta tabla por si esto deja de ser un estándar en el futuro. La tabla T_Ubicacion_Fisica fue llenada con la mitad de las sentencias que se generaron, pues esta tabla almacena la dirección donde se encuentra el documento en formato duro en caso de existir, ya que la mayoría de las sentencias se almacenan solamente en formato digital. Se generaron 1000 datos en la tabla T_Jueces, y en la tabla T_Localizacion fueron generados un total de 500 localizaciones, pues no existen más de 500 tribunales en Cuba.

El resto de las tablas fue llenado con 10000 datos cada una generando un total de 46515 datos en un tiempo de 5:46 minutos y cero errores como se muestra en la figura 6.

Mientras se añadían datos a las tablas no se presentaron problemas de desbordamiento de atributos, columnas o tipo de datos, ni de límite de capacidad. Utilizando esta herramienta se pudo verificar la integridad de los datos, además de garantizar que el diseño de las estructuras de la BD y el gestor utilizado para su desarrollo, soportan el cúmulo de información requerido para su funcionamiento.

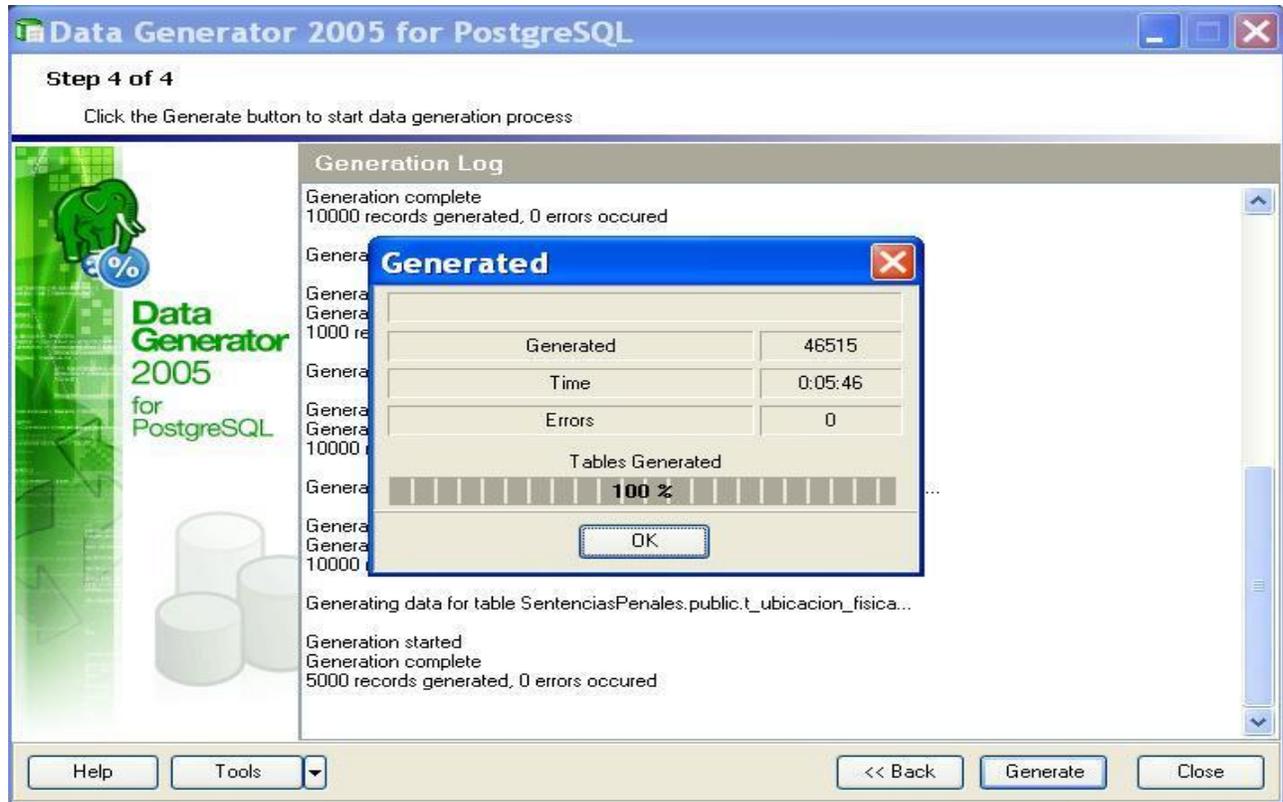


Figura 7. Prueba de Volumen.

3.4.2 Prueba de Carga.

Las pruebas de cargas son utilizadas normalmente para someter a la aplicación al límite de su funcionamiento, mediante la ejecución de un número de usuarios muy superior al esperado. Este test tiene como finalidad el determinar la robustez de una aplicación cuando la carga es extrema y ayuda a administradores a determinar si la aplicación se comportará correctamente en dichas situaciones. Otro de los objetivos de este tipo de pruebas es determinar el límite real de la aplicación en cuanto a número de usuarios concurrentes y número de transacciones por segundo.

