

Universidad de las Ciencias Informáticas

Facultad 1



Título: Módulo para la clasificación automática de documentos en la Biblioteca digital
“Alma Mater”

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Autor: Daniel Enrique Fernández Lorenzo

Tutores: Ing. Nairys Morales Sosa
Ing. Adrian Antonio Garbey Fonseca

Junio del 2012

Declaración de Autoría

Declaro ser el autor de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los ____ días del mes de ____ del año 2012.

Daniel Enrique Fernández Lorenzo

Ing. Nairys Morales Sosa

Ing. Adrián Antonio Garbey Fonseca

A mis padres por su gran apoyo y dedicación en todo momento.

A mi novia Iyugnis por su gran apoyo y amor dado en estos 5 años.

A mi tía Elina por preocuparse tanto por mí.

A mis tutores Adrián y Nairys por toda su guía y esfuerzo.

A mis amigos, mis compañeros, mis hermanos en estos 5 años inolvidables de mi vida que siempre estaban conmigo en las buenas y en las malas Teymor, Dayron, Remy, Lachy, Fernando, "El hijo", "Rega" y todos los del grupo de los M.

A mis amigos de Camagüey que siempre puedo contar con ellos para lo que haga falta Agustín, Rudy, Jorge, Marlon... muchas gracias a todos por su amistad.

A todas las personas que me ayudaron de una forma u otra a confrontar los problemas que se me presentaban.

A todas las personas queridas... que son muchas... muchas gracias a todos....

Les dedico esta tesis especialmente a mis padres ya que sin ellos no estuviera donde estoy ahora...para ti mamá y papá....

...y a mi novia y a todos mis amigos...

Índice

Índice de Figuras	VIII
Índice de Tablas	IX
Resumen	XI
Introducción	12
Capítulo 1: Fundamentación Teórica	15
1.1 Introducción.....	15
1.2 Biblioteca Digital.....	15
1.3 Clasificación de documentos.	17
1.3.1 Tipos de clasificación.....	17
1.3.1.1 Clasificación supervisada o Categorización.....	18
1.4 Sistemas de clasificación de documentos.	19
1.4.1 Partes de un Sistema de Clasificación.	19
1.4.2 Ejemplos de Sistemas de Clasificación de documento.....	20
1.4.3 Sistema de Clasificación Dewey	20
1.5 Clasificación automática de textos o documentos.	20
1.5.1 Software de clasificación automática de documentos. Ejemplos.....	21
1.6 Los Sistemas de Gestión de Contenidos (CMS).....	25
1.6.1 Drupal.	26
1.7 Tecnologías asociadas a Drupal.....	27
1.7.1 Del lado del cliente.....	27
1.7.2 Del lado del servidor.	29
1.8 Gestor de Base datos (SGBD).	30
1.9 Entorno de Desarrollo Integrado.....	30
1.10 Herramienta Case.	30
1.11 Patrón arquitectónico Modelo Vista Controlador (MVC).....	31
1.12 Metodología de desarrollo. Rational Unified Process (RUP).....	31
Conclusiones del capítulo.....	32

Capítulo 2: Descripción del Sistema	33
2.1 Introducción.....	33
2.2 Modelo de negocio.	33
2.2.1 Reglas del negocio.	33
2.2.2 Actores y trabajadores del negocio.	33
2.2.3 Diagrama de casos de uso del negocio.	34
2.2.4 Descripción de los casos de uso del negocio.....	35
2.2.5 Diagrama de actividades.....	35
2.2.6 Modelo de objetos.....	35
2.3 Especificación de Requisitos.	36
2.3.1 Requisitos Funcionales (RF).	36
2.3.2 Requisitos no funcionales (RNF).....	37
2.4 Descripción del sistema propuesto.	38
2.4.1 Actores del sistema.....	39
2.4.2 Diagrama de casos de uso del sistema.....	39
2.4.3 Descripción de los Requisitos Funcionales.	40
2.5 Modelo de análisis.....	50
2.5.1 Diagrama de clases del análisis.....	51
2.6 Arquitectura de Drupal.....	51
2.6.1 Patrones de diseño en Drupal.....	52
2.7 Modelo de diseño.	53
2.7.1 Diagrama de clase de diseño web.	53
2.7.2 Diseño de la Base de datos.	54
2.7.3 Diagrama de clases persistentes.	55
2.7.4 Modelo de datos.	58
2.7.5 Descripción de las tablas de base de datos.	59
Conclusiones del capítulo.	59
Capítulo 3: Implementación y validación del sistema	60
3.1 Introducción.....	60

3.2 Diagrama de despliegue.....	60
3.3 Pruebas.....	61
3.3.1 Pruebas de caja negra.....	61
3.3.2 Resultados de las pruebas internas.....	67
Conclusiones del capítulo.....	67
Conclusiones.....	68
Recomendaciones.....	69
Referencias Bibliográficas.....	70
Bibliografía.....	72
Anexos.....	74

Índice de Figuras

Figura 1: Diagrama de casos de uso del negocio.	34
Figura 3: Modelo de objeto del negocio.	35
Figura 2: Diagrama de actividades del caso de uso Clasificar documentos.	36
Figura 4: Diagrama de casos de uso del sistema.	40
Figura 5: Diagrama de clases del análisis Clasificar documentos.	51
Figura 6: Diagrama de clases del diseño general del CMS Drupal.	55
Figura 7: Diagrama de clases del diseño Clasificar documentos.	56
Figura 8: Diagrama de clases persistente del módulo de clasificación.	57
Figura 9: Diagrama del Modelo de datos del módulo de clasificación.	58
Figura 10: Diagrama de Despliegue del módulo de clasificación.	60
Figura 11: Diagrama de clases del análisis Gestionar clasificación.	74
Figura 12: Diagrama de clases del análisis Configurar módulo.	74
Figura 13: Diagrama de clases del análisis Mostrar documentos clasificados.	75
Figura 14: Diagrama de clases del análisis Mostrar documentos sin clasificar.	75
Figura 15: Diagrama de clases del análisis Generar formulario.	75
Figura 16: Diagrama de clases del análisis Generar base de clasificación.	75
Figura 17: Diagrama de clases del diseño Configurar módulo.	76
Figura 18: Diagrama de clases del diseño Mostrar documentos sin clasificar.	77
Figura 19: Diagrama de clases del diseño Mostrar documentos clasificados.	78
Figura 20: Diagrama de clases del diseño Generar textos de entrenamiento.	79
Figura 21: Diagrama de clases del diseño Gestionar clasificación.	80

Índice de Tablas

Tabla 1: Actores del negocio.	34
Tabla 2: Trabajadores del negocio.	34
Tabla 3: Descripción del caso de uso del negocio Clasificar documentos.	35
Tabla 4: Actores del sistema.	39
Tabla 5: Descripción del RF Clasificar documentos.....	41
Tabla 6: Descripción del RF Generar formulario.....	42
Tabla 7: Descripción del RF Adicionar clasificación.....	43
Tabla 8: Descripción del RF Modificar clasificación.	44
Tabla 9: Descripción del RF Eliminar clasificación.....	45
Tabla 10: Descripción del RF Mostrar clasificación.	46
Tabla 11: Descripción del RF Mostrar documentos clasificados.	47
Tabla 12: Descripción del RF Mostrar documentos sin clasificar.	48
Tabla 13: Descripción del RF Generar base de clasificación.	49
Tabla 14: Descripción del RF Configurar módulo.	50
Tabla 15: Descripción de la tabla de base de datos base_de_conocimiento.	59
Tabla 16: Descripción del caso de prueba para el RF Adicionar clasificación.....	63
Tabla 17: Descripción de variable para el RF Adicionar clasificación.	63
Tabla 18: Descripción del juego a probar para el RF Adicionar clasificación.	65
Tabla 19: Descripción del Registro de defectos y dificultades detectadas para el RF Adicionar clasificación.....	67
Tabla 20: Descripción de la tabla de base de datos relación_doc_clas.	81
Tabla 21: Descripción de la tabla de base de datos selección_tipo_contenido.....	81
Tabla 22: Descripción de la tabla de base de datos node.....	82

Tabla 23: Descripción de la tabla de base de datos node_revisions.....	83
Tabla 24: Descripción de la tabla de base de datos node_type.....	84
Tabla 25: Descripción del caso de prueba para el requisito funcional Modificar clasificación.	86
Tabla 26: Descripción de las variables para el requisito funcional Modificar clasificación.....	86
Tabla 27: Descripción del juego a probar para el requisito funcional Adicionar clasificación.	89
Tabla 28: Descripción del registro de defectos y dificultades detectadas para el requisito funcional Modificar clasificación.....	90
Tabla 29: Descripción del caso de prueba para el requisito funcional Configurar módulo.....	92
Tabla 30: Descripción de las variables para el requisito funcional Configurar módulo.	92
Tabla 31: Descripción del juego a probar para el requisito funcional Configurar módulo.	94
Tabla 32: Descripción del registro de defectos y dificultades detectadas para el requisito funcional Configurar Módulo.....	95

Resumen

En la actualidad el proceso de clasificación de documentos en las bibliotecas ha evolucionado significativamente con la introducción de las nuevas Tecnologías de la Información y las Comunicaciones (TIC). Todo este cambio ha forzado a los bibliotecarios a cambiar los métodos con los que tradicionalmente trabajaban, valiéndose de nuevos procedimientos y técnicas automáticas que elevan la calidad con que sea realiza este proceso.

Tomando lo anteriormente expresado como punto de partida, el presente trabajo aborda aspectos teóricos relacionados con la clasificación y específicamente la clasificación automática de los documentos. Además se desarrolla un módulo para la Biblioteca digital “Alma Mater” de la República Bolivariana de Venezuela, el cual garantiza funciones como: clasificar documentos, generar formulario de clasificación, gestionar clasificación, mostrar documentos clasificados, mostrar documentos sin clasificar y otras funcionalidades que permiten un conjunto de facilidades para los bibliotecarios.

Palabras clave:

proceso, clasificación automática, biblioteca digital, tecnologías de la información.

Introducción

Desde tiempos remotos los sucesos más importantes ocurridos en la historia se han documentado en lugares donde mucho tiempo después las personas puedan informarse de lo sucedido, estos lugares son conocidos como bibliotecas. La biblioteca hoy en día se ha convertido en una institución cuya finalidad consiste en la adquisición, conservación, estudio y exposición de libros y documentos (ESPAÑOLA, 2012).

Con el surgimiento de la informática se ha encontrado una vía para almacenar, compartir, y gestionar grandes colecciones de datos digitalizados. Más adelante con la aparición de la Internet y en especial por la aparición de la World Wide Web¹ aparecieron repositorios de documentos de todo tipo generados por diferentes instituciones, repositorios que aumentaron considerablemente y cambiaron notablemente su visualización, adoptando interfaces cada vez más sencillas de utilizar. Se caracterizaron por incorporar elementos como: texto, imágenes, audio y videos.

Estos avances dieron paso al surgimiento de las bibliotecas digitales que según definiciones son una colección estructurada y organizada de documentos digitales, desarrollada según una política y un esquema conceptual, que ofrece a sus usuarios servicios de valor añadido, fundamentados precisamente en la colección, o en aspectos relacionados con la misma (SaZ, 2005).

Una biblioteca digital presenta varias ventajas como permitir: funcionamiento las 24 horas del día, permitiendo el acceso a distancia, simultáneo de un gran número de personas, a personas con dificultades físicas, a materiales que pueden ser manipulados y distribuidos en formatos que la versión original no permitiría, entre muchas otras.

Las bibliotecas digitales utilizan sistemas informáticos que ayudan en la mejora y ejecución de sus procesos, un ejemplo de esto lo constituyen los sistemas para clasificación automática de documentos, los cuales facilitan el trabajo del bibliotecario y además, permiten que la recuperación de la información sea más sencilla y rápida.

En la República Bolivariana de Venezuela se desarrolla la Biblioteca digital “Alma Mater”, la cual surge con el objetivo de que su documentación ayude a la formación de los universitarios así como a la adquisición de una mayor cultura general e integral. Esta biblioteca contiene un gran volumen de

¹ Expresión inglesa que se basa en un sistema de documentos de hipertexto que se encuentran enlazados entre sí y a los que se accede por medio de la Internet a través de un software conocido como navegador.

información en libros y documentos de diferentes tipos de categorías, además diariamente se indexan grandes volúmenes de documentación.

Actualmente para clasificar un documento el bibliotecario aunque lo hace de forma digital, se realiza manual, o sea documento a documento. Es lógico imaginar lo lento y engorroso que resulta este proceso. Esta situación provoca que se acumulen gran cantidad de libros, documentos y revistas, afectando además la calidad de las búsquedas e investigaciones.

Por esta razón se hace necesario implementar un módulo que se integre a la Biblioteca digital desarrollada, surgiendo el siguiente **problema a resolver**. ¿Cómo garantizar que el proceso de clasificación de documentos en la Biblioteca digital “Alma Mater” se realice de manera sencilla y rápida?

Se define como **objeto de estudio** el proceso de clasificación automática de documentos y como **campo de acción** la clasificación automática de documentos en la Biblioteca digital “Alma Mater”.

Se plantea como **objetivo general**: Desarrollar un módulo que permita la clasificación automática de documentos en la Biblioteca digital "Alma Mater" de manera sencilla y rápida. Para cumplir con este objetivo se definen como **objetivos específicos**:

- Realizar el marco teórico de la investigación.
- Describir las tecnologías, herramientas y metodologías empleadas en el desarrollo del trabajo.
- Realizar el análisis y diseño del módulo clasificación automática de documentos.
- Implementar las funcionalidades de la solución propuesta.
- Validar las funcionalidades del módulo.

Se plantea como **idea a defender** que: con el desarrollo de un módulo de clasificación automática de documentos se garantizaría que el proceso en la Biblioteca digital “Alma Mater” se realice de manera sencilla y rápida.

Métodos teóricos:

- Analítico–sintético: Posibilitando procesar toda la información enfocada hacia la investigación sobre la clasificación de los documentos y específicamente la clasificación automática.
- Histórico–lógico: Para conocer los antecedentes y tendencias actuales en la clasificación automática de los documentos, software que se emplean, las plataformas y lenguajes de programación.

Métodos empíricos:

- La observación: Se utilizó para comprobar el estado real de la clasificación de documentos en la Biblioteca digital "Alma Mater" de la República Bolivariana de Venezuela así como la necesidad de automatizar este proceso.

Como **posible resultado** se obtendrá un módulo de clasificación automática de documentos que permitirá que este proceso sea en la Biblioteca digital "Alma Mater" de la República Bolivariana de Venezuela, sencillo y rápido.

La investigación consta de una introducción, tres capítulos que incluyen todo lo relacionado con el tema investigativo, la solución propuesta, y la validación de la misma, estructurados de manera organizada para un buen entendimiento del lector, conclusiones generales, recomendaciones, bibliografía, referencias bibliográficas utilizadas, y por último los anexos que complementan el desarrollo del trabajo.

Capítulo 1: Fundamentación Teórica.

Se abordarán los conceptos asociados al problema antes expuesto, se realiza un estudio del estado del arte sobre los sistemas informáticos que se utilizan para la clasificación automática de documentos así como las herramientas y metodologías a utilizar.

Capítulo 2: Descripción del Sistema.

Se propone la solución del problema a resolver, describiendo las principales actividades y artefactos que se generan en las fases de modelamiento del negocio, requerimiento, y análisis y diseño.

Capítulo 3: Implementación y validación del sistema.

Se implementan las funcionalidades y se valida la solución aplicando el método de caja negra. Se hace un análisis del resultado obtenido.

Capítulo 1: Fundamentación Teórica

1.1 Introducción.

El campo de las bibliotecas digitales cada vez toma mayor importancia dentro de las áreas de la información, estas representan una interesante combinación entre aplicaciones, sistemas y antiguas teorías tradicionales del manejo de información ordenada y estructurada. Asimismo, representan amplias oportunidades para el campo de la educación al extender y mejorar el acceso a nuevas formas de acercarnos al conocimiento, especialmente en áreas de ciencia y tecnología.

La clasificación automática de documentos en las bibliotecas es de gran importancia ya que agiliza los procesos internos de la misma al permitir ordenar por temas los documentos y colecciones que en ella existen, facilitando así su recuperación por el usuario y consultar directa y libremente sin restricciones.

En el presente capítulo se realiza un estudio del estado del arte de los software para la clasificación automática de documentos, fundamentando la selección de uno de ellos. Se brinda información y conceptos de las bibliotecas digitales, así como tipos y sistemas de clasificación de documentos, además se describe la metodología y herramientas a utilizar para el desarrollo del trabajo, las cuales fueron definidas por el proyecto.

1.2 Biblioteca Digital.

Es de vital importancia para el desarrollo de un país, acercar las fuentes de conocimiento con quienes generan avances tecnológicos, científicos, económicos y políticos. Las bibliotecas son el pilar de la educación de todas las comunidades y el mayor apoyo de conocimiento para los estudiantes de nivel superior, docentes e investigadores.

El desarrollo de bibliotecas digitales es un gran paso de avance al abrir una nueva puerta para la adquisición de conocimientos mediante las tecnologías de la información, y con los conceptos que han perdurado por siglos en el área bibliotecológica. Por esta razón el desarrollo de las mismas ha alcanzado gran auge en la actualidad.

Una biblioteca digital es una colección estructurada y organizada de documentos digitales, desarrollada según una política y un esquema conceptual, que ofrece a sus usuarios servicios de valor añadido, fundamentados precisamente en la colección, o en aspectos relacionados con la misma (SaZ, 2005).

Es un conjunto de recursos informativos electrónicos, organizados en colecciones de acuerdo a la temática que abarque dicha biblioteca y que además brinda al usuario las herramientas y los servicios que lo asistan en el manejo de la información de manera automatizada (Márquez, 2006).

Según las definiciones antes vistas se puede generalizar que una biblioteca digital no es más que un conjunto de documentos digitales organizados y estructurados en un sitio web, donde brinda al usuario la facilidad de investigar y manejar la información que posee la misma.

Funcionalmente, una biblioteca digital ofrece los servicios de búsqueda y recuperación de información. Los documentos que se encuentran en ella pueden ser texto, imágenes, sonido, video o combinaciones de cualquiera de estas. Idealmente se deben almacenar y recuperar documentos completos, y las búsquedas se realizan sobre el contenido completo de los documentos. Es decir, si el documento es texto, es posible realizar la búsqueda sobre cada palabra incluida en el documento completo y una vez localizado es posible obtenerlo de manera inmediata. En el caso de otros formatos como audio y video, la búsqueda se realiza sobre las fichas que definen las características de cada material (Guzmán, 2000).

