Universidad Politécnica del Estado de Morelos



**Sistematización del departamento de gestión tecnológica y vinculación del CENIDET**

Sistema Gestor de Documentos

**T E S I N A**

Que para obtener el título de:

**INGENIERO EN INFORMÁTICA**

P r e s e n t a

**Fraga Cabrera Luis David**

Directores de Tesina

 **INTERNO**

**M. en A. Liliana Márquez Mundo.**

Jiutepec, Morelos Enero de 2011

**LISTA DE FIGURAS**

**LISTA DE TABLAS**

**RESUMEN**

CAPÍTULO 1: INTRODUCCIÓN

**1.1 Antecedentes**

El CENIDET ofrece sus programas de posgrado para egresados de licenciaturas afines que estén interesados en prepararse para la investigación aplicada y el desarrollo tecnológico. Estos programas están estructurados para que el alumno adquiera conocimientos y habilidades que le permitan desarrollarse con éxito en campos afines. El Centro ofrece la posibilidad de realizar trabajos de investigación en acuerdos con otras instituciones como el Instituto de Investigaciones Eléctricas y el Centro de Investigación en Energía; existen además acuerdos con importantes universidades e institutos extranjeros, como la Universidad de Oviedo de Española, la Universidad de Concepción de Chile, la Universidad Politécnica de Madrid, la Universidad Federal de Santa Catarina en Brasil, el Instituto Politécnico de Virginia y la Universidad de Texas A&M. El Centro ofrece la posibilidad de realizar trabajos de investigación en acuerdos con otras instituciones como el Instituto de Investigaciones Eléctricas y el Centro de Investigación en Energía; existen además acuerdos con importantes universidades e institutos extranjeros, como la Universidad de Oviedo de Española, la Universidad de Concepción de Chile, la Universidad Politécnica de Madrid, la Universidad Federal de Santa Catarina en Brasil, el Instituto Politécnico de Virginia y la Universidad de Texas A&M.

**1.2 Planteamiento del problema.**

Actualmente no se cuenta con un sistema que administre los documentos expedidos y recibidos en el Departamento de Gestión Tecnológica y Vinculación (DGTyV). Debido a la gran cantidad de documentos es complicado el manejo estos de manera manual generando así retrasos, perdidas y demás ineficiencias.

**1.3 Panorama General del Proyecto.**

El DGTyV con base en el manual de procedimientos, al hacer uso o prestación de algún servicio, se genera un acuerdo entre las instituciones involucradas.

Previo al desarrollo del acuerdo, se realizan una serie de reuniones para definir los aspectos del acuerdo. En cada reunión es generada una “minuta”, en este documento se plasma lo acordado en dicha reunión y se establecen responsabilidades.

A través del tiempo se han generado una serie de acuerdos con diversas instituciones y la consulta a ellos se torna cada vez más compleja.

La implementación de este sistema facilitará la consulta de dichos documentos, el sistema integrara el uso de XML para la creación de un árbol estructurado de cada documento, independientemente se generará una base de datos que almacene una serie de palabras claves que describan generalmente el contenido del documento.

El uso de un diccionario, nos permitirá ligar los criterios de búsqueda del usuario con palabras o expresiones que se relacionen con dichos criterios, de este modo ayudar a generar búsquedas más precisas.

**1.4 Objetivo General**

Generar un sistema capaz de gestionar los acuerdos y minutas del departamento, mediante uso de tecnologías como las búsquedas semánticas, uso de bases de texto y bases de datos que nos permitan controlar el manejo de la información contenida en estos documentos

1.5 Justificación.

El proyecto tendrá como función principal la administración de acuerdos y minutas mediante métodos de gestión (en algunos semánticos), que permita el manejo eficiente y facilite la consulta y explotación de la información que estos incluyen.

Este proyecto ayudará con la sistematización del departamento en general ya que es parte del proyecto de informatización del DGTyV.