Para la realización de esta prueba primeramente se creó un grupo de hilos donde se definió la cantidad de usuarios que realizarían peticiones al servidor. Luego se configuró el acceso a la BD a través del archivo Configuración de la conexión JDBC. Se definió la consulta que se utilizaría para la

realización de las pruebas, esta devuelve todas las sentencias realizadas por el juez Juan. La misma queda como se muestra a continuación:

```
select t_sentencia_penal.numero, t_sentencia_penal.fecha_entrada_documento,  
t_sentencia_penal.existencia_fisica  
from t_sentencia_penal, t_sentencia_t_jueces, t_jueces  
where t_sentencia_penal.pk_id_sentencia=t_sentencia_t_jueces.t_sentencia_penalpk_id_sentencia  
and t_sentencia_t_jueces.t_sentencia_penalpk_id_sentencia=t_jueces.pk_id_jueces  
and t_jueces.nombre='Juan'
```

Se realizaron varias pruebas utilizando la consulta mostrada anteriormente, pero definiendo una cantidad diferente de usuarios conectados en cada caso, por lo que se arrojaron resultados distintos que se presentan a continuación a través de cuatro variables que son muy importantes a tener en cuenta.

- ✓ **Media:** tiempo promedio de respuesta de todas las peticiones.
- ✓ **Mediana:** tiempo promedio de la mitad de los resultados, la otra mitad tomará un tiempo entre la mediana y el valor máximo.
- ✓ **Mínimo:** mínimo tiempo de respuesta de una petición.
- ✓ **Máximo:** máximo tiempo de respuesta de una petición.

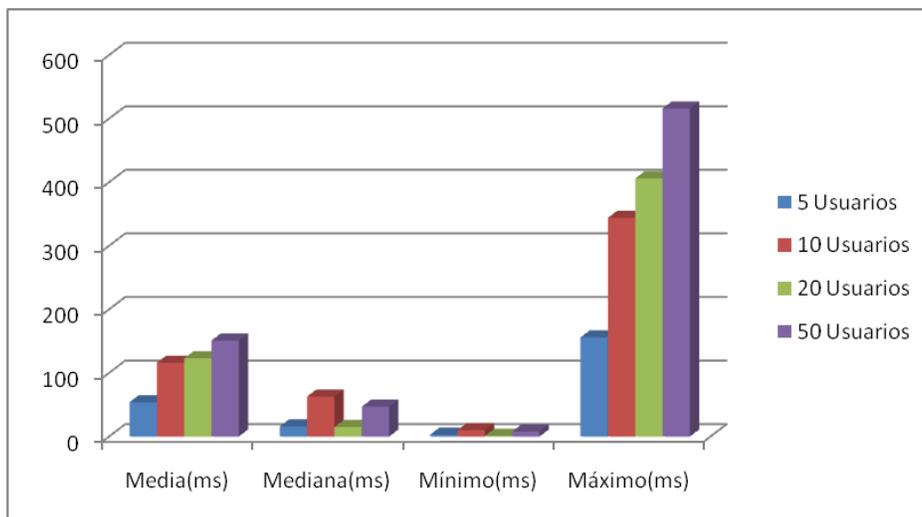


Figura 8. Representación de los resultados de la prueba de Carga.

En la figura anterior se representa gráficamente la prueba de carga realizada. En el eje horizontal se representan las cuatro variables a tener en cuenta en el resultado final, y en el eje vertical los tiempos

de respuestas en milisegundos (ms). Los valores utilizados para la realización de esta prueba se deben a que en el CENDIJ existen aproximadamente 20 computadoras en total, ahí se incluyen las utilizadas por el personal que labora en el CENDIJ y las disponibles para la prestación de servicios. Existen 2 computadoras para la prestación de servicios a los usuarios, estas cuentan con una memoria RAM de 512 Mb. Para la realización de estas pruebas se tuvo en cuenta un valor mínimo (5, azul), dos valores medios (10, rojo y 20, verde) y uno con vista futurista (50, morado). Finalmente se evidencia que el rendimiento de la BD fue satisfactorio para cada una de las cantidades de usuarios concurrentes probadas, evidenciándose de esta forma la eficiencia de la misma.

Nota 5: Algunas de las imágenes del procedimiento realizado para 50 usuarios se encuentran anexadas al presente trabajo.