Las bases de datos y la información pueden ser almacenadas de manera centralizada en caso de colecciones pequeñas. Aunque lo más común es contar con información distribuida, en este caso, las búsquedas y las consultas de un sistema a otro deberán ser transparentes para el usuario. La interfaz por excelencia es la Web, ya que ofrece una plataforma abierta, principalmente para proporcionar el acceso a los usuarios, aunque también es utilizada por los administradores de la biblioteca y por los grupos encargados de generar contenidos.

Las bibliotecas digitales proporcionan sus servicios a través de alguna red, ya sea una red sólo de uso local o tan abierta como Internet, esto dependerá de los servicios y las restricciones que deban hacerse para el acceso a la información. Los beneficios de implementar bibliotecas digitales se pueden dividir para tres sectores. Inicialmente, tienen *beneficios nacionales o globales*:

- Promueve y facilita la expansión de la cultura en una comunidad.
- Preserva en un medio no degradable los acervos culturales y científicos generados para el beneficio social.
- Se hace uso eficiente de los contenidos de los materiales a través de búsquedas sencillas y eficientes.
- Promueve el uso de estándares para el manejo de información digital, incluso a nivel mundial.

Los beneficios institucionales son:

- Elimina duplicidad de actividades, acervos y costos.
- Promueve nuevas áreas de investigación.
- Permite el crecimiento de acervos sin demanda de espacio físico para almacenamiento ni para servicio.
- Prolonga la permanencia de documentos dentro de una colección disponible al público.
- Control total sobre la información.
- Reducción del costo de imprenta, para aquellas que editan sus propios libros.

Los beneficios al usuario:

- Confianza en el contenido de los documentos que se consultan.
- Acceso uniforme desde cualquier punto de la red sin desplazamiento a la biblioteca.
- Equidad en el acceso a la información, toda la comunidad autorizada tiene la misma facilidad y derecho de consultarla.
- Siempre habrá disposición de los materiales existentes que sean solicitados.
- Ahorro en el tiempo de búsqueda por la centralización aparente de los acervos y por la posibilidad de búsqueda en el contenido completo de los documentos.
- Acceso a información interrelacionada, es decir, posibilidad de enlaces hipertextuales, incluso con otros recursos de la biblioteca digital (Guzmán, 2000).

1.3 Clasificación de documentos.

Se entiende por clasificación de documentos al arte de asignar a los libros un lugar exacto en un sistema de clasificación en el cual, los distintos asuntos del saber o la descripción de la vida humana en sus varios aspectos están agrupados de acuerdo con sus semejanzas o con las relaciones de uno con otros (Acosta, 2003).

Es la acción de ordenar por clases, agrupar por ramas del conocimiento, problemas, materias o características formales de los documentos (Sánchez y Vera, 2008).

Según las definiciones anteriores se puede decir que la clasificación de documentos es el arte de asignar los documentos a un lugar exacto en un sistema, ordenándolos tanto por clases, problemas, materias como características formales de los documentos.

1.3.1 Tipos de clasificación.

Existen dos tipos de clasificación, la Supervisada y no Supervisada. Estos escenarios reciben diversos nombres, pero básicamente consisten en lo siguiente: la Supervisada es una situación en la que se parte de una serie de clases o categorías conceptuales prediseñadas a priori, y en la que la labor del clasificador (manual o automático) es asignar cada documento a la clase o categoría que le corresponda. Este escenario requiere la elaboración manual o intelectual del cuadro o esquema de categorías, además de un proceso de aprendizaje o entrenamiento por parte del clasificador, que debe ser supervisado manualmente en mayor o menor medida.

En la clasificación no Supervisada no hay categorías previas ni esquemas o cuadros de clasificación establecidos a priori. Los documentos se agrupan en función de ellos mismos, de su contenido; de alguna manera, se puede decir que se auto-organizan (Figuerola, Berrocal, Angel F. Zazo Rodríguez, Emilio Rodríguez, 2004).

Luego de realizar una investigación sobre cómo funcionan los procesos internos y el desarrollo de la Biblioteca digital “Alma Mater” se pudo corroborar que actualmente se utiliza el Sistema de Clasificación Dewey. Dicho sistema utiliza una serie de clases y categorías conceptuales diseñadas a priori, y por consiguiente el tipo de clasificación empleado es la Clasificación Supervisada. La actividad de clasificar documentos se realiza de forma digital pero manual.

1.3.1.1 Clasificación supervisada o Categorización.

En esta modalidad como se ha explicado anteriormente se parte de una serie de clases o categorías prediseñadas a priori, en las cuales hay que colocar a cada uno de los documentos. Conocida también como categorización, facilita la recuperación ad-hoc², limitando las búsquedas a las clases o categorías que el usuario elige, basándose en el conocimiento intelectual que previamente tiene del tema; pero también se aplica en cosas como el filtrado automático de documentos, el clásico problema del routing³ de documentos o la navegación autónoma en la web.

Aunque hay gran cantidad de algoritmos capaces de hacer clasificación supervisada, la idea o enfoque básico es muy parecido: conseguir, de alguna manera, construir un patrón representativo de cada una de

² Es la tarea más representativa en la que se basan los buscadores web. El conjunto de documentos sobre el que se realizan las consultas permanece estable donde nuevas consultas procedentes de los usuarios llegan al sistema de forma continua.

³ Consiste en servir a una serie de clientes ubicados geográficamente de manera dispersa, y para atender a estos clientes se parte desde un punto inicial, se le asigna una ruta a cada cliente minimizando el coste de tiempo a cada uno de ellos.

las clases o categorías; y aplicar alguna función que permita estimar el parecido o similitud entre el documento a clasificar y cada uno de los patrones de las categorías. El patrón más parecido al documento es el que nos indica a qué clase se debe asignar ese documento.

Para la construcción de los patrones de las categorías se utilizan documentos clasificados manualmente de antemano, que sirven como ejemplo. El proceso de formar esos patrones de cada clase a partir de esos documentos preclasificados se conoce como entrenamiento o aprendizaje; y la colección de documentos preclasificados utilizada se conoce como colección de entrenamiento (Figuerola, Berrocal, Angel F. Zazo Rodríguez, Emilio Rodríguez, 2004).

1.4 Sistemas de clasificación de documentos.

Los sistemas de clasificación son sistemas de catalogación por materias pre-coordinadas y estructura jerárquica, son lenguajes de recuperación de la información. El propósito de cualquier sistema de clasificación en una biblioteca es agrupar de una forma lógica los recursos por disciplinas en común. Dentro de sus características fundamentales se encuentran: sistemáticos, sencillos y flexibles porque van de lo general a lo particular, la notación debe ser fácil de escribir y recordar, y que permitan la combinación de ideas y puntos de vista, respectivamente (Figuerola, Berrocal, Zazo Rodríguez, Emilio Rodríguez, 2004).

1.4.1 Partes de un Sistema de Clasificación.

- Base ideológica: La clasificación de los conocimientos supone una toma de postura ideológica, un punto de vista desde el cual se enfoca la realidad.
- Base teórica: Sirve para explicar el sistema y como guía de utilización.
- Sistema de notación: Puede ser numérico, alfabético y alfanumérico, o mixto, ya que ningún sistema de clasificación es absolutamente puro.
- Posibilidad de expresar materias compuestas: Es la posibilidad de análisis y de síntesis.
- Esquema básico: La estructuración en tablas principales, tablas auxiliares y los índices.
- Microorden: Se trata de una serie de especificaciones interiores para hacer las subdivisiones dentro de cada clase.
- Índice alfabético: Para facilitar el acceso a las notificaciones ya que la estructura principal es sistemática. (Acosta, 2003).

1.4.2 Ejemplos de Sistemas de Clasificación de documento.

Existen en el mundo varios sistemas dedicados a la clasificación de documentos entre ellos se pueden encontrar:

- Sistema de Clasificación de Documento de la Biblioteca del Congreso (LCC).
- Sistema de Clasificación Dewey.
- Bliss Classification (BC).
- Colon Classification (CC).
- Clasificación bibliográfica para bibliotecas públicas de la URSS (BBC).

Como se explicó anteriormente el sistema previamente definido con la creación de la biblioteca es el Sistema de Clasificación Dewey.

1.4.3 Sistema de Clasificación Dewey

Melvin Dewey era bibliotecario en Amherst Collage en Massachusetts cuando tuvo la idea de crear un sistema de clasificación que respondiera a las necesidades de la biblioteca de su colegio. La notación de este sistema se basa en números árabes. Estos tienen la ventaja de ser casi universales, a diferencia de las letras que tienen varios alfabetos, además de otras formas de representación. Una característica que lo distingue bastante es el número de notas que acompañan a las diferentes siglas. Se pueden distinguir cuatro categorías principales:

- Notas que indican el contenido de una clase.
- Notas que indican el emplazamiento de una cierta materia que carece de signo propio, bien porque hay inseguridad en su emplazamiento, bien porque la materia es demasiado específica.
- Notas que señalan el contenido de otras secciones.
- Notas que explican los cambios habidos en el sistema y en el esquema o en las tablas auxiliares.

Para lograr establecer la notación a un documento primeramente se debe realizar el análisis del contenido del documento, logrando así identificar los aspectos o materias que trata el documento. Utilizando el registro o índice relativo del sistema se buscan los signos correspondientes a estos términos de manera que facilite la búsqueda en las tablas y clases del sistema. Una vez seleccionado el índice principal y el número auxiliar necesario, se establece la notación del documento (Acosta, 2003).

1.5 Clasificación automática de textos o documentos.

La Clasificación automática de textos, también denominada Categorización de textos, es la tarea de asignar automáticamente un conjunto de documentos a una o más categorías preexistentes a través de un conjunto predefinido de documentos pre categorizados, sobre los que el sistema lleva a cabo un proceso de aprendizaje supervisado (MULTIDOC, 2008).

Por clasificación automática de documentos se entiende, en sentido amplio, un conjunto de algoritmos, técnicas y sistemas capaces de asignar un documento a una o más clases, construidos según su afinidad temática (Gastell y Ceruto, 2008).

Se podría definir como clasificación automática de documentos a la tarea de ordenar documentos de forma automática en un sistema utilizando un conjunto de algoritmos y técnicas capaces de asignar dichos documentos a una o más clases.

La clasificación automática de textos tiene sus orígenes en la recuperación de información y últimamente ha recibido más atención debido al incremento en la cantidad de información disponible en formato electrónico. Es por esta razón, que cada vez es mayor la necesidad de herramientas que ayuden a satisfacer las necesidades del usuario en cuanto a la información que busca, y además que pueda encontrarla en un tiempo adecuado.

El objetivo de la clasificación automática de textos es archivar documentos dentro de un número fijo de categorías predefinidas en función de su contenido. Un mismo documento puede pertenecer a una, varias, todas o ninguna de las categorías dadas. Se considera de gran importancia la automatización de este proceso en las bibliotecas ya que ayuda a los bibliotecarios a la hora de gestionar documentos, y los usuarios logran un rápido y efectivo acceso a la información que buscan (MULTIDOC, 2008).

1.5.1 Software de clasificación automática de documentos. Ejemplos.

Se podría definir que un software para la clasificación de documentos es aquel que posee la programación y características esenciales para clasificar documentos de forma automática. En la actualidad existe una gran variedad de software dedicados a esta importante actividad. A continuación se explican algunos de ellos.

Visual bibliotecas.

Programa sencillo de manejar, extremadamente rápido y potente, válido para gestionar grandes bibliotecas, es adaptable a las necesidades del usuario, contiene funcionalidades como: códigos de barras para los libros con impresión en impresora normal y posibilidad de modificar el diseño de

impresión, escaneo de documentos, listado de libros, posibilidad de exportar los informes a otros formatos como PDF, Word, Excel, permite guardar tantos documentos por libro como se quiera, catálogo de libros, clasificación por categorías y subcategorías de las obras de la biblioteca, configurables utilizando Dewey. Es un software propietario y está dirigido fundamentalmente a las bibliotecas tradicionales (Visual Bibliotecas, 2012).

Agente de recuperación de información Just-in-time (JITIR Agent).

Los agentes JITIR son aplicaciones que proactivamente recuperan y presentan información basada en un contexto local del usuario, reducen el esfuerzo mental requerido para encontrar, evaluar y acceder a la información que se necesita para incrementar el sentido de una información. Poseen un sistema de recuperación de información, Savant, el cual consiste en dos programas fundamentales: ra-retrieve y ra-index.

El programa ra-retrieve realiza la recuperación de la información basada en una consulta, y ra-index crea el índice de los archivos donde se realiza la búsqueda. Savant actúa como un artefacto de recuperación de información que realiza sus acciones por “detrás” del agente JITIR, dada una consulta se crea una lista de documentos que tienen relación con ella. Esta aplicación está principalmente dirigida a la recuperación de información en los centros de prensa (Marchena y Pérez, 2007).

Sistema de Recuperación de Información Geográfica para el motor de búsqueda Orión.

Este sistema cuenta con dos interfaces de usuarios, una interfaz de usuario de administración, destinada a realizar el mantenimiento de la estructura de indexación del sistema en general, y otra para realizar consultas al sistema e interactuar con los resultados.

El módulo de recuperación de información geográfica se encarga de recuperar el documento solicitado de la fuente de datos donde se encuentra almacenado. Para realizar una consulta el usuario interactúa con el sistema mediante la interfaz de usuario de consulta, esta interfaz le permite al usuario indicar las palabras claves a buscar en la consulta, y también navegar por el mapa para visualizar el área de interés para la consulta.

Esta navegación se presenta mediante la petición de mapa que, aunque aparece una única vez, se realiza para cada movimiento en el mapa y para cada ajuste del nivel zoom. Las peticiones al servidor de mapas provocan que este componente tenga que recuperar la información geográfica necesaria de las

fuentes de datos. Este sistema está enfocado a la búsqueda de información en la rama de la geografía (García, 2011).

Motor de Clasificación Inteligente de Contenidos (MOCIC).

MOCIC es un sistema que tiene como objetivo proporcionar una solución integral para la categorización, automática e inteligente de grandes volúmenes de información. Los contenidos a analizar podrán encontrarse en distintos formatos permitiendo al sistema adaptarse a diversos entornos, como las bibliotecas digitales (Rodríguez y Rios, 2009).

El sistema brinda a los administradores la posibilidad de auditar información, de forma automática, almacenada en un dispositivo o lugar en un sistema. Permitirá a un usuario común organizar su información personal, además de generar listas de URLs⁴ categorizadas abasteciendo a cualquier sistema de filtrado en especial a Smart Keeper⁵. Está desarrollado sobre tecnologías libres y cuenta con las funcionalidades que a continuación se describen:

- **Módulo Clasificador de Rostros:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador⁶ el cual indica la clasificación de las imágenes de una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el directorio de sus imágenes y le realizará un procesamiento para determinar el número de rostros de personas encontrados en cada imagen. Seguidamente se elabora un mensaje que se colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.
- **Módulo Clasificador de Desnudez:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador el cual indica la clasificación de las imágenes de una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el directorio de sus imágenes y le realizará un procesamiento para determinar por cada imagen la existencia o no de desnudez (presencia de piel humana). Seguidamente se elabora un

⁴ Secuencia de caracteres que se utiliza para nombrar y localizar recursos, documentos e imágenes en Internet.

⁵ Filtro cubano de contenido web desarrollado en la Universidad de las Ciencias Informáticas orientado a servidores para medianas redes de usuarios como las disponibles generalmente en colegios, empresas y bibliotecas.

⁶ Módulo Controlador de MOSIC que ejerce el control sobre dicho sistema.

mensaje que se colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.

- **Módulo Clasificador de Objetos:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador el cual indica la clasificación de las imágenes de una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el directorio de sus imágenes y le realizará un procesamiento para determinar por cada imagen el número de símbolos encontrados de cada categoría (Ciencias, Computadoras, Deporte, Juegos, Pornografía, Violencia). Seguidamente se elabora un mensaje que se colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.
- **Módulo Reconocimiento Óptico de Caracteres:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador el cual indica la clasificación de las imágenes de una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el directorio de sus imágenes y le realizará un procesamiento para extraer por cada imagen todo el texto asociado y elaborar un único fichero de texto a partir del cual, y utilizando los mismos algoritmos del clasificador de textos determinará las categorías de contenido asociadas a esta URL, elaborando un mensaje que colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.
- **Módulo Clasificador de Enlaces:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador el cual indica la clasificación de los enlaces relacionados a una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el fichero donde están depositados todos los enlaces correspondientes, le realizará un procesamiento para determinar las categorías de contenido asociadas a esta URL, elaborando un mensaje que colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.
- **Módulo Clasificador de Texto:** Este módulo tiene como función recibir a través de la consulta de un fichero de solicitudes, un mensaje proveniente del MOD-Controlador el cual indica la clasificación del texto de una URL. Para dicha clasificación accederá al directorio de localización de la URL, cargará el fichero .html correspondiente, le realizará un preprocesamiento para identificar el idioma al que

pertenece y seguidamente determinará las categorías de contenido previamente definidas asociadas a esta URL, elaborando un mensaje que colocará en un fichero de respuestas. Este mensaje será posteriormente utilizado por el Módulo Decisor, para decidir a qué categoría finalmente pertenece la URL.

- **Módulo Decisor:** Este módulo tiene como función recibir por parte del MOD-Controlador, toda la información proveniente de los módulos clasificadores y devolverle la categoría más probable a la que pertenece la URL.

Selección del sistema

A partir de todos los sistemas analizados anteriormente, se selecciona el sistema MOCIC por su adaptabilidad a los diversos entornos, específicamente a las bibliotecas digitales, posee las funcionalidades principales que se necesitan para un clasificador automático de documentos y además, está desarrollado sobre tecnologías libres.

El uso del software MOCIC en la presente investigación se centrará en el Módulo Clasificador de Texto. El módulo cuenta con algoritmos de clasificación automática de documentos que se usan hoy en día como el algoritmo de Rocchio, Vecino más próximo en su variante KNN, entre otros, los cuales funcionan a través de un sistema de aprendizaje previo. A continuación se explicará para un mejor entendimiento cómo se utilizará la funcionalidad Clasificar Texto:

El primer paso para realizar dicha tarea es crear una base de conocimiento con todas las categorías por las cuales en la biblioteca digital se categorizan los documentos. Como segundo paso, se crean una serie de textos de entrenamiento por cada categoría para que los algoritmos aprendan las distintas formas de clasificación de documentos que se rigen en la biblioteca digital. Es válido aclarar que estos dos pasos lo realizan los bibliotecarios con un nivel avanzado en la materia, de manera manual pero de forma digital. Al finalizar los pasos definidos se hace uso del sistema MOCIC para que clasifique los documentos. Dicha clasificación contiene un margen de error, que depende tanto de lo bien programado que esté el algoritmo como de la información que posean los textos de entrenamiento.