1.6 Alcances y Limitaciones.

Derivado de un análisis integral al DGTyV, así como su relación con el CENIDET, las empresas y los institutos de investigación, se define el siguiente alcance que tendrá el sistema.

-Generar un programa de búsquedas dentro de contenido de los documentos del DGTyV, tales como:

* Acuerdos: convenio realizados con empresas e instituciones públicas y privadas.
* Minutas.
* Oficios.
* Actas
* Constancias
* Convocatorias
* Formatos

Dentro de esta primera etapa se realizara el sistema para el manejo de “acuerdos” y “minutas”, manejando una programación modular para que en etapas posteriores puedan integrarse el resto de los documentos.

* 1. Definiciones, Acrónimos y abreviaturas.

 *CENIDET*: Centro Nacional de Investigación y Desarrollo Tecnológico

*UPEMOR*: Universidad Politécnica del Estado de Morelos

*DGTyV*: Dirección de Gestión Tecnológica y Vinculación

*Empresas*: empresas que cuentan con convenio con el Cenidet.

*Institutos de Investigación*: institutos de Investigación que han realizado convenios con el Cenidet.

**Sistematización**: conjunto de fenómenos, elementos o ideas que se ordenan de forma racional y jerárquica con referencia a uno o varios principios generales

**Acuerdo**: texto que contiene lo acordado entre dos o más partes (instituciones)

**Automatización**: conjunto de métodos y procedimientos para la substitución del operario en tareas físicas y mentales previamente programadas

**Informatizar**: aplicar los métodos de la informática en un negocio, un proyecto, etc.

* 1. Organización de la tesina.

En este documento se describirán los principales aspectos y funcionalidades del sistema.

El documento cuenta con tres secciones en las que se describen, detalladamente las funciones con las que el sistema contará:

* En la primera sección describe una introducción del sistema y se proporciona una visión general de los recursos del sistema.
* En la segunda sección describe el contexto del sistema, las principales funciones, condiciones y restricciones del sistema.

En la última sección describe las construcciones físicas, desarrollo, durabilidad, adaptabilidad y condiciones ambientales.

**CAPITULO 2: MARCO TEORICO**

**2.1 Gestor de Documentos**

**Gestión documental es un “Conjunto de actividades administrativas y técnicas tendientes a la planificación, manejo y organización de la documentación producida y recibida por las entidades, desde su origen hasta su destino final, con el objeto de facilitar su utilización y conservación”** (1)**.**

Uno de los mayores problemas a los que se enfrentan las grandes corporaciones empresariales es la dificultad de compartir y conservar en la empresa el conocimiento generado por las sinergias del negocio, debido a la dispersión y duplicación de la documentación empresarial. Ello conlleva un descenso de la productividad, que se traduce en una pérdida de competitividad de la empresa en el mercado.

El gestor de documentos es un sistema que permite almacenar, buscar, gestionar, compartir y controlar los documentos corporativos de una forma práctica y sencilla

(2).

* + 1. Principales Gestores de Documentos en el mercado.
* SWAG: es un sistema que permite almacenar, buscar, gestionar, compartir y controlar los documentos corporativos de una forma práctica y sencilla.
* KnowledgeTree™ : es un Sistema de Gestión Documental -DMS- (Document Management System) de código fuente abierto y basado en entorno Web que a lo largo de los años ha ido mostrando su solvencia y cubriendo las más altas expectativas en la gestión documental en todas las organizaciones que ha sido implementada frente a otras soluciones propietarias mucho más costosas y complejas.
* Documentos MSD es una completísima herramienta de gestión de documentos y correspondencia que incluye un sistema avanzado de gestión de datos.

Consta de dos módulos principales, uno de documentos y otro de registro de correspondencias

* 1. Uso de Herramientas de Indexación.