3.5 Conclusiones parciales.

Al finalizar el capítulo se concluye lo siguiente:

- La realización de las pruebas teóricas permitió validar el modelo de datos, llevando la BD obtenida a 3ra FN, eliminando la redundancia y garantizando la integridad de los datos almacenados.
- Con las pruebas funcionales se logró conocer el comportamiento de la BD con diferente cantidad de usuarios realizando peticiones a la misma concurrentemente.
- Las validaciones realizadas arrojaron resultados que dan por sentado el pleno funcionamiento de la base de datos y el total cumplimiento de los requisitos establecidos.

Conclusiones.

Una vez finalizada esta investigación se arriba a las siguientes conclusiones:

- La fundamentación teórica sobre el campo de acción en el cual se enmarca el objetivo general de la investigación permitió identificar los principales elementos a tener en cuenta en el desarrollo de una BDD para el almacenamiento de las sentencias penales en el CENDIJ.
- Se logró implementar una BDD para el almacenamiento de las sentencias penales, la cual sirve como punto de partida para el desarrollo de una BDD que permita almacenar todos documentos jurídicos en formato digital existentes en el CENDIJ.
- Se demostró mediante las pruebas teóricas y funcionales el pleno funcionamiento de la base de datos y el total cumplimiento de los requisitos establecidos.

Recomendaciones.

Se recomienda:

- Desarrollar una herramienta capaz de automatizar la inserción de documentos en la BD.
- Basándose en la BD desarrollada, realizar un sistema que permita insertar los demás tipos de sentencias y documentos que almacena el CENDIJ.
- Continuar desarrollando la BDD obtenida en el trabajo, con el objetivo de en un futuro poder almacenar en la misma todos los documentos en formato digital existentes en el CENDIJ.

Bibliografía Referenciada.

- ✓ **Blaha, Michael. 2010.** *PATTERNS OF DATA MODELING*. FL 33487-2742 : CRC Press, 2010. FL 33487-2742.
- ✓ **Collector, Garbage. 2004.** Garbage Collector. *Garbage Collector*. [En línea] 1 de noviembre de 2004.
- ✓ *Confeción*. **Emilio Laguna Serrano, Antonio Irazabal Nerpel y Adolfo Valle Bacero. 2000.** 2, Madrid : Revista Española de Documentación Científica, 2000, Vol. 12.
- ✓ *Constructor de Tesoros CAT. Confeción Automática de Tesoros*. **Emilio Laguna Serrano, Adolfo Valle Bracero, Antonio Irazazabal Nerpel. 2007.** 2, Madrid : Revista Española de Documentación Científica, 2007, Vol. 12.
- ✓ **Cruzata., Yordanis García Leiva Héctor Luis Duran. 2011.** *Procedimiento para el diseño de un Sistema de Información Basado en Ontologías Jurídicas*. Habana, Cuba : UCI, 2011.
- ✓ **Date, C.J. 2001.** *INTRODUCCIÓN A LOS SISTEMAS DE BASES DE DATOS 7ma Edición*. México, : PEARSON EDUCACIÓN, , 2001. ISBN: 968-444-419-2.
- ✓ **Definición.DE. 2008-2012.** Definición.DE. *Definición.DE*. [En línea] Gestionado con WordPress, 2008-2012. <http://definicion.de/metodologia/>. Copyright 2008-2012.
- ✓ **Ernest Abadal, Lluís Codina. 2005.** *Recuperación de Información en Bases de datos documentales: Características, funciones y métodos*. Madrid : Síntesis, 2005. 84-9756-263-1.
- ✓ *Estrategia de modelado de procesos de negocio para el desarrollo de sistemas de*. **Diana Valdés González, Yoena Domínguez Rivero, Keidy García Lira. 2010.** 3, Habana, Cuba : Revista Infociencia, 2010, Vol. 14.
- ✓ **Gavilán, C.M. 2009.** *Lenguajes documentales Principales tipos de clasificación Encabezamientos de materia, descriptores y tesauros*. 2009.
- ✓ **Group, The PostgreSQL Global Development. 1996-2012.** PostgreSQL. *The PostgreSQL Global Development Group*. [En línea] Copyright Privacy Policy|About PostgreSQL, 1996-2012. <http://www.postgresql.org/docs/8.3/static/textsearch-indexes.html>.
- ✓ **Herrera, Antonio Gabriel. 2006.** *Modelos de sistemas de recuperación de información documental basados en información lingüística difusa*. Granada : Escuela Técnica Superior de Ingeniería Informática, 2006.