1.6 Los Sistemas de Gestión de Contenidos (CMS).

Se conocen como CMS (del inglés Content Management System) a las aplicaciones altamente configurables. Se utilizan principalmente para publicar, editar, eliminar, actualizar y administrar la información de un portal web sin conocimiento alguno de HTML, ya sea en Internet o en una intranet,

proporcionan la capacidad para gestionar la estructura del sitio y la navegación a los usuarios. Basada en un entorno cien por ciento web, la finalidad de un CMS es la creación de aplicaciones y sitios web dinámicos de forma rápida y eficiente (Batista y Rosario, 2010).

Los CMS permiten crear sitios web de forma sencilla, estos se encargan de la parte más engorrosa, la programación web y trabajo con la base de datos. Permiten que usuarios con pocos conocimientos del tema puedan gestionar sus páginas web. Actualmente existe un gran número de CMS que se pueden clasificar de la siguiente forma (Guervos, 2005).

- Portal: Es una aplicación web que posee contenidos y diversas funciones, se utiliza para brindar información. Ejemplo: Drupal, Plone, Joomla, GX Portal, entre otros.
- Foros: Aplicaciones web donde se les permite a los usuarios intercambiar opiniones de temas de interés. Ejemplo: SMF y phpBB.
- Blogs: Utilizados para publicar artículos y noticias ordenadas cronológicamente, contienen un espacio para realizar comentarios, por ejemplo Drupal, WordPress.
- Wikis: Sitios web de colaboración entre usuarios, brindan la posibilidad de crear y editar artículos. Ejemplo: TikiWiki y MediaWiki (Batista y Rosario, 2010).

1.6.1 Drupal.

Drupal es un CMS modular, gratuito y de código abierto para todos. El mismo permite crear páginas web de cualquier tipo, adaptándose a las necesidades particulares de cada caso (Mercer, 2008).

Presenta un conjunto de módulos con diversidad de funcionalidades, no solo en la descarga inicial, sino que además se puede descargar una amplia variedad de módulos del repositorio oficial. Presenta una total independencia con el tipo de base de datos. Posee una capa de abstracción de base de datos que actualmente está implementada para MySQL⁷ y PostgreSQL⁸, aunque permite adicionar con facilidad soporte a otras bases de datos (Guervos, 2005).

Entre el núcleo y los módulos contribuidos Drupal permite hacer prácticamente cualquier cosa. Si Drupal no tiene éxito en una función en especial, es probable que se integre con un servicio que pueda prestar la función para su utilización.

⁷ Es un sistema de gestión de bases de datos (SGBD, por sus siglas en inglés) multiusuario, multiplataforma y de código abierto.

⁸ Es un sistema de gestión de base de datos relacional orientada a objetos y libre, publicado bajo la licencia distribución de software Berkeley (BSD, por sus siglas en inglés).

Este CMS está publicado bajo la licencia GNU/GPL 2⁹, por tanto se puede descargar libremente, modificarlo y reutilizarlo. Esta es una de las características fundamentales que hacen que Drupal tenga una comunidad de desarrollo gigante, lo cual es una de sus principales ventajas. Es multiplataforma, puede funcionar como servidor web en los sistemas operativos Linux, Windows, Mac OS X, BSD y Solaris. Por otro lado al estar implementado en el lenguaje de programación PHP es totalmente portable (Palenzuela, 2010). Durante el desarrollo del módulo de clasificación se utilizará el CMS Drupal en su versión 6.x.

1.7 Tecnologías asociadas a Drupal.

Como sistema estructurado de contenidos, Drupal está soportado sobre una arquitectura cliente-servidor, cuya composición se describe en los siguientes apartados.

1.7.1 Del lado del cliente.

Las tecnologías del lado del cliente son utilizadas para acceder a los servicios, estas se encargan de gestionar la comunicación cliente-servidor, tal como solicitar un servicio concreto y recibir los datos enviados por el servidor. Son un conjunto de herramientas que presentan al usuario los datos en pantalla. A continuación se muestran las tecnologías utilizadas para desarrollar la aplicación.

➤ Lenguaje de marcado de hipertexto.

Su traducción al español sería Lenguaje de marcado de hipertexto (HTML, por sus siglas en inglés), no es más que un sistema de etiquetas las cuales tienen como objetivo estructurar el documento en forma de textos, vínculos, imágenes y otros elementos que compongan una página. Este lenguaje no presenta ningún compilador, su entorno para trabajar puede ser simplemente un editor de texto.

XHTML 1.0 fue el primer cambio importante desde el HTML 4.0 lanzado en 1997. Con el XHTML 1.0, el rigor de XML a las páginas web es la piedra angular en el trabajo del W3C¹⁰ para crear estándares que hacen más ricas las páginas web en una creciente gama de las plataformas de navegador, incluyendo teléfonos celulares, televisores, coches, etc. XHTML 1.0 fue el primer paso: la renueva HTML como una

⁹ La Licencia Pública General de GNU(GNU/GPL, por sus siglas en inglés)

¹⁰ El World Wide Web Consortium (W3C), es un consorcio internacional que produce recomendaciones para la World Wide Web.

aplicación XML. Esto hace que sea más fácil de procesar y de mantener. XHTML 1.0 toma prestados elementos y atributos de la anterior obra del W3C sobre HTML 4, y puede ser interpretado por los navegadores existentes, siguiendo unas simples directrices (Palenzuela, 2010).

➤ **Hojas de Estilo en Cascada (CSS).**

El CSS (del inglés Cascading Style Sheets) es un lenguaje de hojas de estilos creado para controlar el aspecto o presentación de los documentos electrónicos definidos con HTML y XML. El uso de CSS brinda muchas ventajas frente al HTML puesto que además de permitir dar estilo más depuradamente, permite definir estilos a determinados grupos de elementos.

CSS funciona a base de reglas, es decir, declaraciones sobre el estilo de uno o más elementos. Las hojas de estilo están compuestas por una o más de esas reglas aplicadas a un documento HTML. La regla tiene dos partes: un selector y la declaración. A su vez la declaración está compuesta por una propiedad y el valor que se le asigne (Palenzuela, 2010).

El selector funciona como enlace entre el documento y el estilo, especificando los elementos que se van a ver afectados por esa declaración. La declaración es la parte de la regla que establece cuál será el efecto. Se utilizará CSS en su versión 2.1 durante el desarrollo del módulo de clasificación.

➤ **JavaScript.**

El objetivo fundamental de JavaScript es la creación de páginas dinámicas que interactúen más con el usuario. Con JavaScript se puede incorporar a cualquier sitio web diferentes efectos para hacerlo más ameno. Este lenguaje es interpretado por lo que no es necesario compilarlo y se puede probar directamente en el navegador. Vale destacar que el mismo posee librerías con funciones ya implementadas que resuelven muchos problemas asociados con la Web (Palenzuela, 2010).

JavaScript es, con diferencia, el lenguaje de programación que más se utiliza en la Web. Hay publicadas millones de páginas que incorporan elementos que lo utilizan. La mayoría de los exploradores web pueden trabajar con JavaScript. Gracias a estos factores se puede asegurar que la mayoría de los usuarios de Internet podrán trabajar con JavaScript.

➤ **AJAX.**

El termino AJAX es un acrónimo de Asynchronous JavaScript + XML que se puede traducir como "JavaScript asíncrono + XML". AJAX no es una tecnología como tal, sino que es la mezcla de varias

tecnologías independientes que se unen de manera sorprendente (Palenzuela, 2010). Las tecnologías que la conforman son las siguientes:

- XHTML y CSS, para crear una presentación basada en estándares.
- Document Object Model (DOM), para la interacción y manipulación dinámica de la presentación.
- XML, XSLT y JSON, para el intercambio y la manipulación de información.
- XMLHttpRequest, para el intercambio asíncrono de información.
- JavaScript, para unir todas las demás tecnologías.

AJAX permite mejorar la interacción entre el usuario y el sistema además de evitar las posibles recargas constantes al sistema, ya que el intercambio entre la página y el servidor pasan a un segundo plano. La recarga constante de la página AJAX la elimina mediante la creación de un elemento intermedio entre el usuario y el servidor. Esta capa tiene como objetivo fundamental que el usuario nunca encuentre una ventana vacía esperando por la posible respuesta del servidor.

1.7.2 Del lado del servidor.

➤ **PHP.**

Hypertext Pre-processor (en español Procesador de Hipertextos). Es un lenguaje script interpretado en el lado del servidor para generar páginas web dinámicas, embebidas en páginas HTML (Palenzuela, 2010). Para que funcione el sitio, PHP necesita estar montado sobre un servidor tanto Apache como Servidor de Información de Internet (IIS, por sus siglas en inglés) con las librerías de PHP.

De manera general este lenguaje es fácil de aprender, se caracteriza por ser muy rápido, además de que presenta muchas funciones. Tiene una amplia capacidad de conexión con la mayoría de los manejadores de bases de datos MySQL, PostgreSQL, Oracle, MSSQL Server, entre otros. Es libre por lo que es una opción muy fácil para todos, es multiplataforma y tiene la capacidad de expandir su potencial utilizando módulos (Palenzuela, 2010). Se utilizará PHP en su versión 5.3 durante el desarrollo del módulo de clasificación.

➤ **Apache.**

Drupal requiere un servidor web, que pueda ejecutar scripts php. Apache es un servidor que además de ser flexible es rápido, eficiente, frecuentemente actualizado y adaptado a los nuevos protocolos (Guervos, 2005). En su versión 2 presenta, entre otras características, un elaborado índice de directorios y puede

ser adaptado a diferentes entornos y necesidades. Otras características son la negociación de contenidos, el informe de errores, la gestión de recursos para procesos hijos, la integración de imágenes del lado del servidor, la reescritura de las URL, la comprobación de la ortografía de las URL y manuales online (Meriño y Castro, 2011).

1.8 Gestor de Base datos (SGBD).

Un sistema gestor de base de datos es un programa que permite gestionar de manera clara y sencilla la información contenida en una base de datos (Guervos, 2005).

➤ **PostgreSQL.**

Es un Sistema gestor de base de datos relacional. Está considerado como el SGBD de código abierto más avanzado del mundo. Posee consultas SQL declarativas, control de concurrencia multiversión, soporte multiusuario, transacciones, optimización de consultas, herencia y arreglos son ejemplos de su avanzada funcionalidad. La API de acceso al SGBD se encuentra disponible en C, C++, Java, Perl, PHP, Python y TCL fundamentalmente. Cuenta con un rico conjunto de tipos de datos, permitiendo además su extensión mediante tipos y operadores definidos y programados por el usuario. Su administración se basa en usuarios y privilegios. Sus opciones de conectividad abarcan TCP/IP, sockets UNIX y sockets NT (Guervos, 2005). Para el desarrollo de la investigación se utilizará la versión 8.4.

1.9 Entorno de Desarrollo Integrado.

➤ **NetBeans.**

Es un IDE (del inglés Integrated Development Environment) multiplataforma de código abierto, pensado para escribir, compilar, depurar y ejecutar software. Sus funciones están provistas por módulos. Está escrito y pensado para Java, pero tiene soporte para lenguajes como C/C++, Ruby, Python, HTML, CSS y PHP. Posee un fuerte debugger¹¹ integrado para PHP 5 con soporte para el framework Symfony (Guervos, 2005). Durante el desarrollo del módulo de clasificación se utilizará el NetBeans en su versión 7.x.

1.10 Herramienta Case.

¹¹ Un depurador (en inglés, *debugger*), es un programa usado para probar y depurar (eliminar los errores) de otros programas (el programa "objetivo").

Las herramientas CASE (del inglés Computer Aided Software Engineering, Ingeniería de Software Asistida por Computadora) son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el costo de las mismas en términos de tiempo y de dinero. Estas herramientas nos pueden ayudar en todos los aspectos del ciclo de vida de desarrollo del software (wordpress, 2008).

➤ **Visual Paradigm.**

Es una herramienta CASE que soporta UML (del inglés Unified Modeling Language, Lenguaje de Modelado Unificado) la cual lleva el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue (Guervos, 2005). Tiene como ventaja que permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. También proporciona abundantes tutoriales, demostraciones interactivas y proyectos UML. Presenta licencia gratuita y comercial. Es fácil de instalar, actualizar y es compatible entre ediciones. Durante el desarrollo del módulo de clasificación se utilizará el Visual Paradigm en su versión 6.4.

1.11 Patrón arquitectónico Modelo Vista Controlador (MVC).

El MVC es un patrón de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario, y la lógica de negocio en tres componentes distintos. Se ve frecuentemente en aplicaciones web, donde la vista es la página HTML y el código que provee de datos dinámicos a la página. El modelo es el sistema de gestión de base de datos y la lógica de negocio, y el controlador es el responsable de recibir los eventos de entrada desde la vista (WIKIPEDIA, 2012b).

1.12 Metodología de desarrollo. Rational Unified Process (RUP).

RUP es una metodología de desarrollo de software muy utilizada que puede ser adaptada prácticamente a cualquier proyecto. El proceso de software que propone RUP presenta tres características esenciales:

- Dirigido por casos de uso.
- Centrado en la arquitectura.
- Iterativo e incremental.

RUP divide el proceso en cuatro fases, dentro de las mismas se realizan varias iteraciones en dependencia del proyecto y en las que se hace mayor o menor hincapié en las distintas actividades. Estas fases son:

- **Inicio:** Durante esta fase se define el alcance del proyecto y el modelo de negocio. Además se identifican los actores y los casos de uso, y se diseñan los casos de uso arquitectónicamente significativos. Por último se elabora un plan de negocio para determinar los recursos que deben ser asignados al proyecto.
- **Elaboración:** El objetivo de esta fase no es más que analizar el dominio del problema, establecer la base de la arquitectura, eliminar los mayores riesgos y desarrollar el plan del proyecto.
- **Construcción:** Durante esta fase todos los componentes, características y requisitos deben ser implementados, integrados y probados, obteniendo una versión del proyecto que sea aceptable. La finalidad principal de esta fase es lograr que el producto tenga una capacidad operacional de forma incremental a través de iteraciones.
- **Transición:** La fase de transición se encarga de entregar el producto a los usuarios finales. Además se debe completar la documentación, entrenar al usuario en el manejo del producto finalizado. Por lo general se ejecutan tareas de ajuste, configuración, instalación y facilidad de uso del producto.

En RUP se han agrupado las actividades en grupos lógicos definiéndose 9 flujos de trabajo principales. Los 6 primeros son conocidos como flujos de ingeniería y los tres últimos como de apoyo (Meriño y Castro, 2011).

Conclusiones del capítulo.

En el presente capítulo se abordaron diferentes conceptos, entre ellos los relacionados con las bibliotecas digitales, clasificación y clasificación automática de documentos, lo que permitió el entendimiento del área en la que se desarrolla la investigación.

Además se estudiaron y analizaron las principales características del software para la clasificación de documentos lo que permitió la selección adecuada para la ejecución del módulo que se propone. Se describieron los lenguajes, herramientas y metodología para el desarrollo del módulo.

Capítulo 2: Descripción del Sistema

2.1 Introducción.

Este capítulo tiene como propósito modelar artefactos que tienen lugar durante los flujos de trabajo: Modelamiento del negocio, Requerimientos y Análisis y diseño. Dentro de estos flujos de trabajo se modela el negocio, obteniendo actores, trabajadores y casos de uso. Se definen además los requisitos funcionales, se obtienen diagramas de actividades, diagramas de clases del análisis, los cuales se refinan en las clases del diseño. Además se diseña la base de datos, se conforma el diagrama de clases persistentes y el modelo de datos.

2.2 Modelo de negocio.

La Modelación del Negocio brinda una visión amplia de la organización permitiendo definir roles, responsabilidades y procesos. Al realizar el modelo de negocio se pretende:

- Comprender la dinámica y estructura de la organización.
- Comprender problemas que presente la organización y posibles mejoras que se pudieran realizar.
- Asegurar que los usuarios finales, desarrolladores y clientes converjan en una idea común.
- Obtener los requerimientos del sistema a partir del modelo de negocio.

2.2.1 Reglas del negocio.

- Solamente el bibliotecario podrá interactuar directamente con el módulo a la hora de clasificar los documentos, así como la gestión de clasificación.
- El administrador es el único usuario capaz de configurar el módulo en el sistema.

2.2.2 Actores y trabajadores del negocio.

Un actor del negocio es cualquier individuo, grupo, entidad, organización, máquina o sistema de información externos; con los que el negocio interactúa. Lo que se modela como actor es el rol que se juega cuando se interactúa con el negocio para beneficiarse de sus resultados (EcuRed, 2012a). Los actores del negocio interactúan con el mismo enviando y recibiendo mensajes, el papel que desempeñan está relacionado íntimamente con el o los procesos en los que se involucra.

Un trabajador del negocio representa un rol que juega una persona (o grupo de personas), una máquina o un sistema automatizado; actuando en el negocio. Son los que realizan las actividades, interactuando

con otros trabajadores del negocio y manipulando entidades (EcuRed, 2012a). Los nombres de los trabajadores deben estar directamente relacionados con las actividades que realizan para ganar en claridad en la modelación.

Para la modelación del negocio y como parte del desarrollo de la solución propuesta se identificaron los siguientes actores y trabajadores:

Actores del negocio.

Actor	Descripción
Director	Es la persona encargada de iniciar el negocio de la clasificación de documentos en la biblioteca.

Tabla 1: Actores del negocio.

Trabajadores del negocio.

Trabajador	Descripción
Bibliotecario	Es el responsable de clasificar los documentos en la biblioteca y de gestionar dicha clasificación y los documentos.
Administrador	Es el responsable de realizar las configuraciones previas al módulo.

Tabla 2: Trabajadores del negocio.

2.2.3 Diagrama de casos de uso del negocio.

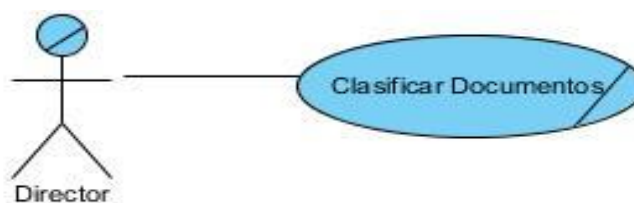


Figura 1: Diagrama de casos de uso del negocio.

2.2.4 Descripción de los casos de uso del negocio.

Caso de Uso del Negocio	Clasificar Documentos	
Actores	Director	
Resumen	El caso de uso inicia cuando el director le comunica al bibliotecario que comience a clasificar los documentos en la biblioteca.	
Acción del actor	Respuesta del proceso de negocio	
1. El director comunica al bibliotecario que comience la clasificación de documentos.	El bibliotecario crea las clasificaciones o categorías por el cual los documentos se van a clasificar en la biblioteca.	
	El bibliotecario escoge un documento y lee el resumen del mismo.	
	El bibliotecario le asigna una clasificación o categoría dado su conocimiento y experiencia en el campo de clasificación de documento.	