Para conseguir que un usuario que introduce una palabra o frase en un buscador obtenga unos resultados, es necesario previamente indexar la información en los buscadores (introducir y preparar las estructuras de búsqueda), y por otro lado buscar la información a través de los índices generados en la indexación, también llamados índices de texto invertidos.(3)

2.2.1 Índices de texto invertidos

La información que recogen las arañas queda almacenada en los índices. A partir de esos índices, los buscadores generan los índices de texto invertidos para que las páginas queden clasificadas a partir de las palabras más significativas que se encuentran en éstas. De esta forma, cuando el buscador realiza la búsqueda de un término, accede directamente a estos índices de texto invertidos donde puede localizar las páginas en las que se encuentra el término solicitado.(3)

* 1. Uso de XML

XML es un formato basado en texto para describir datos. Incluye nuevas formas de distribuir datos a aplicaciones propias de Web, en lo que permite que casi cualquier tipo de dato se envíe a cualquier lugar. Esto suena como algún tipo de varita mágica, pero la utilidad de XML se centra en dos sencillas características. Primero, XML es extensible, lo que significa que puede extenderlo con facilidad con sus propias etiquetas y estructura. segundo, está basado en texto, lo que significa que usted podrá generar o leer XML con cualquier editor de textos.

Tal como HTML, XML utiliza simples etiquetas de marcado para describir su contenido. A diferencia de HTML, no hay etiquetas estándar.(4)

* 1. Lectura y Escritura de Archivos

2.4.1 Lectura de Archivos de Word con Jakarta POI - API Java

El proyecto POI consiste en APIs para manipular varios formatos de ficheros basados en el formato de Documento Compuesto OLE 2 de Microsoft, utilizando Java puro. En concreto, se pueden leer y escribir ficheros MS Excel utilizando Java. Pronto se podrá leer y escribir ficheros Word utilizando Java. POI es su solución Java Excel así como su solución Java Word. En cualquier caso, tenemos un API completo para portar otros formatos de Documento Compuesto OLE 2 y todo aquel que quiera participar será bienvenido.

Entre los ficheros basados en el formato de Documento Compuesto OLE 2 de Microsoft se incluyen la mayor parte de los ficheros de Microsoft Office tales como XLS y DOC así como formatos de fichero basados en el API de serialización MFC.

Como regla general se intentará colaborar lo más posible con otros proyectos para proporcionar esta funcionalidad. Algunos ejemplos: Cocoon para el que hay serializadores para HSSF; Open Office.org con quienes colaborará en la documentación del formato XLS; y Lucene para el que pronto se tendrá intérpretes del formato de fichero. Cuando es práctico, donamos componentes directamente a los proyectos para dotarles de capacidad-POI.(5)

* + 1. Lectura de XML

JDOM es un API para leer, crear y manipular documentos XML de una manera sencilla y muy intuitiva para cualquier programador en Java, en contra de otras APIs tales como DOM y SAX, las cuales se idearon sin pensar en ningún lenguaje en concreto, de ahí que resulte un poco incomodo su utilización.

la API JDOM no es un parser, de hecho, usa un parser para su trabajo, JDOM "solo" nos aporta una capa de abstracción en el tratado de documentos XML facilitándonos bastante la tarea como veremos enseguida, de hecho no tendremos que ser unos gurús de DOM y SAX para poder trabajar con XML desde Java.(6)

* + 1. Escritura de XML

Jaxen es un software libre de la librería XPath escrito en java. Es adaptable a muchos diferentes modelos de objetos, incluyendo DOM, XOM, dom4J y JDOM.(7)

CAPITULO 3: **DESCRIPCIÓN GENERAL DEL SISTEMA**

El sistema se encargará de la gestión y manejo de los documentos que recibe el DGTyV.

El sistema realizará búsquedas dentro del contenido del documento, mediante el uso de métodos de indexación, y diccionarios de datos que ayuden al usuario realizar búsquedas más precisas.

3.1 Contexto del sistema

El sistema deberá sistematizar el manejo de los documentos, reduciendo tiempos de búsquedas y mejorando el control sobre estos.

Dentro de esta primera etapa el sistema se basara sólo en la administración de acuerdos y minutas, en etapas posteriores se integraran los demás documentos.