- ✓ **Javier Alcolea Velázquez, Felipe Álvarez Arrieta, Lara Moreno Iglesias. 2007.** *Generador del Modelo Relacional y Esquemas de BD a partir del modelo Entidad / Relación.* Madrid : Universidad Complutense de Madrid, 2007.
- ✓ **Lapuente., María Jesús Lamarca. 2011.** Hipertexto: El nuevo concepto de documento en la cultura de la imagen. Tesis doctoral. *Tesauros.* [En línea] Universidad Complutense de Madrid, 19 de noviembre de 2011. <http://www.hipertexto.info>.
- ✓ **Morato, Jorge. 2009.** Glossarium-BITri. [En línea] 20 de abril de 2009. <http://www.tesauro.com/>.
- ✓ **Pérez-Luño, Antonio Enrique. 2006.** *Ensayos de Informática Jurídica. Biblioteca de Ética, Filosofía del Derecho y Política.* C.V. Coyoacán, México. : Ediciones Coyoacán, S.A., 2006.
- ✓ **Valdes, Julio Téllez. 2005.** *III. Informática Jurídica: Documentaria, Control y Gestión y Metadocumentaria.* Mexico : s.n., 2005.
- ✓ **Varona, Yandy Peñate Peña y Sulivan Lominchar de. 2011.** Análisis del Sistema de Información y diseño de una base de datos documental para la Unión Nacional de Juristas de Cuba. *Análisis del Sistema de Información y diseño de una base de datos documental para la Unión Nacional de Juristas de Cuba.* Habana : UCI, 2011.
- ✓ **Yunta, Luis Rodríguez. 2001.** *Bases de datos documentales: estructura y principios de uso.* Maldonado, Angeles : CINDOC-CSIC, 2001.
- ✓ **Danysoft. 2006.** Desarrollo |Diseño |Infraestructura |Ofimática |Artículos |Servicios Profesionales. [En línea] julio de 2006. <http://www.danysoft.com/free/0607rev.pdf>.
- ✓ **Ian H. Witten, Stefan Boddie y Jhon Thompson. 2010.** Biblioteca digital GREENSTONE. [En línea] Dpto. de Informática. Universidad de Waikato, Nueva Zelanda, 2010. http://www.greenstone.org/manuals/gsd12/es/html/User_es_index.html.
- ✓ **MySQL. 2011.** MySQL The world's most popular open source database. *MySQL 5.0 Reference Manual.* [En línea] Copyright ©1997, 2011. <http://dev.mysql.com/doc/refman/5.0/es/index.html>.
- ✓ **Oracle. 2011.** Sitio Oficial de Oracle. [En línea] febrero de 2011. http://www.oracle.com/solutions/business_intelligence/feature_dw_leadership.html.
- ✓ **PremiumSoft. 1999-2012.** Navicat. The gateway to your database. [En línea] PremiumSoft CyberTech Ltd., 1999-2012. <http://www.navicat.com/>. Copyright © 1999-2012 .PremiumSoft™
- ✓ **ProCite. 2005.** *Guía ProCite® 5.0.* Madrid : Social Sciences Citation Index y Arts & Humanities Citation Index"., 2005. 206-220..

- ✓ **S.A., Micronet. 2011.** Knosys Blue. *Knosys Blue, una nueva solución para la gestión de la información y el incremento de la productividad.* San Sebastián de los Reyes Madrid. : s.n., 2011.
- ✓ **Server, Microsoft SQL. 2012.** Microsoft SQL Server. *Microsoft SQL Server.* [En línea] Microsoft Corporation, 2012. <http://www.microsoft.com/es-xl/sqlserver/default.aspx>.
- ✓ **Simon Riggs, Hannu Krosing. 2010.** *PostgreSQL 9 Administration Cookbook.* BIRMINGHAM - MUMBAI : Packt Publishing Ltd., 2010. ISBN 978-1-849510-28-8.
- ✓ **SQLManager.net. 1999-2012 .** EMS Database Management Solutions. [En línea] Solutions, Inc. , 1999-2012 . Copyrigh © 1999-2012. All rights reserved.

Bibliografía Consultada

- ✓ **bombadil. 2009.** Bosque ViejoBase de Datos Relacionales: SQLite, MySQL y PostgreSQL. [En línea] 2009. <http://bosqueviejo.net/2009/01/07/base-de-datos-relacionales-sqlite-mysql-y-postgresql/>.
- ✓ **BRUERA, A. G. Y. O. M. 1982.** *Análisis del lenguaje jurídico.* Buenos Aires Argentina : Belgrano, 1982. ISBN 950.077.030.
- ✓ **Carrasco, Lic. Lourdes María. 2006.** Revista Jurídica JUSTICIA Y DERECHO. *III Encuentro Internacional Justicia y Derecho 2006 Mejor preparados para administrar justicia.* Ciudad de la habana : s.n., 2006. No. 7.
- ✓ **Codina, Lluís. abril, 1995.** *Metodología de creación de bases de datos docuemntales.* Barcelona : España, abril, 1995. ISBN: 1386-6710.
- ✓ *Desarrollo de la gestión documental.* **Mexicana, Cámara Nacional de la Indutria Editorial. 2001.** Naucalpan de Juárez : Pearson Educación de México, S.A de C.V, 2001, Vol. 7ma Edición. 1031.
- ✓ **FERNÁNDEZ, J. M. I. A. Y. L. M. M. 1994.** *Listas de encabezamientos de materia y Thesauro en perspectiva comparada.* . s.l. : Universidad de Murcia, 1994.
- ✓ **LAS TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN Y LA GESTIÓN DEL CONOCIMIENTO EN EL SECTOR SALUD.** **Susana Beatriz Llanusa Ruiz, Nereida Rojo Pérez, Magali Caraballosa Hernández, Roberto Capote Mir y Julia Pérez. 2005.** 2005, Rev Cubana Salud Pública.