Tabla 3: Descripción del caso de uso del negocio Clasificar documentos.

2.2.5 Diagrama de actividades.

Véase en la figura 2 más adelante.

2.2.6 Modelo de objetos.



Figura 3: Modelo de objeto del negocio.

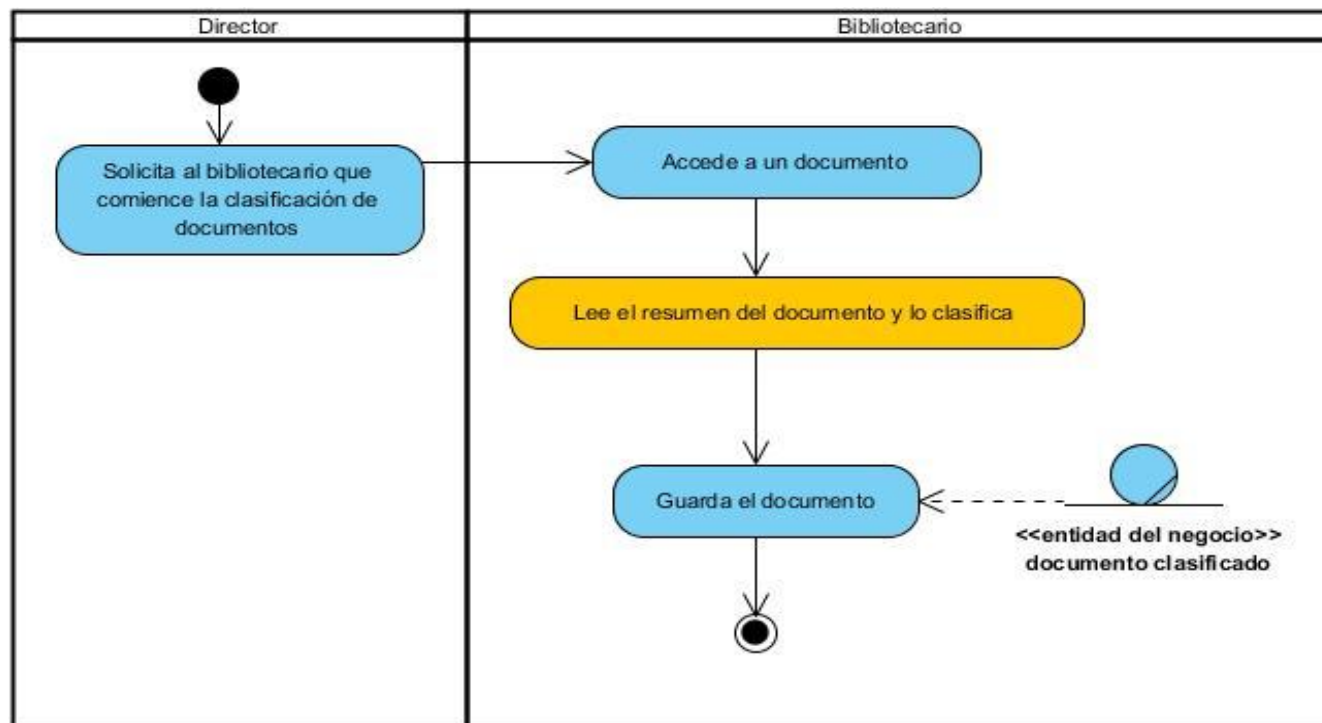


Figura 2: Diagrama de actividades del caso de uso Clasificar documentos.

2.3 Especificación de Requisitos.

Un requisito no es más que una declaración abstracta de alto nivel de un servicio que debe proporcionar el sistema o una restricción de este (EcuRed, 2012a). En el otro extremo, es una definición detallada y formal de una función del sistema. Los requisitos se pueden clasificar en **funcionales** y **no funcionales**.

A continuación se detallan los requisitos funcionales del módulo a desarrollar:

2.3.1 Requisitos Funcionales (RF).

RF1 Clasificar documentos: El sistema clasifica los documentos de forma automática.

RF2 Generar formulario: El sistema mostrará un resumen con los detalles de la clasificación de los documentos.

RF3 Gestionar clasificación:

- RF3.1 Adicionar clasificación: El bibliotecario adiciona la nueva clasificación en el sistema.
- RF3.2 Modificar clasificación: El bibliotecario modifica la clasificación previamente seleccionada.

➤ RF3.3 Eliminar clasificación: El bibliotecario elimina la clasificación que seleccione convenientemente.

➤ RF3.4 Mostrar clasificación: El sistema muestra las tablas con las clasificaciones previas.

RF4 Mostrar documentos clasificados: El sistema muestra los documentos clasificados existentes en la biblioteca.

RF5 Mostrar documentos sin clasificar: El sistema muestra los documentos sin clasificar que existen en la biblioteca.

RF6 Generar base de clasificación: El sistema genera textos para que el clasificador aprenda a categorizar.

RF7 Configurar módulo: El administrador configura el módulo con respecto a las características del sistema de clasificación de la biblioteca digital.

2.3.2 Requisitos no funcionales (RNF).

Los Requisitos No Funcionales son propiedades o cualidades que el producto debe tener. Debe pensarse en estas propiedades como las características que hacen al producto atractivo, usable, rápido y confiable. Son los aspectos del sistema visibles para el usuario, que no están relacionados de forma directa con el comportamiento funcional del sistema.

RNF1 Usabilidad:

- Será usable por cualquier tipo de usuario con experiencia, básica, media o avanzada en los procesos de clasificación de documentos.
- Se mostrará la información de forma lógica y correctamente estructurada.
- El módulo debe permitir que el usuario administrador lo configure de la forma deseada.

RNF2 Apariencia o interfaz externa:

- El módulo contará con un menú de herramientas para que el usuario pueda disponer de él en la medida de sus necesidades, visible todo el tiempo para propiciar el fácil acceso a las funcionalidades del software.
- El sistema tendrá un ambiente agradable y sencillo, con una estructura amigable que permita que el usuario se adapte con facilidad.

- El sistema permitirá claridad en los servicios que brinda, con términos asociados a los procesos de la clasificación de documentos, propiciando que el usuario tenga un buen entendimiento de las utilidades que brinda.
- Deberá visualizarse de manera correcta en los navegadores más usados en la actualidad, como: Internet Explorer versión 6.0 y 7.0, Mozilla Firefox, Opera, Netscape Navigator.

RNF3 Hardware:

- Para el buen funcionamiento del módulo en el sistema y del servicio MOCIC, se requiere de una PC Core (i3 | i5 | i7) con un mínimo de 4Gb de RAM¹². Mientras mayor memoria y mejor microprocesador¹³ posea la PC mejor rendimiento tendrá el sistema.
- Se requiere contar con los periféricos mouse y teclado.

RNF4 Portabilidad:

- El módulo debe ser independiente de plataforma siendo compatible con cualquier tipo de sistema operativo siempre que se integre al CMS Drupal en su versión 6.x.

RNF5 Software:

- Se utilizará como lenguaje de programación PHP 5.3.
- Se utilizará como Sistema Operativo, Windows 2000/XP/Seven, Linux, Unix.
- Se utilizará para el modelado la herramienta Visual Paradigm.
- Se utilizarán el IDE NetBeans 7.x.

RNF6 Reusabilidad:

- El módulo debe permitir posteriores modificaciones y actualizaciones, así como modificar elementos ya existentes.

2.4 Descripción del sistema propuesto.

¹² Siglas en inglés para Random Access Memory (Memoria de Acceso Aleatorio)

¹³ El microprocesador Constituye la unidad central de procesamiento (*CPU*) de un PC y es el encargado de ejecutar los programas desde el sistema operativo hasta las aplicaciones de usuario.

El módulo para la clasificación automática de documentos estará destinado a la Biblioteca digital “Alma Mater” de la República Bolivariana de Venezuela. En el mismo existirán usuarios con diferentes permisos con el objetivo de garantizar que los datos sean accedidos únicamente por las personas que cuentan con privilegios para ello.

El sistema deberá cumplir con determinadas funcionalidades, como son: Gestionar clasificación, Mostrar documentos clasificados y sin clasificar, Generar textos de entrenamiento, Configurar módulo, Generar formulario y Clasificar documentos. Para la selección de estas funcionalidades se tuvo en cuenta principalmente el nivel de importancia que estas tuvieran para el correcto funcionamiento del mismo.

2.4.1 Actores del sistema.

Los actores del sistema son terceros que interactúan con este sin ser parte del mismo. Cada trabajador del negocio es candidato a ser actor del sistema (Cruz, 2010).

Actor	Descripción
Administrador	El administrador es el responsable de realizar las configuraciones necesarias al módulo, el cual se va a integrar al sitio de la biblioteca digital.
Bibliotecario	Es el responsable de gestionar las clasificaciones de los documentos que existen en la biblioteca.

Tabla 4: Actores del sistema.

2.4.2 Diagrama de casos de uso del sistema.

A continuación se muestra el diagrama de casos de uso del sistema definido para el desarrollo del módulo de clasificación automática de documentos en la Biblioteca digital “Alma Mater”. Véase en la figura 4 más adelante.

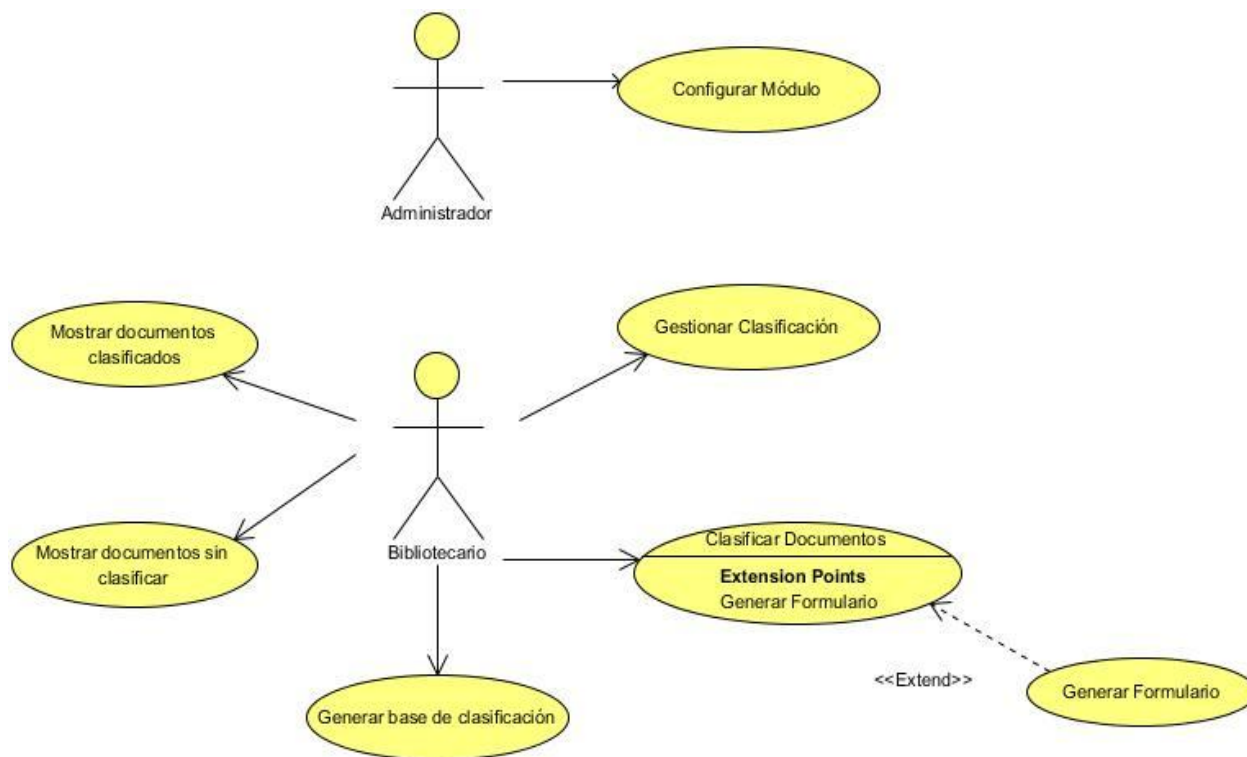


Figura 4: Diagrama de casos de uso del sistema.

2.4.3 Descripción de los Requisitos Funcionales.

RF1 Clasificar documentos.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF1	Clasificar documentos	El bibliotecario accede a una interfaz donde ejecuta la opción para que el sistema comience la clasificación de documentos de forma automática.	Alta	Alta
Prototipo				



Campos	Tipos de Datos	Reglas o Restricciones
No procede	No procede	El usuario autenticado debe ser el bibliotecario.

Tabla 5: Descripción del RF Clasificar documentos.

RF2 Generar formulario.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF 2	Generar formulario	Inicia cuando el sistema clasifica los documentos y se muestra una tabla con la cantidad de documentos clasificados, los que no pudo clasificar y el margen de error de clasificación que se obtuvo.	Media	Alta
Prototipo				

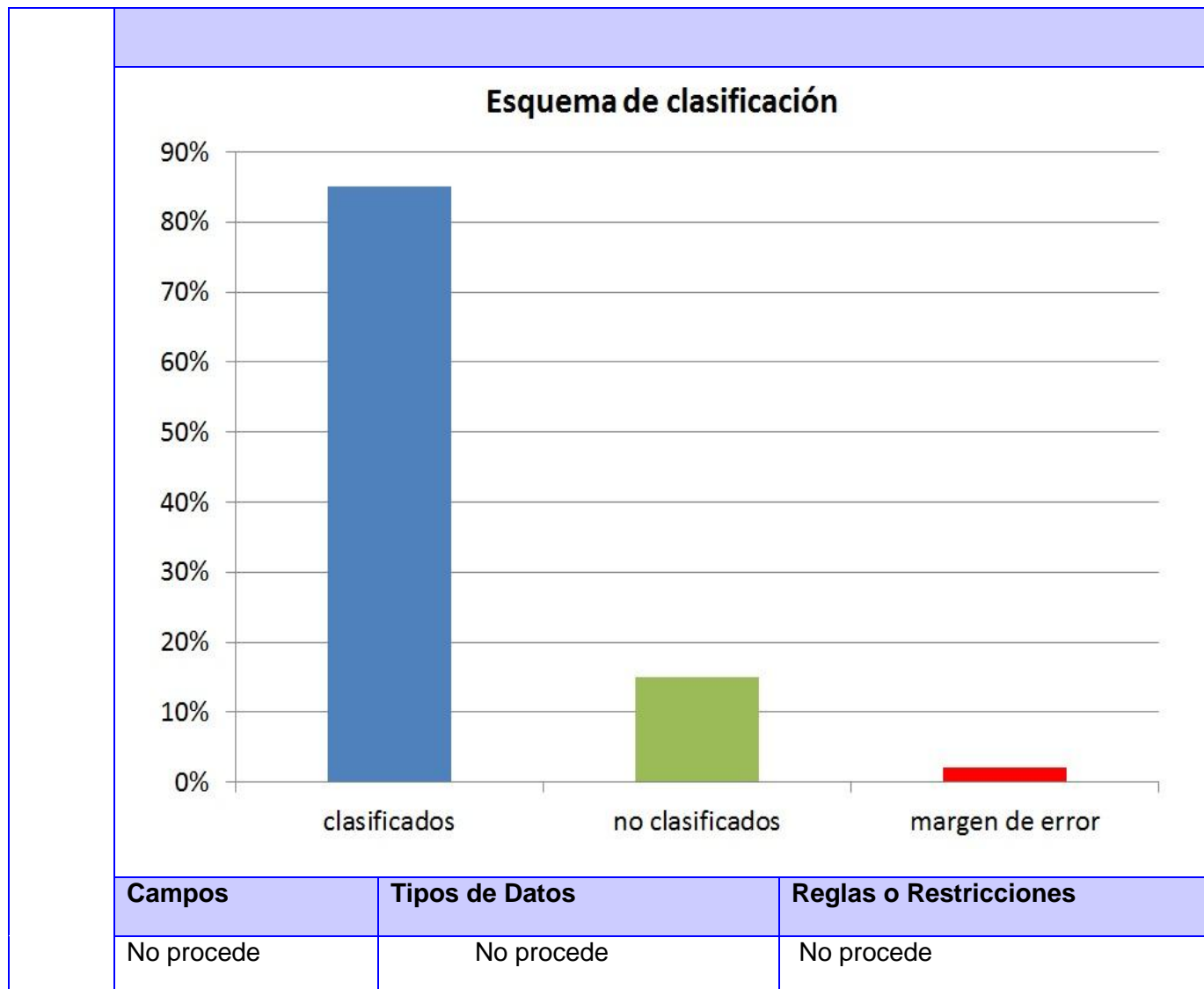


Tabla 6: Descripción del RF Generar formulario.

RF3 Gestionar clasificación.

RF3.1 Adicionar clasificación.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF3.1	Adicionar	Inicia cuando el bibliotecario accede a la opción de gestionar	Alta	Alta

	clasificación	las clasificaciones y comienza a crear las mismas.		
Prototipo				
<p>Identificador: *</p> <input type="text"/>				
<p>Nombre: *</p> <input type="text"/>				
<p>Texto de entrenamiento:</p> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 5px;"> <input type="text"/> </div> <p style="text-align: center; color: #4a7ebb; font-size: small;">Cambiar a editor de texto enriquecido</p> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 5px;"> <input type="text"/> </div> <p style="text-align: center; color: #4a7ebb; font-size: small;">Cambiar a editor de texto enriquecido</p> <p style="text-align: center; margin-top: 10px;"><input type="button" value="Agregar"/></p>				
	Campos	Tipos de Datos	Reglas o Restricciones	
	<ul style="list-style-type: none"> Identificador Nombre Resumen 	<ul style="list-style-type: none"> Varchar 	El usuario autenticado debe ser el bibliotecario.	

Tabla 7: Descripción del RF Adicionar clasificación.