3.2 Modos y estados del sistema

Dentro del sistema habrá diferentes estados para cada usuario del DGTyV.

El modo administrador es el encargado de brindar el mantenimiento, llevar un control de los usuarios, gestión de los documentos contenidos en el sistema (alta y baja).

El modo usuario solo podrá visualizar la información contenida así como la realización de búsquedas que le faciliten la obtención de esta.

3.3 Principales capacidades del sistema

 Dentro de las principales capacidades con las que el sistema deberá contar se encuentran:

 -Búsquedas dentro del contenido de los documentos.

 -Importación de archivos.

 -Generación de Reportes

En la Figura 3.1 se muestra gráficamente las principales actividades en el sistema.



**Figura 3.1**

3.4 Principales condiciones del sistema

 El sistema debe cumplir las siguientes características:

* *Acceso Ubicuo*: que cuente con sincronía.
* *Acceso Universal*: que cualquier persona con permisos pueda acceder a él.
* *Exhaustivo*: que cuente con todos los elementos necesarios para su buen funcionamiento.
* *Actualizable*: deberá tener la capacidad de mantener al día la información.
* *Automatizado*: deberá contar con la capacidad de aplicar la automática en los procesos que realice.

3.5 Principales restricciones del sistema

* El desarrollo de cada etapa del sistema deberá cumplir con los lineamientos y estándares que rigen al CENIDET.
* El uso de los lenguajes de programación está limitado a los servicios activos en el servidor.
* El diseño del sitio web debe manejar la plantilla oficial.

3.6 Características de los usuarios

El sistema será desarrollado para usuarios con conocimientos básicos en computación e internet. Para buen uso del sistema los usuarios deberán saber leer, escribir básicamente.

El personal que labora dentro de este departamento cuenta con al menos una maestría o doctorado, teniendo conocimiento en el manejo de cómputo e internet.

3.7 Suposiciones y dependencias

* Los documentos deberán mantener el formato oficial.
* Antes de usar el sistema el usuario deberá estar registrado previamente.

3.8 Escenarios de operación

Pantalla principal del usuario: en este escenario el usuario podrá introducir parámetros de búsqueda y visualizar el resultado.

Pantalla principal del administrador: en esta sección el administrador podrá configurar y dar mantenimiento al sistema.

CAPITULO 4: Capacidades, condiciones y restricciones del sistema

4. Capacidades, condiciones y restricciones del sistema

Capacidades del sistema son aquellas tareas que el sistema es capaz de realizar y no obstante deben cumplirse una serie de condiciones para el buen funcionamiento.

4.1 Condiciones físicas

4.1.1 Construcción

El sistema será basado en una arquitectura web brindando el acceso a él de forma remota y garantizando la disponibilidad de los datos.

4.1.2 Durabilidad

El sistema dependerá de los lineamientos y reglas establecidas en el instituto, deberá ser susceptible a cambios, actualizaciones y/o modificación en caso de ser necesario.

4.1.3 Adaptabilidad

 El diccionario de datos manejado por el sistema podrá ampliarse para mejorar el rendimiento y los tiempos de búsqueda.

4.1.4 Condiciones ambientales

El sistema estará diseñado para el DGTyV, el personal que labora dentro de este departamento cuenta con alguna maestría o doctorado, teniendo conocimiento en el uso de cómputo e internet.

El personal cuenta con una amplia experiencia en el área, ya que han estado en el cargo durante un periodo superior a los cinco años y en algunos casos han ocupado cargos dentro de otros departamentos de la institución.

Cada área del departamento cuenta con al menos un equipo de cómputo con conexión a internet.

4.1.5 Restricciones del sistema

Debido a que el CENIDET contrata un servicio de hospedaje para su sitio web, el uso de algunas tecnologías o funciones está sujeto a la disponibilidad de los servicios disponibles. Por ello el sistema se adaptara a los servicios soportados por el servidor.

4.2 Características de desempeño del sistema

El sistema por ser basado en una plataforma web, permite la conexión de múltiples usuarios realizando diversas tareas simultáneamente.