- ✓ **abril, 2009.** *Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia, descriptores y tesauros.* abril, 2009.
- ✓ **MÁCIÁ, M. 1995.** *El tesoro EUROVOC.* . Madrid : Universidad Complutense. , 1995.
- ✓ **O., Prof. Pedro Luis Hippolyte. 2011.** *Base de Datos Documental: Textos, Gráficos, Videos y sonido para consulta on-line desde internet.* Caracas : Facultad de Arquitectura y Urbanismo. Universidad Central de Venezuela, 2011.
- ✓ **2010.** ONE Oficina Nacional de Estadísticas. [En línea] 2010. www.one.cu.
- ✓ **Pérez-Luño, Antonio Enrique. 2006.** *Ensayos de Informática Jurídica. Biblioteca de Ética, Filosofía del Derecho y Política.* C.V. Coyoacán, México. : Ediciones Coyoacán, S.A., 2006.
- ✓ **PLACES, M. Á. S. 2003.** *Arquitectura para Federación de Bases de Datos Documentales basada en Ontologías.* Coruña : DEPARTAMENTO DE COMPUTACIÓN. UNIVERSIDADE DA CORUÑA, 2003.
- ✓ **2003.** Radio Rebelde. [En línea] 2003. www.radiorebelde.cu.
- ✓ **Rincón, María Inés Arias de. 2009.** *Consideraciones sobre la Informática Jurídica Documental.* Zulia : Universidad del Zulia, 2009. B:1990.36.
- ✓ **Rodríguez., A.D. 2009.** *Diseño e implementación de la base de datos del sistema de planificación de entrenamiento deportivo de judo femenino.* Ciudad Habana : UCI, 2009.
- ✓ **Salomón, Omar Pérez. 2005.** Cuba Socialista. [En línea] 2005. www.cubasocialista.cu.
- ✓ *Sistemas Gestores de Bases de Datos. México, Universidad Autónoma de.* 2007. Universidad Autónoma de México : Revista de Posgrado, 2007.
- ✓ **Tabla Comparativa entre Endnote, ProCite y Reference Manager. 2005..** [softwarecientifico.com. http://www.softwarecientifico.com/paginas/comparativa.htm](http://www.softwarecientifico.com/paginas/comparativa.htm).
- ✓ **Torres., Charles Edward Bedón Cortázar. Andrés Felipe Arboleda. 2004.** *Sistema de Detección de Intrusos Utilizando Inteligencia Artificial.* s.l. : Universidad del Cauca Fac de Ing Electrónica y Telecomunicaciones, 2004.
- ✓ **Tramullas, Dr. Jesús. 2010.** *Aplicaciones informáticas para tratamiento documental. Sistemas electrónicos de tratamiento documental.* Panamá : s.n., 2010.

Anexos

Tablas del Diccionarios de Datos

Tabla T_Localizacion:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
PK_Id_Localizacion	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
Lugar	Específico del lugar donde se dictó la sentencia.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Fecha	Específico de la fecha en que dictó la sentencia.	Date	Si		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

Tabla T_Jueces:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
Pk_Id_Jueces	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
Nombre	Nombre del Juez	Alfanumérico	Si	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no

					se adicionará a la base de datos.
Apellidos	Apellidos del Juez	Alfanumérico	Si	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

Tabla **T_Resultando**:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
Pk_Id_Resultando	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
FK_Id_Sentencia2	Identificativo de la tabla T_Sentencia_Penal como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Sentencia_Penal .
Nombre	Nombre del Resultando, ej. Primer Resultando, Segundo Resultando, etc.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Descripcion	Descripción o cuerpo del resultando.	Texto	Si	Lenguaje Libre	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Orden	Orden de resultando	Integer	No		No puede quedar como campo en blanco, debe

	dentro de la sentencia, valor entre uno y cuatro.				validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
--	---	--	--	--	---