RF3.2 Modificar clasificación.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
----	--------	-------------	-------------	------------------------

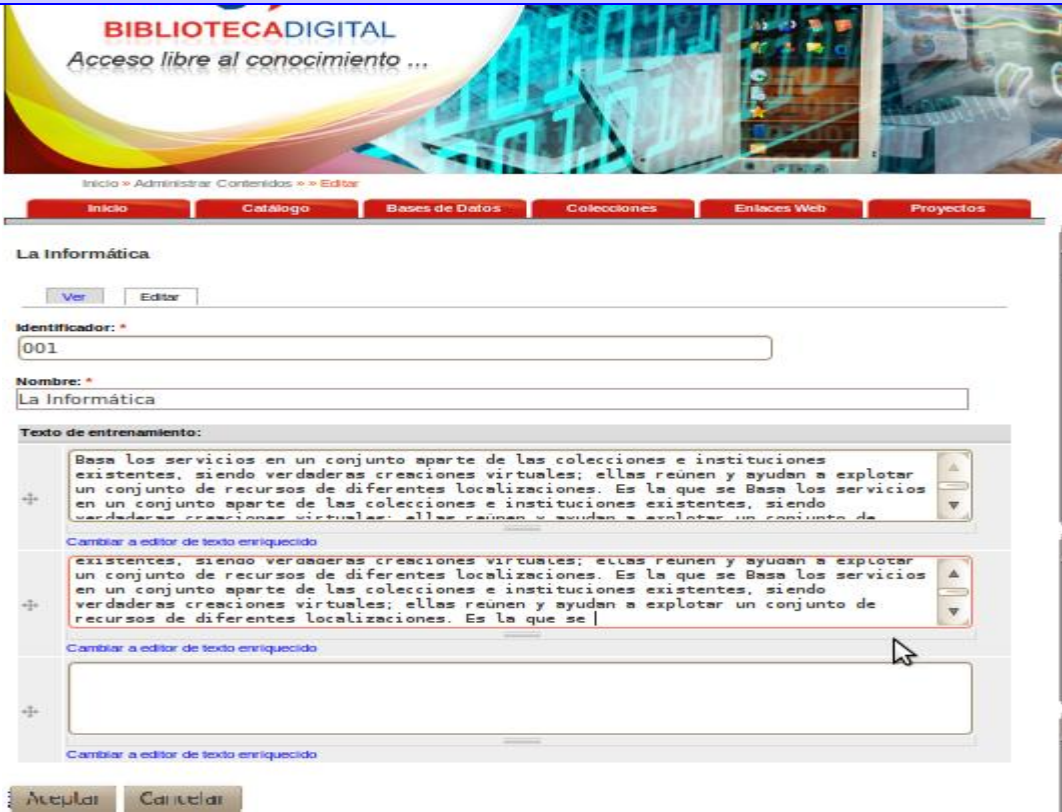
RF3.2	Modificar clasificación	Inicia cuando el bibliotecario muestra todas las clasificaciones y elige la que quiere modificar.	Alta	Alta
Prototipo				
				
Campos		Tipos de Datos	Reglas o Restricciones	
<ul style="list-style-type: none"> • Identificador • Nombre • Resumen 		<ul style="list-style-type: none"> • Varchar 	El usuario autenticado debe ser el bibliotecario.	

Tabla 8: Descripción del RF Modificar clasificación.

RF3.3 Eliminar clasificación.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
----	--------	-------------	-------------	------------------------

RF3.3	Eliminar clasificación	Inicia cuando el bibliotecario muestra todas las clasificaciones y elige la que quiere eliminar.	Alta	Alta
Prototipo				
				
Campos		Tipos de Datos	Reglas o Restricciones	
No procede		No procede	El usuario autenticado debe ser el bibliotecario.	

Tabla 9: Descripción del RF Eliminar clasificación.

RF3.4 Mostrar clasificación.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF3.4	Mostrar clasificación	Inicia cuando el bibliotecario accede al sistema para mostrar las clasificaciones que ya	Alta	Alta

	existan.														
Prototipo															
 <p>Inicio</p> <p>Inicio Catálogo Bases de Datos Colecciones Enlaces Web Proyectos</p> <p>Gestionar Clasificación</p> <table border="1"> <thead> <tr> <th>Identificador</th> <th>Nombre de la clasificación</th> <th>Editar</th> <th>Eliminar</th> </tr> </thead> <tbody> <tr> <td>001</td> <td>La Informática</td> <td>Editar</td> <td>Eliminar</td> </tr> <tr> <td>asdsda</td> <td>asdasdas</td> <td>Editar</td> <td>Eliminar</td> </tr> </tbody> </table>				Identificador	Nombre de la clasificación	Editar	Eliminar	001	La Informática	Editar	Eliminar	asdsda	asdasdas	Editar	Eliminar
Identificador	Nombre de la clasificación	Editar	Eliminar												
001	La Informática	Editar	Eliminar												
asdsda	asdasdas	Editar	Eliminar												
Campos	Tipos de Datos	Reglas o Restricciones													
No procede	No procede	El usuario autenticado debe ser el bibliotecario.													

Tabla 10: Descripción del RF Mostrar clasificación.

RF4 Mostrar documentos clasificados.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF4	Mostrar documentos clasificados	Inicia cuando el bibliotecario accede al sistema para mostrar los documentos clasificados en	Alta	Alta

	la biblioteca digital.		
Prototipo			
			
Campos	Tipos de Datos	Reglas o Restricciones	
No procede	No procede	El usuario autenticado debe ser el bibliotecario.	

Tabla 11: Descripción del RF Mostrar documentos clasificados.

RF5 Mostrar documentos sin clasificar.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF5	Mostrar documentos	Inicia cuando el bibliotecario accede al sistema para mostrar	Alta	Alta

sin clasificar	los documentos sin clasificar en la biblioteca digital.		
Prototipo			
			
Campos	Tipos de Datos	Reglas o Restricciones	
No procede	No procede	El usuario autenticado debe ser el bibliotecario.	

Tabla 12: Descripción del RF Mostrar documentos sin clasificar.

RF6 Generar base de clasificación.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF6	Generar base de clasificación	Inicia cuando el bibliotecario accede al sistema para generar la base de clasificación o textos	Alta	Alta

		de entrenamiento por el cual el clasificador automático se entrenará.		
Prototipo				
				
Campos	Tipos de Datos		Reglas o Restricciones	
No procede	No procede		El usuario autenticado debe ser el bibliotecario.	

Tabla 13: Descripción del RF Generar base de clasificación.

RF7 Configurar módulo.

Nº	Nombre	Descripción	Complejidad	Prioridad para cliente
RF7	Configurar módulo	Inicia cuando el administrador accede al sistema para realizar las configuraciones pertinentes	Alta	Alta

	al módulo de clasificación automática.		
Prototipo			
			
Campos	Tipos de Datos	Reglas o Restricciones	
<ul style="list-style-type: none"> • Dir_servicio • Dir_clasificación 	<ul style="list-style-type: none"> • Varchar 	El usuario autenticado debe ser el bibliotecario.	

Tabla 14: Descripción del RF Configurar módulo.

2.5 Modelo de análisis.

Es un modelo conceptual puesto que lo que se ilustra en el mismo son una serie de clases y relaciones entre las mismas que permiten una interpretación correcta del funcionamiento de la parte del sistema que se esté modelando. Durante el análisis se reestructuran y refinan los requisitos. Está descrito en el

lenguaje del desarrollador con el objetivo de que se comprenda claramente cómo debería diseñarse e implementarse (Cruz, 2010).

2.5.1 Diagrama de clases del análisis.

Un Diagrama de clases del análisis es un artefacto en el que se representan los conceptos en un dominio del problema. Representa el funcionamiento del mundo real, no de la implementación automatizada del mismo (Cruz, 2010).

2.5.2 Diagrama de clases del análisis Clasificar documentos.

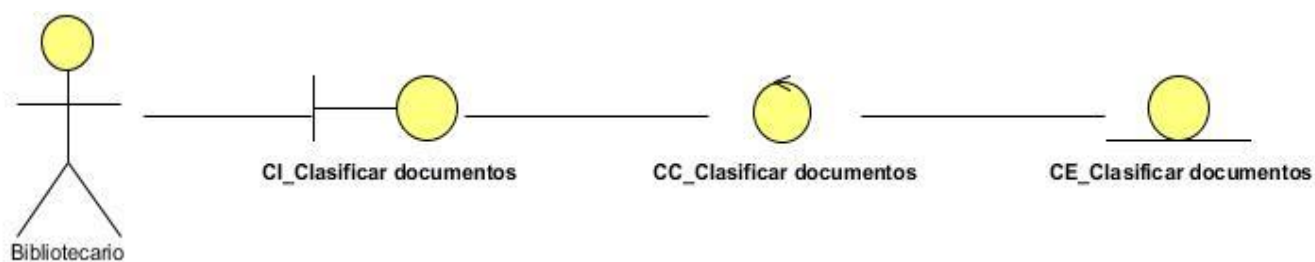


Figura 5: Diagrama de clases del análisis Clasificar documentos

Se puede consultar los demás diagramas en los Anexos más adelante.

2.6 Arquitectura de Drupal.

Drupal está diseñado para funcionar sobre, prácticamente, cualquier entorno independientemente del sistema operativo, servidor web o sistema de gestión de base de datos (Asamblea, 2011). El mismo proporciona un framework responsable de proveer las funcionalidades básicas necesarias en las otras partes del sistema, así como de servir de soporte para las nuevas funcionalidades.

El framework que Drupal proporciona, facilita la ampliación de nuevas funcionalidades a partir de unidades auto contenidas denominadas “módulos”. A través de dichos módulos Drupal marca el camino que los desarrolladores han de seguir para ampliarlo funcionalmente, ya sea creando nuevos tipos de contenidos con nuevas funcionalidades como por ejemplo autenticación de usuarios frente a LDAP, integración con el gestor documental Drupal, soporte para e-commerce, etc.

Para el desarrollo de la solución propuesta el patrón arquitectónico utilizado se asemeja al patrón Modelo Vista Controlador (MVC), donde el Modelo está representado por la capa de abstracción de la base de datos que emplea el CMS Drupal, la Vista mediante el tema implementado y el Controlador a través de los módulos desarrollados (Asamblea, 2011).

El patrón arquitectónico MVC separa los datos, la lógica del negocio y la interfaz de usuario en tres capas distintas:

- **Modelo:** corresponde a la parte de la aplicación que implementa la lógica del dominio de datos de la aplicación.
- **Vista:** es la capa de la aplicación que ve el usuario en un formato adecuado para interactuar.
- **Controlador:** es la capa que controla todo lo que puede realizar la aplicación. Responde a eventos, usualmente acciones del usuario e invoca cambios en el modelo y probablemente en la vista.

2.6.1 Patrones de diseño en Drupal.

Dentro del core¹⁴ de Drupal se utilizan varios de patrones de diseño, como por ejemplo:

Controlador: El patrón controlador es un patrón que sirve como intermediario entre una determinada interfaz y el algoritmo que la implementa, de tal forma que es la que recibe los datos del usuario y la que los envía a las distintas clases según el método llamado.

Este patrón sugiere que la lógica de negocios debe estar separada de la capa de presentación, esto para aumentar la reutilización de código y a la vez tener un mayor control.

Se recomienda dividir los eventos del sistema en el mayor número de controladores para poder aumentar la cohesión y disminuir el acoplamiento.

Observer (Observador): Define una dependencia de uno a muchos entre objetos, de forma que cuando un objeto cambie de estado se notifique y actualicen automáticamente todos los objetos que dependen de él.

Comando: Muchos de los ganchos de Drupal utilizan el patrón de comando para reducir el número de funciones que son necesarias para la aplicación, pasando por la operación como un parámetro, junto con los argumentos. De hecho, el sistema de enganche utiliza este modelo, por lo que los módulos no tienen que definir cada gancho, sino sólo los que les importa poner en práctica.

Puente: El Drupal capa de abstracción de base de datos se lleva a cabo de manera similar al patrón de diseño del puente. Los módulos deben ser escritos en una forma que es independiente del sistema de base de datos en uso, y la capa de abstracción que lo provee. Nuevas capas de base de datos se pueden escribir que se ajusten a la API definida por el puente, añadiendo soporte para los sistemas de bases de datos adicionales sin necesidad de modificar el código del módulo.

¹⁴ También llamado el núcleo de Drupal.

Decorador: Drupal hace un uso extensivo del patrón decorador. El polimorfismo de los objetos de nodo es sólo una pequeña parte del poder del sistema de nodo. Más interesante es el uso del nodo con varios ganchos (`hook_node_load ()`, `hook_node_view ()`, etc), que permiten ampliar los módulos que arbitraría la conducta de todos los nodos.

2.7 Modelo de diseño.

Es un modelo de objetos que describe la realización física de los casos de usos, centrándose en cómo los requisitos funcionales y no funcionales tienen un fuerte impacto en el sistema a desarrollar (Palenzuela, 2010).

El diseño es un proceso iterativo a través del cual se traducen los requisitos en una representación del software. Se representa a un alto nivel de abstracción, un nivel que se puede seguir hasta los requisitos específicos de datos, funcionales y de comportamiento (Meriño y Castro, 2011).

Con el diseño se pretende construir un sistema que:

- Satisfaga determinada especificación del sistema.
- Se ajuste a las limitaciones impuestas por el medio de destino.
- Respete los requisitos sobre forma, rendimiento, utilización de recursos y costo.

El diseño es un refinamiento de los requisitos, que tiene en cuenta los requisitos no funcionales, en pocas palabras, cómo cumple el sistema sus objetivos.

2.7.1 Diagrama de clase de diseño web.

Un diagrama de clases es un tipo de diagrama de estructura estática que describe la estructura de un sistema mostrando sus clases, atributos y las relaciones entre ellos. Los diagramas de clases son utilizados durante el proceso de diseño de los sistemas, es donde se crea el diseño conceptual de la información que se manejará en el sistema y los componentes que se encargarán del funcionamiento y la relación entre uno y otro (WIKIPEDIA, 2012). El diagrama de clases de diseño describe gráficamente las especificaciones de las clases de software y de las interfaces en una aplicación.

El CMS Drupal provee un tipo de contenido genérico llamado “node”, y permite la creación, edición y publicación de este tipo de contenido. Los documentos creados se depositan en una base de datos central, donde también se guardan el resto de datos de la web. Posee un potente sistema de seguridad, basado en la creación de usuarios y roles, en dependencia de los permisos que se necesiten para administrar el sitio y acceder a los módulos.

Para que se comprenda de forma sencilla el diagrama de clases de diseño, se decidió describir por paquetes los módulos de Drupal.

- **Includes:** contiene un conjunto de ficheros imprescindibles para el funcionamiento, como el fichero Database, que provee las funcionalidades de acceso a la base de datos del CMS Drupal.
- **Themes:** contiene las plantillas de diseño, cuando se desee incluir uno nuevo, se copia la plantilla dentro de esta carpeta.
- **Scripts:** contiene un conjunto de ficheros necesarios orientados al aspecto visual, CSS y JavaScript para el correcto funcionamiento de la aplicación.
- **Modules:** comprende todos los módulos, que permiten las distintas funcionalidades del CMS, cuando desee incorporar un nuevo módulo sólo se tiene que copiar en dicha carpeta.

Luego de explicado el CMS Drupal, se modela primeramente el diagrama general del Drupal y luego los diagramas correspondientes al módulo de clasificación. El diagrama general muestra el funcionamiento del CMS Drupal con sus clases y módulos.

Diagrama de clases del diseño general del CMS Drupal:

Véase en la figura 6 más adelante.

Diagrama de clases del diseño Clasificar documentos:

Véase en la figura 7 más adelante.

2.7.2 Diseño de la Base de datos.

La meta más importante que se persigue al diseñar un modelo de base de datos, es dividir la información en tablas organizadas por temas, para reducir los datos redundantes y proporcionar el acceso a la información necesaria y así reunir la información de las tablas cuando se precise. Un correcto diseño de la base de datos ayuda a garantizar la exactitud e integridad de la información y satisfacer las necesidades de procesamiento de los datos y de generación de informes (Meriño y Castro, 2011).

El objetivo principal del diseño de bases de datos es generar tablas que modelan los registros en los que se guarda la información del sistema. Posteriormente se ilustra el diagrama de clases persistentes correspondiente (Rodas, 2009).

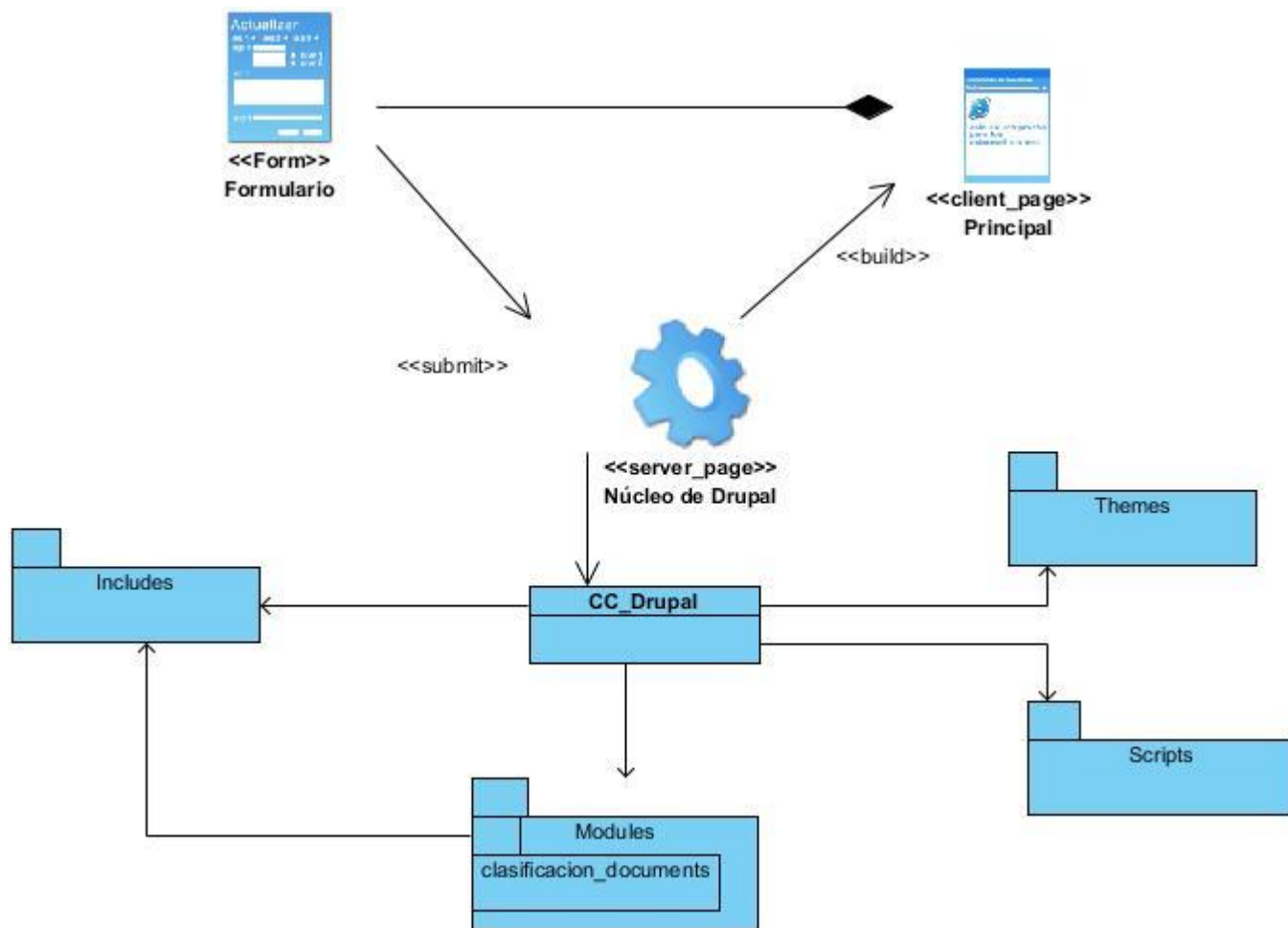


Figura 6: Diagrama de clases del diseño general del CMS Drupal.