4.3 Seguridad

El uso de portales web genera una serie de riesgos que se deben contemplar para evitar fallos en la seguridad y garantizar la integridad de la información, las medidas de seguridad contempladas para el desarrollo del sistema son las siguientes:

* El acceso al sistema deberá estar restringido mediante una validación de datos del usuario. Esta contará con diferentes privilegios o funciones dentro del sistema, las cuales serán establecidas con base en las necesidades de cada usuario.
* El uso de contraseñas tendrá una encriptación, esto con el fin de asegurar la confidencialidad de las mismas al momento de establecer la comunicación con el servidor.
* El sistema contará con distintos niveles de usuarios, cada usuario contará con ciertos privilegios dentro del sistema y no podrá hacer uso de aquellas funciones en las que no esté autorizado.

4.4 Administración de la información

El sistema manejará la información mediante bases de datos, bases de texto, XML y metadatos que aproximen a las búsquedas semánticas.

4.5 Operaciones del sistema

El sistema deberá contemplar aspectos de mantenibilidad para facilitar el manejo a los usuarios.

4.5.1 Factores humanos

El sistema será manipulado por los empleados del DGTyV, contará con un administrador general.

El sistema contará con un usuario general: que podrá hacer uso de la información contenida como realizar búsquedas y visualizar los documentos.

El administrador general se encargará del mantenimiento del sistema, importar, eliminar, consultar y generar reportes de los documentos.

4.5.2 Mantenibilidad

El sistema será elaborado por módulos para futuros cambios y/o adaptaciones dando una mejor organización y estandarización de la programación en nuestro sistema.

Contará con un manual técnico especificando el uso de las funciones y/o módulos generados en el sistema.

4.5.3 Confiabilidad

El sistema contará con las principales medidas de seguridad, para la prevención de ataques y protección del sistema.

Estará disponible las 24hrs del día funcionando en su totalidad y podrá ser usado remotamente.

El sistema estará bajo las normas, reglamentos y procedimientos del CENIDET, garantizando la fiabilidad de los datos.

4.6 Políticas y regulaciones

Las políticas y regulación estarán sujetas a las establecidas por el CENIDET.

4.7 Ciclo de vida del sistema

El *modelo de ciclo de vida de objetos* es utilizado para describir como un objeto cambia de estados en el tiempo y que eventos producen dichos cambios de estado.

RUP (Racional Unified Process). Proceso Unificado Racional se caracteriza por ser iterativo e incremental, estar centrado en la arquitectura y guiado por los casos de uso, emplearemos esta metodología ya que las características y requerimientos de nuestro sistema se acoplan en gran forma a esta metodología por sus etapas de desarrollo y las características con las que cuenta dicha metodología

El RUP divide el proceso de desarrollo en ciclos, teniendo un producto final al culminar cada una de ellos, estos a la vez se dividen en fases y donde se debe tomar una decisión importante:

- **Concepción**: se hace un plan de fases, se identifican los principales casos de uso y se identifican los riesgos

- **Elaboración**: se hace un plan de proyecto, se completan los casos de uso y se eliminan los riesgos

- **Construcción**: se concentra en la elaboración de un producto totalmente operativo y eficiente y el manual de usuario

- **Transición**: se Instala el producto en el cliente y se entrena a los usuarios. Surgen nuevos requisitos a ser analizados.

- **Mantenimiento**: una vez instalado el producto, el usuario realiza requerimientos de ajuste, esto se hace de acuerdo a solicitudes generadas como consecuencia del interactuar con el producto.

**Características de RUP**

- Forma disciplinada de asignar tareas y responsabilidades (quién hace qué, cuándo y cómo).

- Pretende implementar las mejores prácticas en Ingeniería de Software.

- Desarrollo iterativo.

- Administración de requisitos.

- Uso de arquitectura basada en componentes.

- Control de cambios.

- Modelado visual del software.

- Verificación de la calidad del software.