Tabla **T_Considerando**:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
Pk_Id_Considerando	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
FK_Id_Sentencia1	Identificativo de la tabla T_Sentencia_Penal como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Sentencia_Penal .
Nombre	Nombre del Considerando , ej. Primer Considerando , Segundo Considerando , etc.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Descripcion	Descripción o cuerpo del considerando.	Texto	Si	Lenguaje Libre	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Orden	Orden de considerando dentro de la sentencia, valor entre uno y cinco.	Integer	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en

					blanco, la sentencia no se adicionará a la base de datos.
--	--	--	--	--	---

Tabla T_Tipo_Sentencia_Penal:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
PK_Id_Tipo_Sentencia	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
Nombre	Nombre del tipo de la sentencia, ejemplo: Sentencia administrativa, Sentencia civil, Sentencia Económica, etc.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Descripcion	Descripción del tipo de Sentencia.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

Tabla T_Tipo_Tribunal:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
PK_Id_TipoTribunal	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
Nombre	Nombre del tipo de tribunal, ejemplo: Tribunal Supremo, etc.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no

					se adicionará a la base de datos.
Descripcion	Descripción del tipo de Tribunal.	Alfanumérico	No	Lenguaje Libre	Puede quedar en blanco, campo no obligatorio.

Tabla T_Ubicacion_Fisica:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
PK_Id_Ubicacion	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
FK_Id_Sentencia	Identificativo de la tabla T_Sentencia_Penal como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Sentencia_Penal .
Seccion	Específico de la ubicación de la sentencia en el centro.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Estante	Específico de la ubicación de la sentencia en el centro.	Alfanumérico	No	Lenguaje Controlado	No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.
Nivel	Específico de la ubicación de la sentencia en el centro.	Numérico	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

					de datos.
Numero	Específico de la ubicación de la sentencia en el centro.	Numérico	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

Tabla **T_Sentencia_T_Jueces**:

Atributo o Etiqueta	Dominio	Tipo de Datos	Indexación	Tratamiento documental	Controles de Validación
Pk_Id_Jueces_T_Sentencia	Número identificativo del registro.	Integer	No		No puede quedar como dato vacío, se genera y se autoincrementa internamente de forma automática.
T_Sentencia_PenalPK_Id_Sentencia	Identificativo de la tabla T_Sentencia_Penal como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Sentencia_Penal .
T_JuecesPK_Id_Jueces	Identificativo de la tabla T_Jueces como llave foránea en esta tabla.	Integer	No		Se obtiene de la tabla T_Jueces .
Ponente	Define si el Jueces ponente en esa Sentencia	Bit	No		No puede quedar como campo en blanco, debe validarse la entrada del mismo, para que nunca llegue en blanco, en caso de que llegue en blanco, la sentencia no se adicionará a la base de datos.

Imágenes de las pruebas de carga

Para la realización de la prueba funcional de carga primeramente se creó un grupo de hilos donde se definieron 50 usuarios que realizarían peticiones al servidor. A continuación se muestran las imágenes que describen el desarrollo de esta prueba:

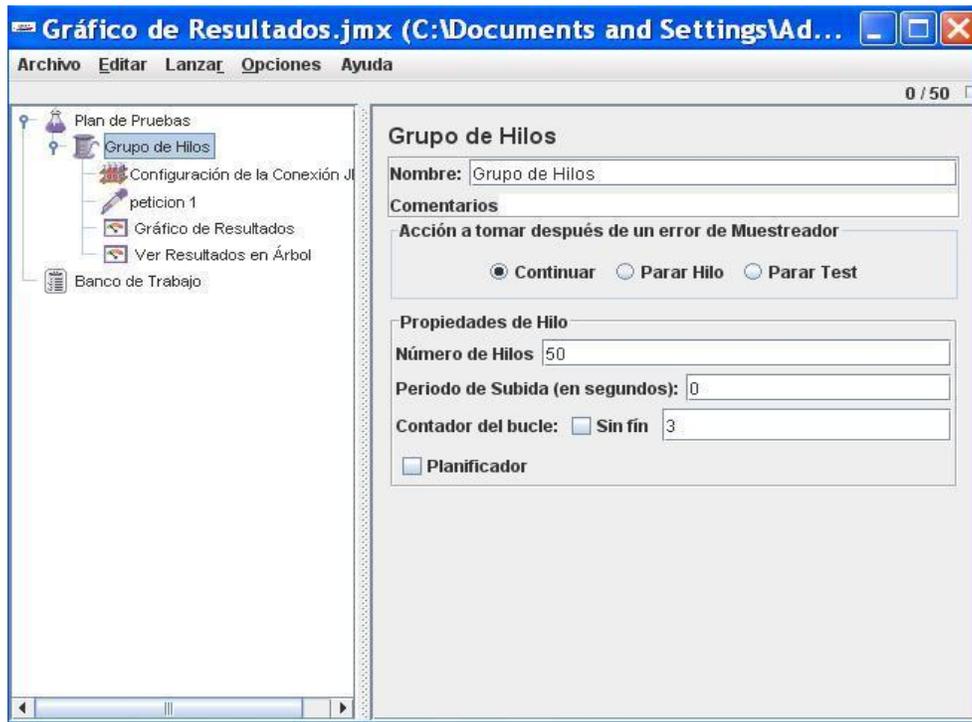


Figura. Configuración de Grupo de Hilos.