2.7.3 Diagrama de clases persistentes.

El diagrama de clases persistentes muestra todas las clases capaces de mantener su valor en el tiempo y el espacio. Es necesario aclarar que el CMS Drupal cuenta con su propia base de datos por lo que se hace representación del diagrama de clases persistentes a partir de las entidades que se le adicionaron para el módulo de clasificación automática de documentos.

Véase en la figura 8 más adelante.

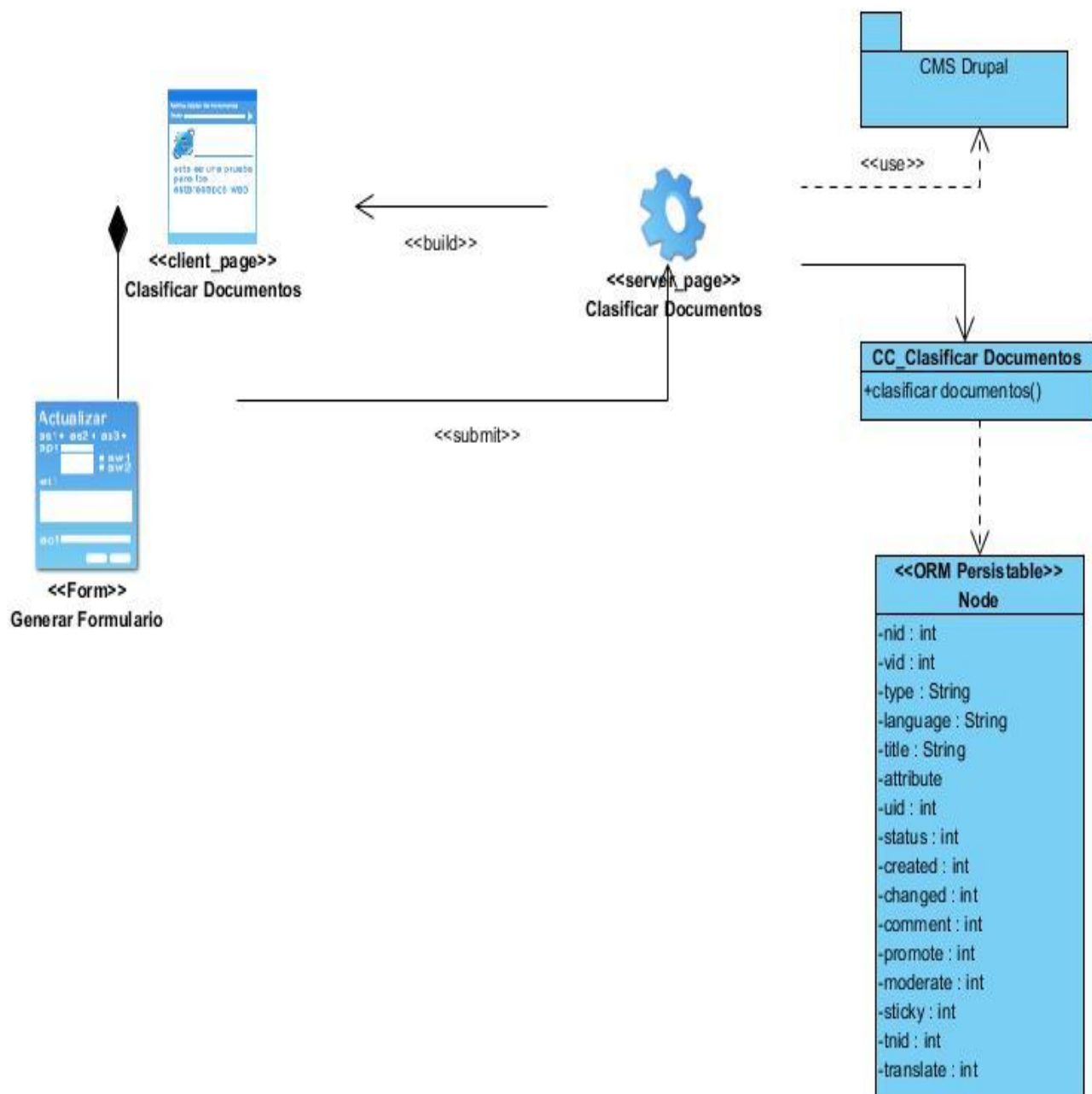


Figura 7: Diagrama de clases del diseño Clasificar documentos.

Se puede consultar los demás diagramas en los Anexos más adelante.

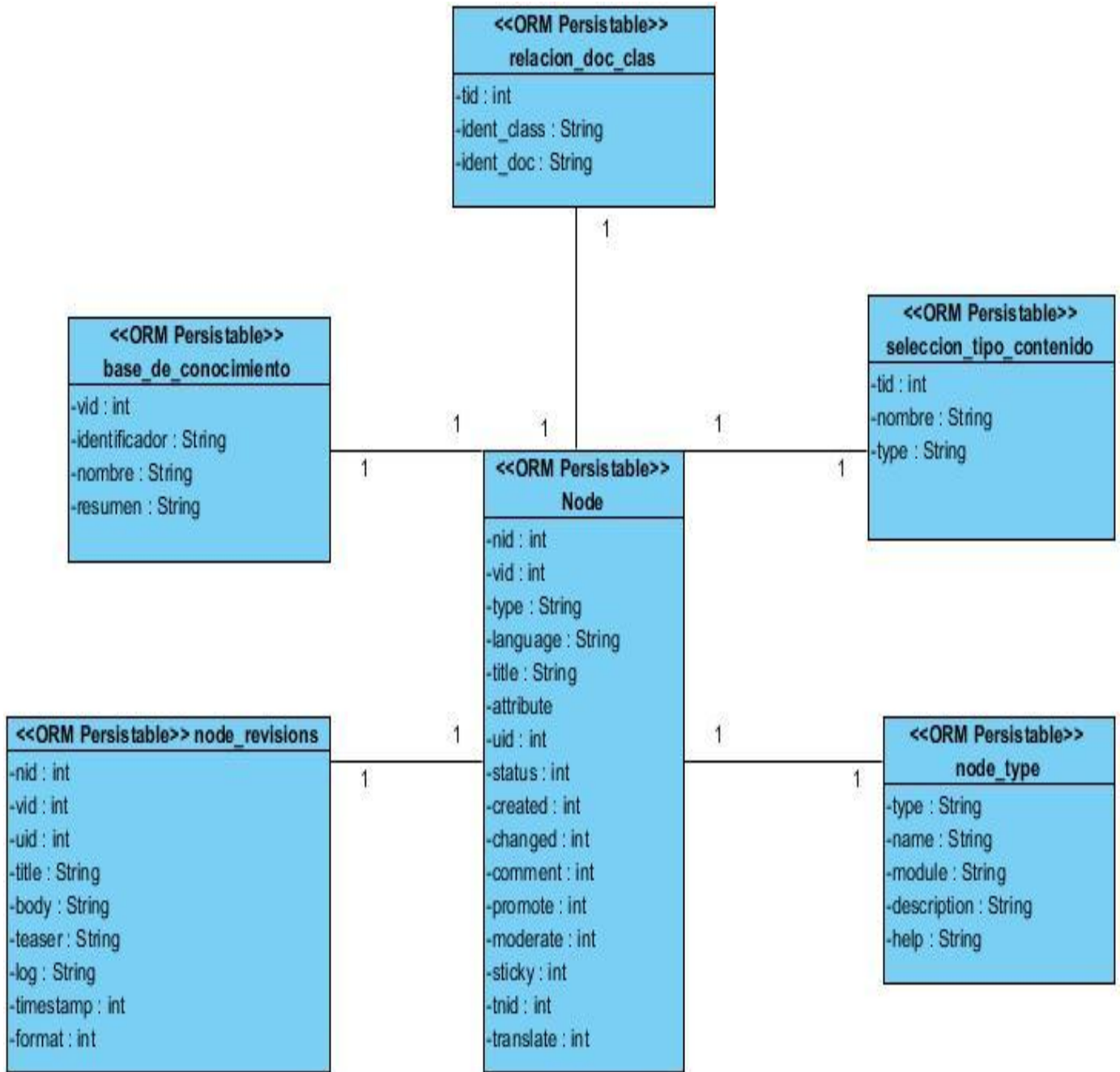


Figura 8: Diagrama de clases persistente del módulo de clasificación.

2.7.4 Modelo de datos.

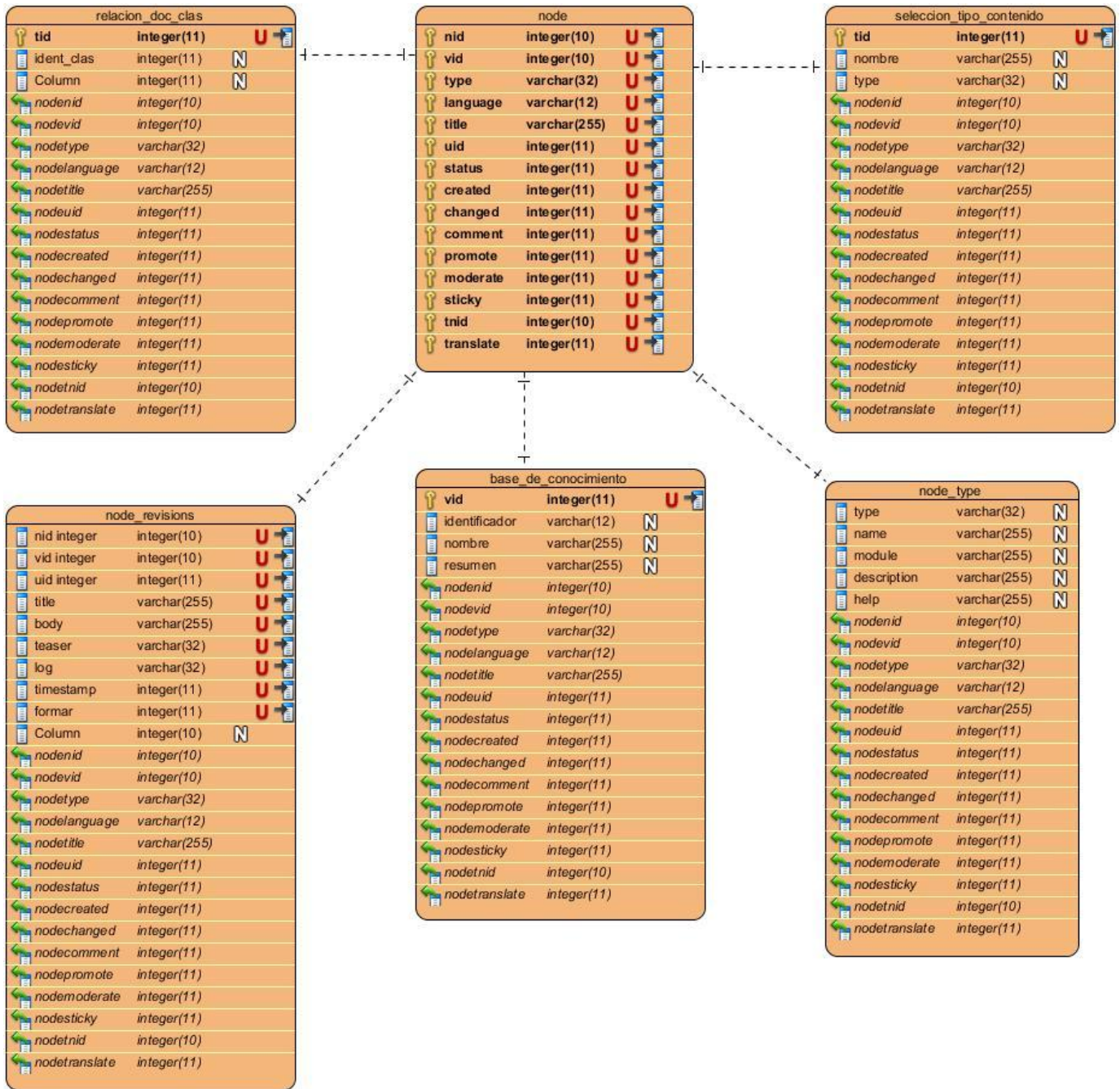


Figura 9: Diagrama del Modelo de datos del módulo de clasificación.

2.7.5 Descripción de las tablas de base de datos.

Luego de ilustrado el diagrama de clases persistentes y el modelo de datos, se procede a describir cada una de las tablas de la base de datos, donde se detallan los atributos correspondientes a cada tabla, el tipo y una breve descripción de los mismos para lograr una mayor comprensión de cómo están estructurados los datos.

Nombre: base_de_conocimiento		
Descripción: Tabla para almacenar los distintos tipos de categorías para clasificar los documentos		
Atributo	Tipo	Descripción
vid	integer	Identificador de la posición de la tabla de la clasificación
identificador	varchar	Identificador de la clasificación
nombre	varchar	Nombre de la clasificación
Resumen	varchar	Resumen o texto de entrenamiento de la clasificación

Tabla 15: Descripción de la tabla de base de datos base_de_conocimiento.

Se puede consultar las demás tablas en los Anexos más adelante.

Conclusiones del capítulo.

En este capítulo se realizó el modelamiento del negocio para el módulo de clasificación automática de documentos en la Biblioteca digital “Alma Mater”. Este proceso permitió comprender la dinámica y estructura de la organización, así como definir los requisitos funcionales y elaborar el Modelo de casos de uso del sistema, donde se describen los actores y de los casos de uso, dando así un mejor entendimiento de las funcionalidades. Se realizó además, el modelo de diseño, obteniendo como resultado los diagramas de clases del diseño del sistema, usándose como patrón el MVC, con el cual se obtiene una mejor comprensión para la implementación del sistema; y el diagrama de clases persistentes de la base de datos, que permitió conocer la información física que se utilizó para construir el módulo.

Capítulo 3: Implementación y validación del sistema

3.1 Introducción.

Este capítulo describe cómo los elementos del modelo de diseño son implementados en términos de componentes y cómo los mismos se organizan de acuerdo con los nodos referidos en el modelo de despliegue. Se exponen las pruebas realizadas a cada caso de uso, siguiendo particularmente, el método de pruebas de Caja Negra.

3.2 Diagrama de despliegue

Permite modelar la disposición física o topología de un sistema. Muestra el hardware usado y los componentes instalados en el hardware. Muestra las conexiones físicas entre el hardware y las relaciones entre componentes.

A continuación se muestra el diagrama de despliegue correspondiente a la aplicación, donde se puede observar que la misma se encontrará desplegada en una unidad de procesamiento, contenida dentro del Servidor de Aplicaciones web (Apache). La base de datos estará desplegada en una unidad de procesamiento corriendo el servicio de PostgreSQL y la asociación entre los nodos representa la ruta de comunicación entre ellos.

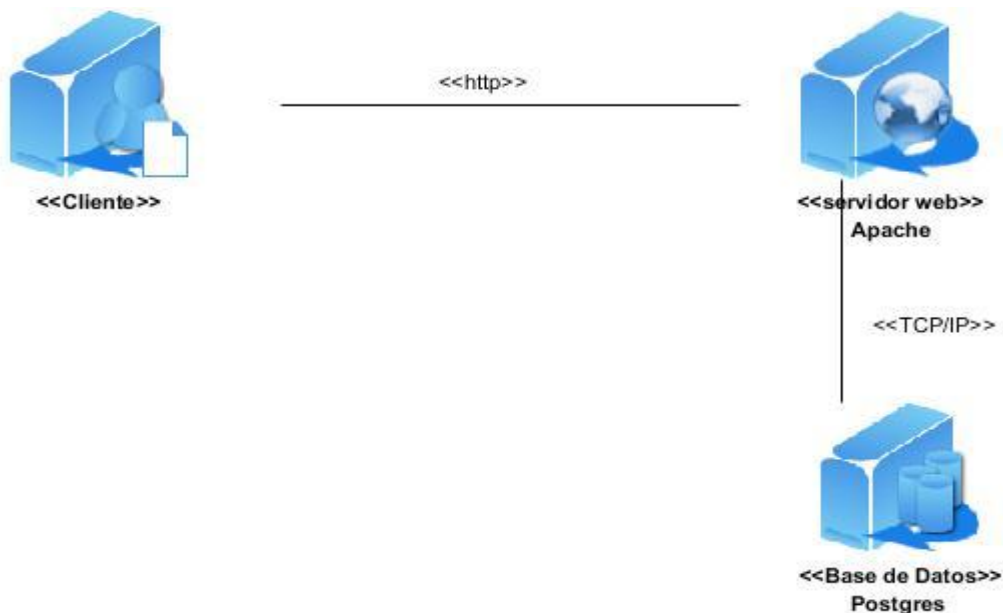


Figura 10: Diagrama de Despliegue del módulo de clasificación.

3.3 Pruebas.

Las pruebas de software se pueden traducir como una revisión del sistema con el objetivo de encontrar problemas antes que estos sean encontrados por el cliente final. El éxito de las pruebas puede mejorar la percepción de calidad del usuario final, evitando que un defecto grave se entregue juntamente con el sistema. Las pruebas verifican que el producto funcione como se diseñó y que los requerimientos son satisfechos cabalmente, además de brindar soporte para encontrar, documentar y solucionar defectos del sistema (Meriño y Castro, 2011).

3.3.1 Pruebas de caja negra.

Se refiere a las pruebas que se llevan a cabo sobre la interfaz del software o sistema. Los casos de pruebas pretenden demostrar que las funciones del software son operativas, que la entrada se acepta de forma adecuada y que se produce un resultado correcto (EcuRed, 2012b).

En las siguientes tablas se le harán pruebas de funcionalidad, que entran en el alcance de las pruebas de caja Negra, a los casos de usos antes detallados.

Descripción del caso de prueba para el requisito funcional Adicionar clasificación.

Nombre del requisito	Descripción general	Escenarios de pruebas	Flujo del escenario
1:Adicionar clasificación	Se adiciona una nueva clasificación con los campos: <ul style="list-style-type: none"> • identificador • nombre • resumen 	EP 1.1: Adicionar una clasificación introduciendo datos válidos.	Se presiona el link Adicionar clasificación. Se introducen correctamente los datos. Se presiona el botón Aceptar

		<p>EP 1.2: Adicionar una clasificación introduciendo datos no válidos.</p>	<ul style="list-style-type: none"> - Se presiona el link Adicionar. - Se introducen incorrectamente los datos. - Se presiona el botón Aceptar.
		<p>EP 1.3: Adicionar una clasificación dejando campos vacíos.</p>	<ul style="list-style-type: none"> - Se presiona el link Adicionar clasificación. - Se introducen los datos dejando campos vacíos. - Se presiona el botón Aceptar.
		<p>EP 1.4: Adicionar una clasificación con datos que existen en la biblioteca.</p>	<ul style="list-style-type: none"> - Se presiona el link Adicionar clasificación. - Se introducen los nuevos datos - Se presiona el botón Aceptar.
		<p>EP 1.5 Cancelar una clasificación.</p>	<ul style="list-style-type: none"> - Se presiona el link Adicionar clasificación. - Se introducen o no los nuevos datos. - Se presiona el botón

			cancelar
--	--	--	----------

Tabla 16: Descripción del caso de prueba para el RF Adicionar clasificación.

Descripción de las variable para el requisito funcional Adicionar clasificación:

No	Nombre de campo	Tipo	Válido	Inválido
1	Identificador	Textfield	Números y puntos.	Letras y caracteres especiales.
2	Nombre	Textfield	Letras y caracteres especiales.	Números
3	Resumen	Textarea	Números, letras y caracteres especiales.	NA

Tabla 17: Descripción de variable para el RF Adicionar clasificación.

Descripción del juego a probar del requisito funcional Adicionar clasificación:

Id del escenario	Escenario	Identificador	Nombre	Resumen	Respuesta del sistema	Resultado de la prueba
EP 1.1	Adicionar una clasificación introduciendo datos válidos.	V(111.23)	V(Daniel)	V(daniel12)	El sistema muestra un mensaje de información de color verde donde explica que el documento ha sido adicionado satisfactoriamente.	Satisfactorio
EP 1.2	Adicionar una clasificación	I(Daniel)	I (123,'#%.)	NA	El sistema muestra un mensaje de	Satisfactorio

	introduciendo datos no válidos.				color rojo informando que existen datos incorrectos en el campo(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Adicionar clasificación.	
EP 1.3	Adicionar una clasificación dejando campos vacíos.	I(Vacío)	I(Vacío)	I(Vacío)	El sistema muestra un mensaje de color rojo informando que existen campo(s) vacío(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Adicionar clasificación.	Satisfactorio

EP 1.4	Adicionar una clasificación con datos que ya existan	V(111)	V(Daniel)	V(daniel12)	El sistema muestra un mensaje de color rojo informando que hay datos que ya existen en la biblioteca marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Adicionar clasificación.	Satisfactorio
EP 1.5	Cancelar una clasificación.	NA	NA	NA	Se cancela la operación y se re direcciona a la página de inicio	No satisfactorio

Tabla 18: Descripción del juego a probar para el RF Adicionar clasificación.

Descripción del Registro de defectos y dificultades detectadas para el requisito funcional Adicionar clasificación:

Requisito	No.	No conformidad	Aspecto correspondiente	Etapa de detección	Significativa	No significativa	Recomendación	Estado NC	Respuesta del equipo de desarrollo
Adicionar clasificación	1	El nombre del link que comienza el requisito no es el mismo que el que se describe, este dice "Insertar clasificación"	Es en la interfaz general, en el bloque inferior derecho.	1ra		x		Pendiente	Resuelta
	2	Cuando se da clic en el botón de cancelar no se redirecciona para la página prevista	Es en la interfaz de adicionar una clasificación.	1ra		x		Pendiente	Resuelta
	3	Cuando se adiciona una	Es en la interfaz de adicionar una	1ra		x		Pendiente	Resuelta

		clasificación el mensaje que se muestra no tiene un punto final.	clasificación.						
	4	Cuando se adiciona una clasificación el nombre en el resumen está mal escrito.	Es en la interfaz de adicionar una clasificación.	1ra	x				Resuelta

Tabla 19: Descripción del Registro de defectos y dificultades detectadas para el RF Adicionar clasificación

Se pueden ver las demás tablas en los Anexos más adelante.

3.3.2 Resultados de las pruebas internas.

Para comprobar la calidad y funcionalidad del módulo desarrollado y con el empleo de los diseños de casos de pruebas para la aplicación, se realizó una iteración de revisión por el grupo de calidad.

Durante la iteración se detectaron 8 no conformidades, de ellas 2 fueron significativas, lo que representa un 25 % del total. Todas las no conformidades fueron resueltas.

Conclusiones del capítulo.

En este capítulo se aplicaron pruebas que permitieron evaluar funcionalmente la solución propuesta obteniendo como resultado que las funcionalidades implementadas mostraron el correcto funcionamiento y respuesta de los requisitos.

Conclusiones

Con el desarrollo de este trabajo se puede afirmar que:

- Los conceptos fundamentales estudiados en el marco teórico permitieron el entendimiento del área en el que se desarrolla la investigación. Además, el estudio del estado del arte posibilitó la selección adecuada del software para la clasificación de documentos.
- La descripción de tecnologías, herramientas y metodologías permitió sentar las bases para el desarrollo del trabajo de diploma.
- El diseño del módulo para la clasificación automática de documentos cumplió con los requerimientos descritos obteniendo una mejor comprensión para la implementación del sistema.
- Las pruebas realizadas permitieron evaluar funcionalmente la solución propuesta, como resultado, las funcionalidades implementadas mostraron un correcto funcionamiento.

Recomendaciones

Para futuras versiones del módulo, se recomienda:

- Ofrecer mayor flexibilidad en la configuración del módulo, en la selección de los textos del documento que serán usados para la clasificación.
- Incluir los métodos de clasificación como funcionalidades del módulo, evitando así, posibles problemas de integración con otros sistemas.
- Se propone usar el módulo en la biblioteca digital de la Universidad de las Ciencias Informáticas con el fin de acelerar el proceso de clasificación automática en dicha biblioteca.

Referencias Bibliográficas

ASSEMBLA, 2011. Diseño de Drupal [en línea]. 2011. S.l.: s.n. Disponible en: my-svn.assembla.com/svn/.../Capítulo%207%20-%20Diseño.odt.

ECURED, 2012a. Flujo de Trabajo Modelo del Negocio - EcuRed. In: [en línea]. 2012. [Accedido 18 Febrero 2012]. Disponible en: http://www.ecured.cu/index.php/Flujo_de_Trabajo_Modelo_del_Negocio#Actores_del_Negocio.

ECURED, 2012b. Prueba de RUP - EcuRed. In: [en línea]. 2012. [Accedido 22 Febrero 2012]. Disponible en: http://www.ecured.cu/index.php/Prueba_de_RUP.

ECURED, 2012c. Requisitos de Software - EcuRed. In: [en línea]. 2012. [Accedido 16 Enero 2012]. Disponible en: http://www.ecured.cu/index.php/Requisitos_de_Software.

ESPAÑOLA, DICCIONARIO DE LA LENGUA, 2012. 2012. [Accedido 15 Enero 2012]. Disponible en: http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=biblioteca.

FIGUEROLA, Carlos G., BERROCAL, José L. Alonso, RODRÍGUEZ, Angel F. Zazo and RODRÍGUEZ, Emilio, 2004. Cuadernos de Documentación Multimedia - N.º 15 (2004). In: [en línea]. 2004. [Accedido 18 Abril 2012]. Disponible en: <http://multidoc.rediris.es/cdm/viewarticle.php?id=28&layout=html>.

GUERVOS, Juan Julian Merelo, 2005. Introducción a los sistemas de gestión de contenidos. In: [en línea]. 2005. [Accedido 20 Abril 2012]. Disponible en: <http://geneura.ugr.es/~jmerelo/tutoriales/cms/>.

GUZMÁN, Clara López, 2000. Biblioteca Digital. In: [en línea]. March 2000. [Accedido 15 Mayo 2012]. Disponible en: <http://www.cyta.com.ar/ta0101/biblio.htm>.

MULTIDOC, 2008. Clasificación automática de textos [en línea]. 2008. S.l.: s.n. Disponible en: www.ucm.es/info/multidoc/.../tema-8-clasificacion-automatica.doc.

RODAS, Christian, 2009. Diseño de una Base de Datos [en línea]. 2009. S.l.: s.n. Disponible en: <http://www.slideshare.net/ChristianR/diseo-de-base-de-datos>.

SAZ, Dr. Jesús Tranullas, 2005. 2005. [Accedido 10 Mayo 2012]. Disponible en: <http://www.absysnet.com/tema/tema41.html>.

SAZ, Dr. Jesús Tranullas, 2011. Definición Biblioteca Digital - TICS. In: [en línea]. 2011. [Accedido 25 Mayo 2012]. Disponible en: <https://sites.google.com/site/posbibliotecadigital/definicion>.

VISUAL BIBLIOTECAS, 2012. Visual Bibliotecas ®. Programa de gestión de bibliotecas, aplicación para la gestión de bibliotecas, programa para bibliotecas escolares, software a medida... In: [en línea]. 2012. [Accedido 13 Mayo 2012]. Disponible en: <http://www.floressoft.com/htm/programavb.htm>.

WIKIPEDIA, 2012a. Diagrama de clases - Wikipedia, la enciclopedia libre. In: [en línea]. 2012. [Accedido 8 Marzo 2012]. Disponible en: http://es.wikipedia.org/wiki/Diagrama_de_clases.

WIKIPEDIA, 2012b. El patrón MVC « Código Unix. In: [en línea]. 2012. [Accedido 24 Abril 2012]. Disponible en: <http://facusdelacruz.wordpress.com/2008/09/16/el-patron-mvc/>. Véase referencias en WIKIPEDIA que son confiables.

WORDPRESS, 2008. *Herramientas Case* [online]. 2008. S.l.: s.n. Available from: angelmpe.files.wordpress.com/2011/08/herramientas_case.docx.

Bibliografía

ACOSTA, Dra. Reina Estrella Herrera, 2003. Clasificación de documentos [en línea]. La Habana: s.n. Disponible en: http://archivohistoricoiuetueb.wikispaces.com/file/view/CLASIFICACION+DE+DOCUMENTOS_v2.pdf.

BATISTA, Arturo Leandro Sánchez and ROSARIO, Osney Mateo, 2010. Personalización de Drupal para teléfonos móviles. La Habana: Universidad de las Ciencias Informáticas.

CRUZ, Eduardo Ramírez, 2010. Análisis y Diseño del Sistema para Graficar Columnas Litológicas de Pozos de Petróleo [en línea]. La Habana: Universidad de las Ciencias Informáticas. Disponible en: http://repositorio_institucional.uci.cu/jspui/bitstream/ident/TD_03653_10/1/TD_03653_10.pdf.

GARCÍA, Anay Medina, 2011. Sistema de Recuperación de Información Geográfica para el motor de búsqueda Orión. La Habana: Universidad de las Ciencias Informáticas.

GASTELL, Daimerys Ceballo and CERUTO, Yanet del Carmen Diego, 2008. Categorización de textos usando Redes Neuronales artificiales [en línea]. La Habana: Universidad de las Ciencias Informáticas. Disponible en: http://repositorio_institucional.uci.cu/jspui/bitstream/ident/TD_1315_08/1/TD_1315_08.pdf.

Mercer, David. 2008. Building powerfull and robust websites with Drupal 6. s.l. : PACKT PUBLISHING, 2008.

MARCHENA, Dionisio Estrada and PÉREZ, José Ramón Fernández, 2007. Agente de recuperación de información Just-intime [en línea]. La Habana: Universidad de las Ciencias Informáticas. Disponible en: http://repositorio_institucional.uci.cu/jspui/bitstream/ident/TD_0879_07/1/TD_0879_07.pdf.

MERIÑO, Joel Paulino Rodríguez and CASTRO, Yanieska Cortina, 2011. Biblioteca Digital “Alma Mater” [en línea]. La Habana: Universidad de las Ciencias Informáticas. Disponible en: http://repositorio_institucional.uci.cu/jspui/bitstream/ident/TD_04955_11/1/TD_04955_11.pdf.

PALENZUELA, Abel Techea, 2010. Oficina Virtual del Servicio Autónomo de Registros y Notarías (SAREN). La Habana: Universidad de las Ciencias Informáticas.

RODRÍGUEZ, Cesar Cuba and RIOS, Julio César Morejón, 2009. Almacenamiento y Gestión de Contenido en el Motor de Clasificación Inteligente de Contenidos (MOCIC). [En línea]. La Habana: Universidad de las Ciencias Informáticas. Disponible en: http://repositorio_institucional.uci.cu/jspui/bitstream/ident/TD_2413_09/1/TD_2413_09.pdf.

SÁNCHEZ, Msc. Ileana Regla Alfonso and VERA, Msc. Ileana Armenteros, 2008. Clasificación de documentos [en línea]. 2008. S.l.: s.n. Disponible en: <http://www.cmiplaya.sld.cu/Archivos/clasificacion.pdf>.

Anexos

Diagrama de clases del análisis Gestionar clasificación.

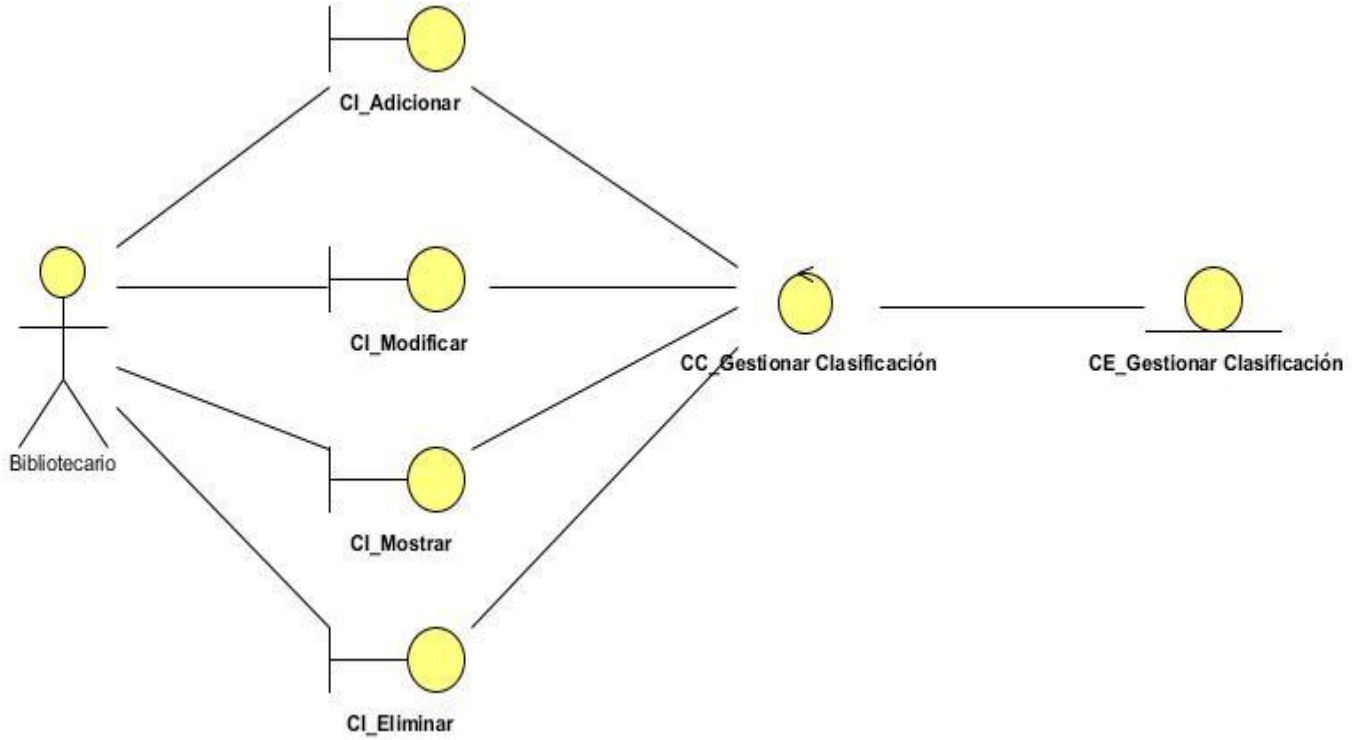


Figura 11: Diagrama de clases del análisis Gestionar clasificación

Diagrama de clases del análisis Configurar módulo

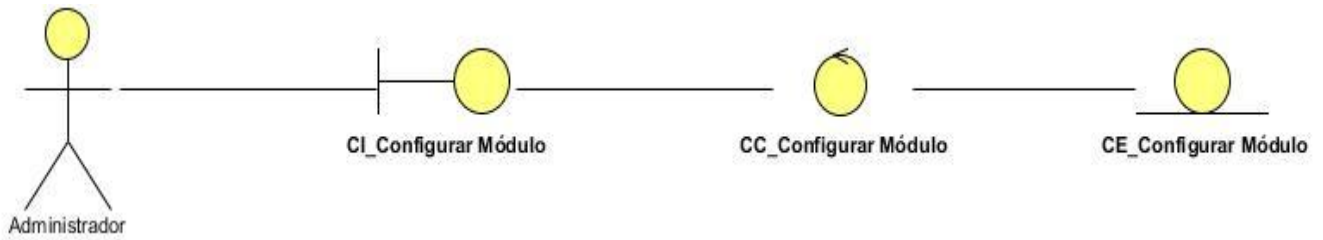


Figura 12: Diagrama de clases del análisis Configurar módulo.

Diagrama de clases del análisis Mostrar documentos clasificados

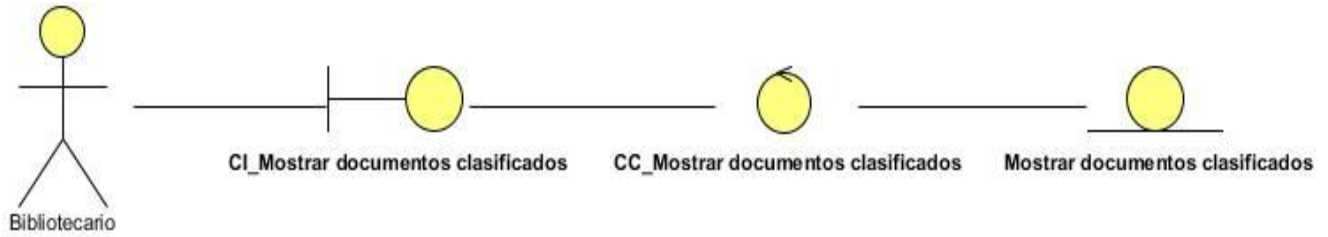


Figura 13: Diagrama de clases del análisis Mostrar documentos clasificados.