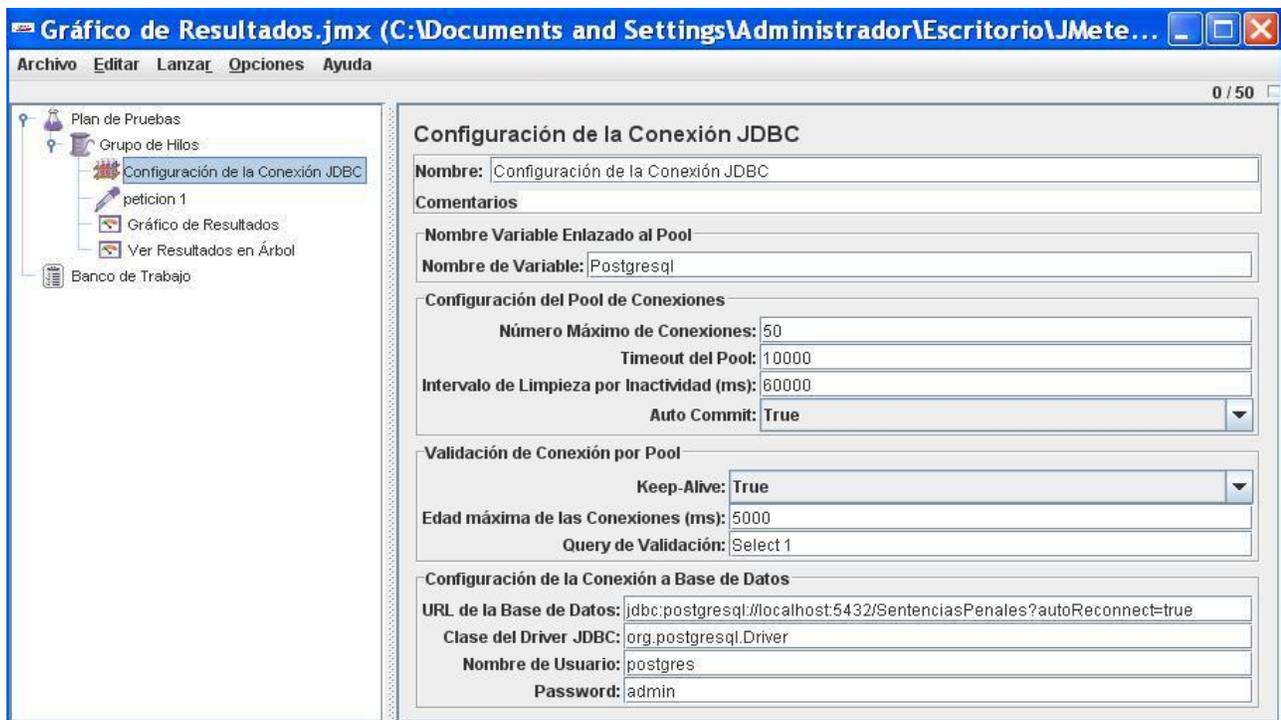


Figura. Configuración de la Conexión JDBC.

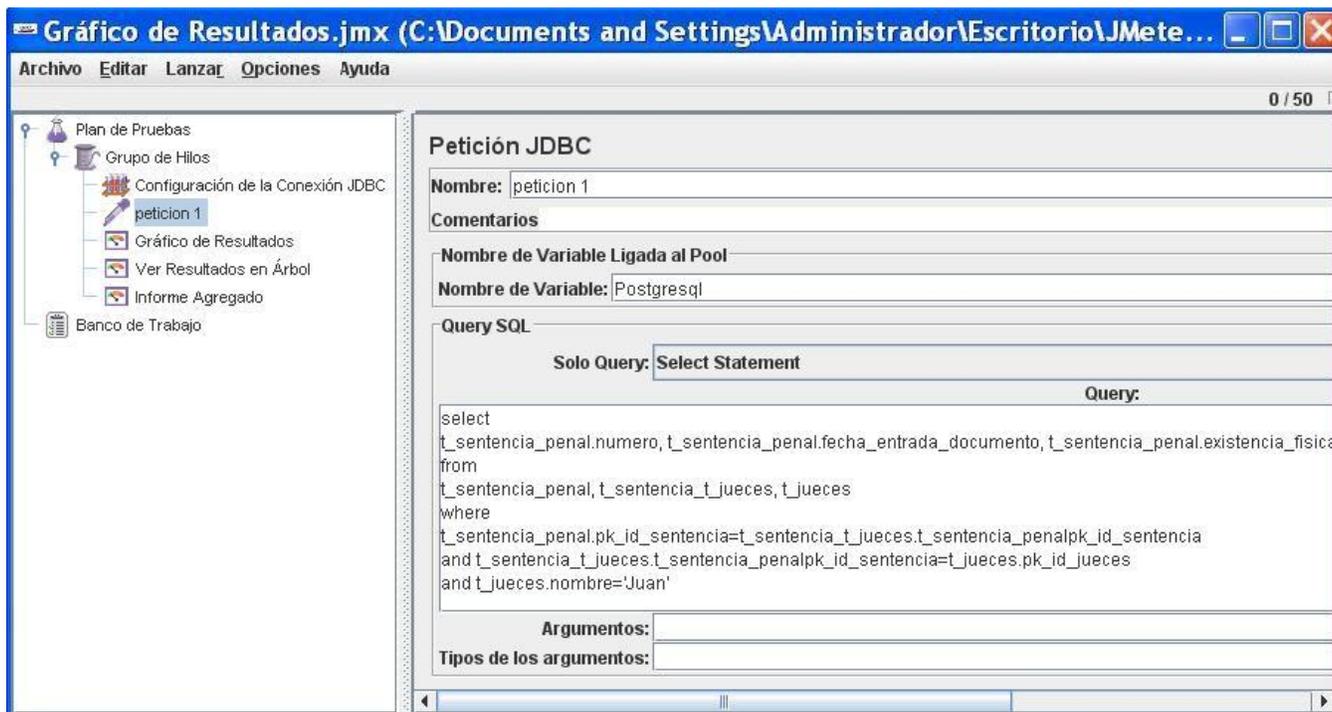


Figura. Petición JDBC realizada a la BD.

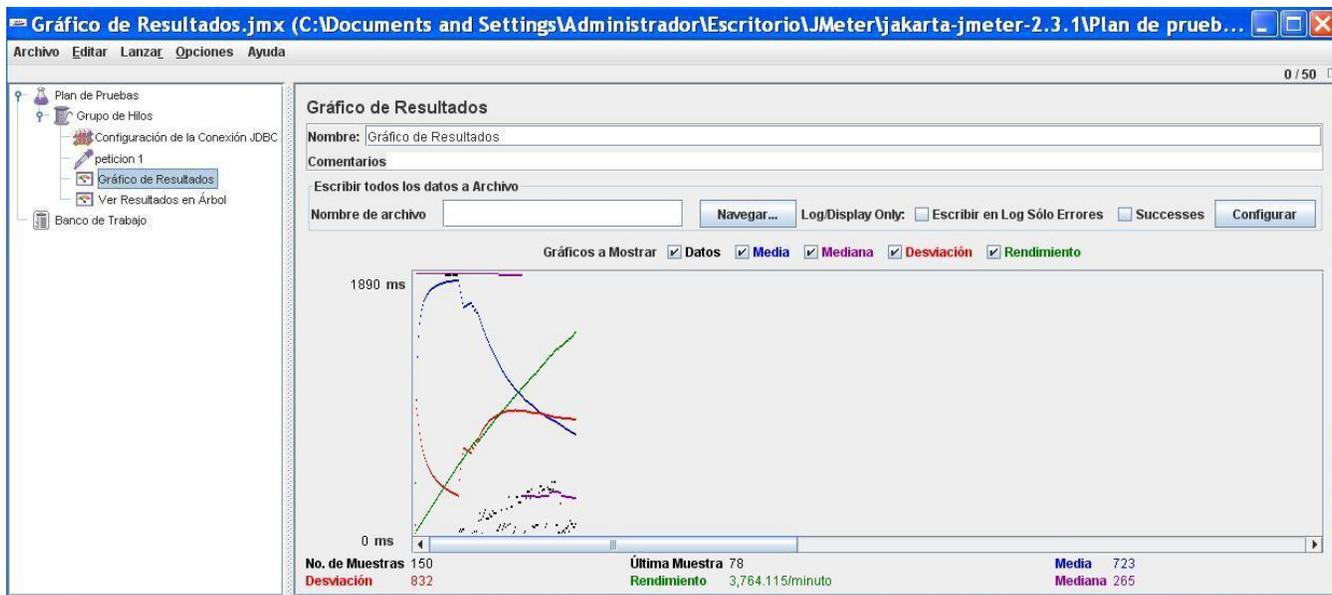


Figura. Gráfico de resultados.