Diagrama de clases del análisis Mostrar documentos sin clasificar

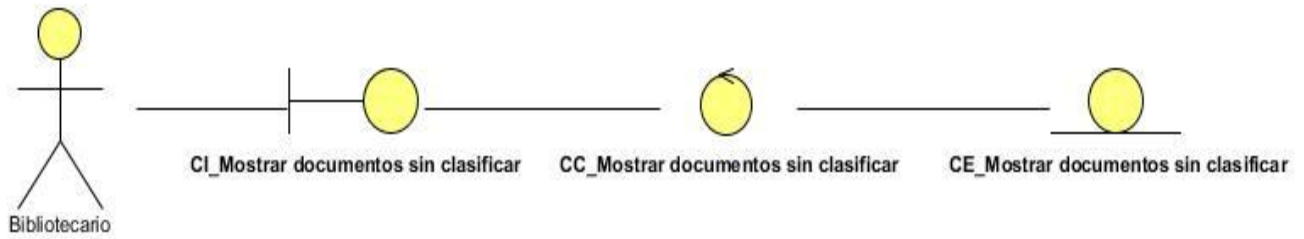


Figura 14: Diagrama de clases del análisis Mostrar documentos sin clasificar.

Diagrama de clases del análisis Generar formulario

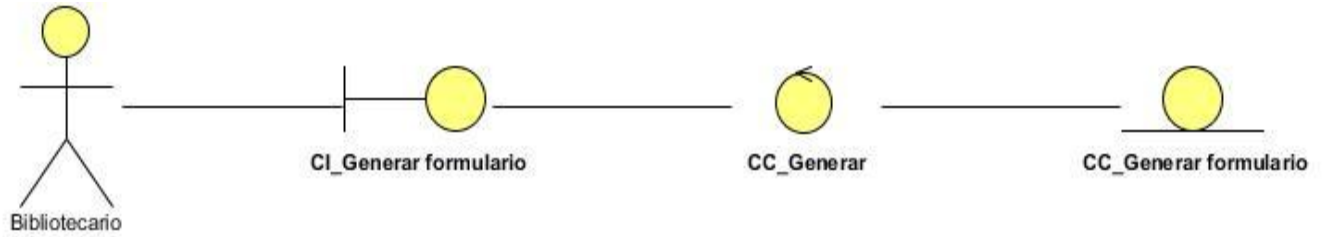


Figura 15: Diagrama de clases del análisis Generar formulario.

Diagrama de clases del análisis Generar base de clasificación

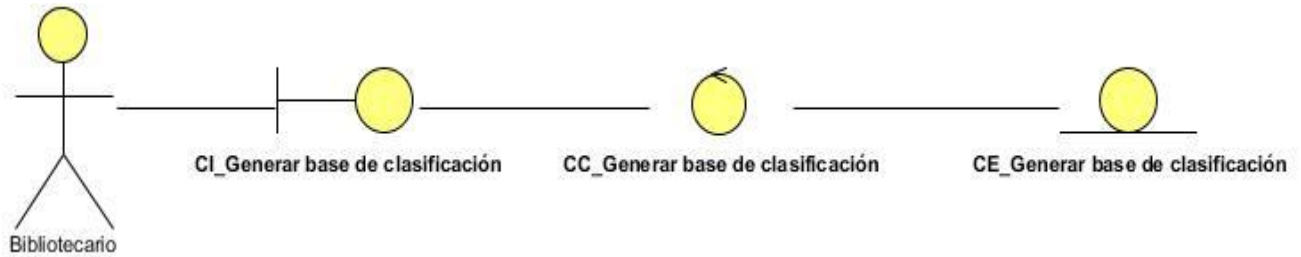


Figura 16: Diagrama de clases del análisis Generar base de clasificación.

Configurar Módulo

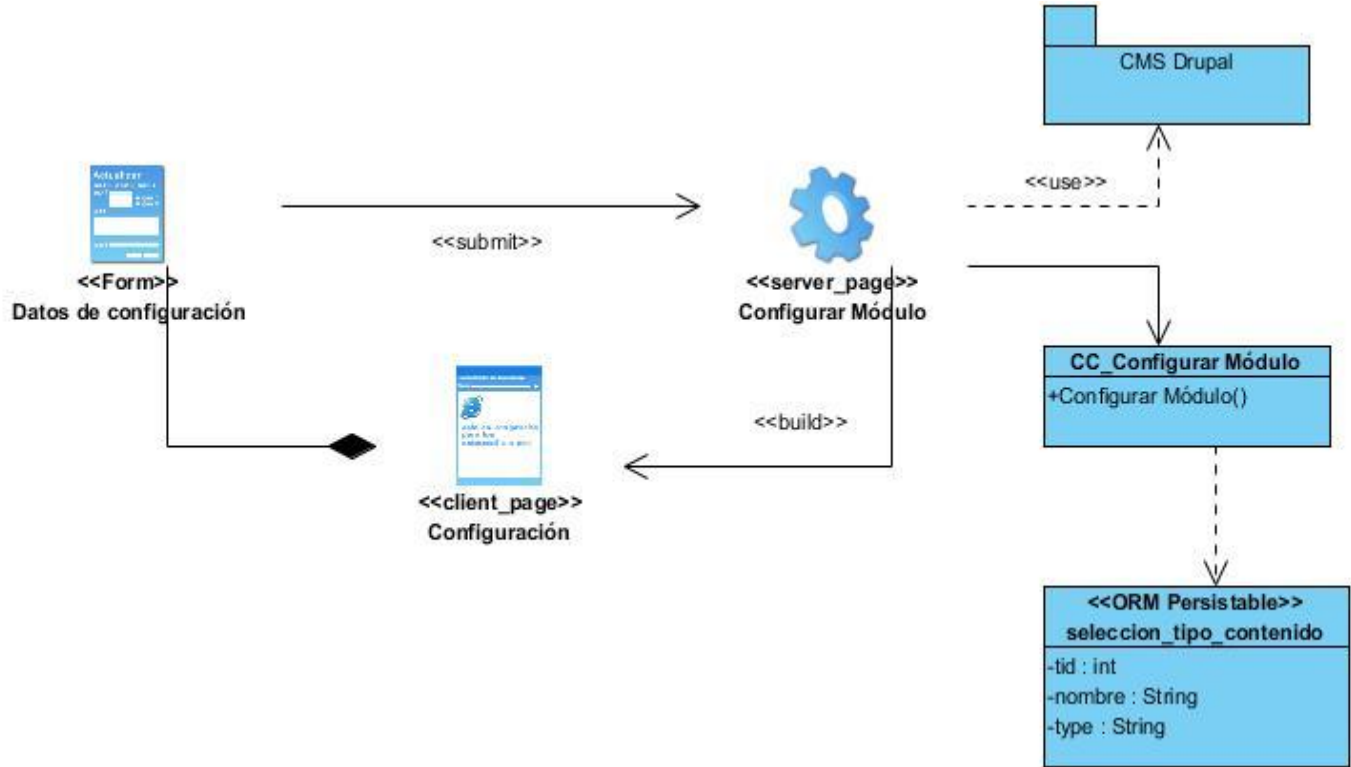


Figura 17: Diagrama de clases del diseño Configurar módulo.

Mostrar documentos sin clasificar

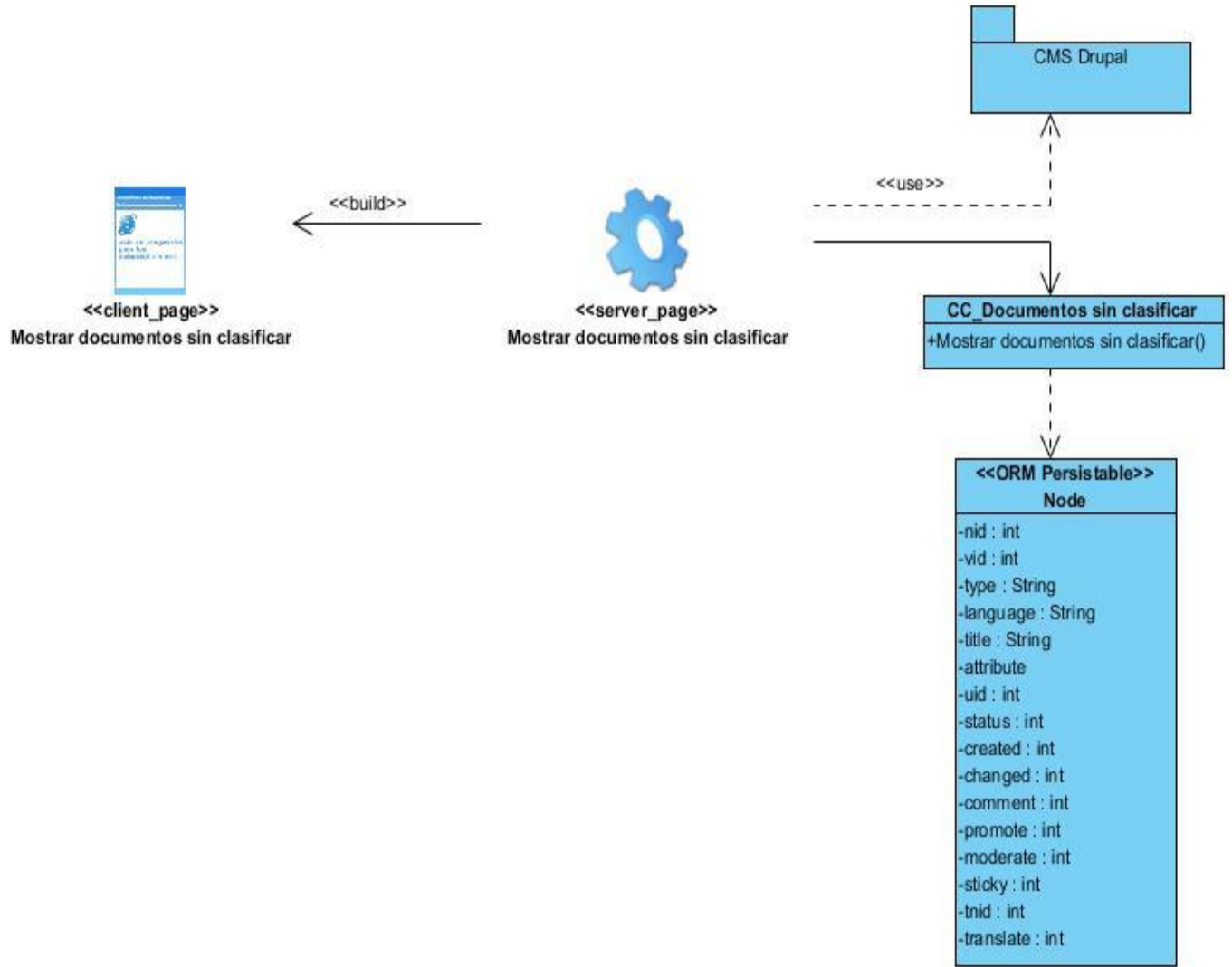


Figura 18: Diagrama de clases del diseño Mostrar documentos sin clasificar.

Mostrar documentos clasificados

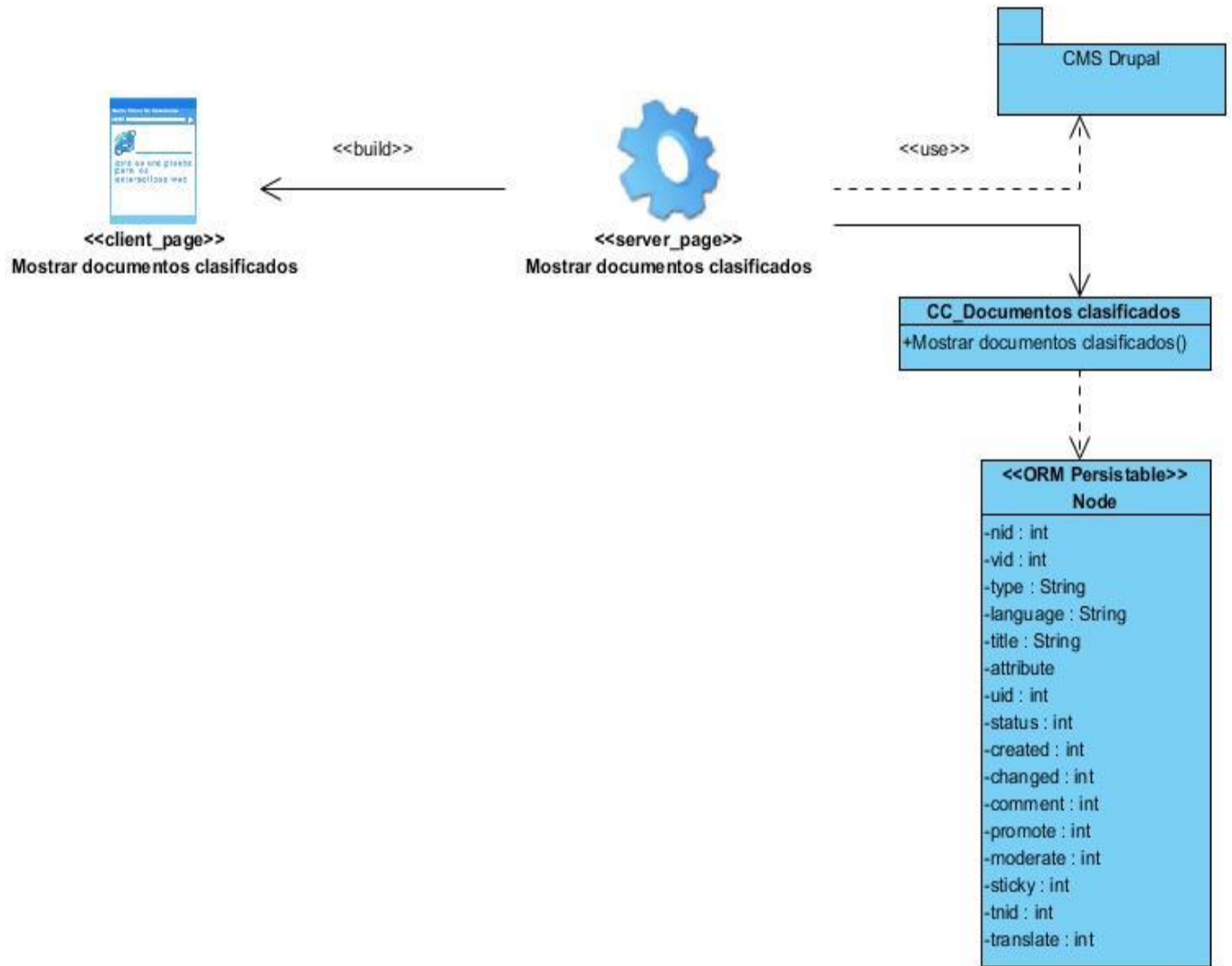


Figura 19: Diagrama de clases del diseño Mostrar documentos clasificados.

Generar textos de entrenamiento

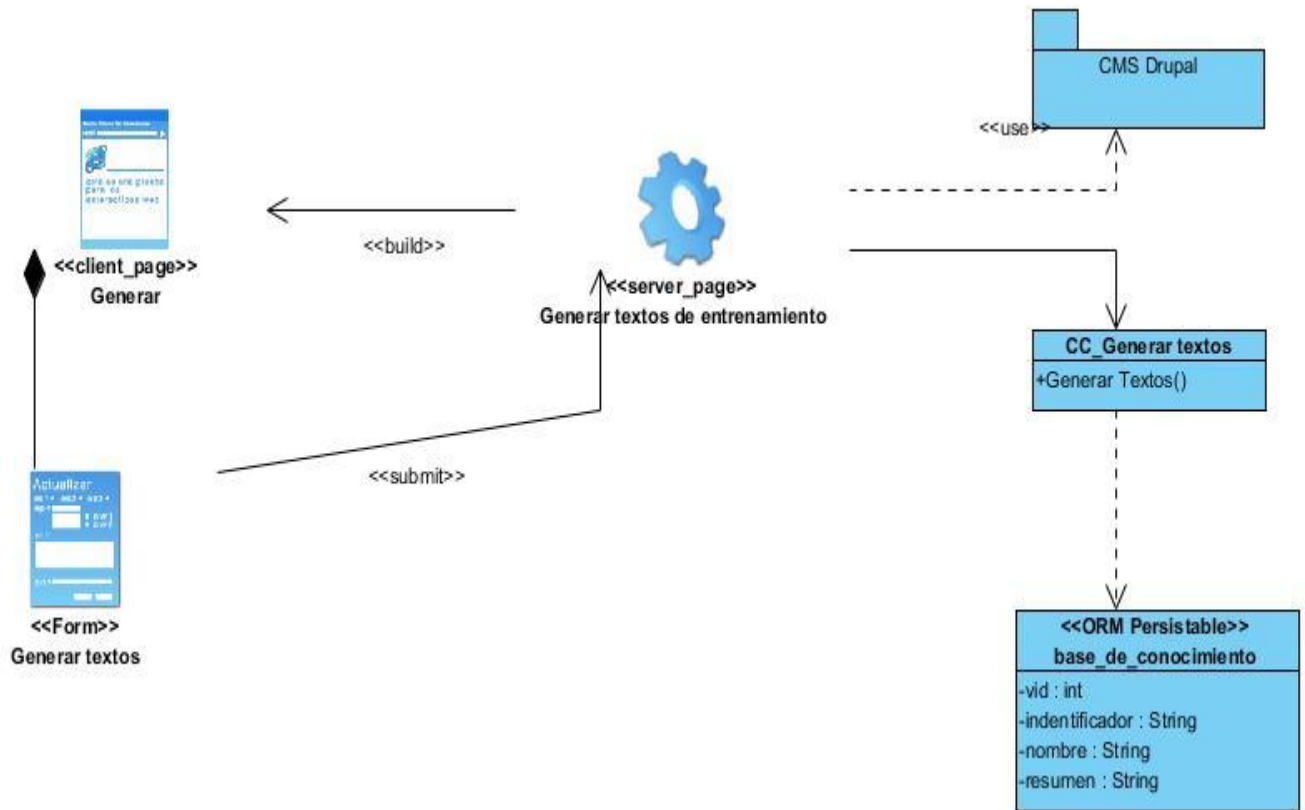


Figura 20: Diagrama de clases del diseño Generar textos de entrenamiento.

Gestionar Clasificación

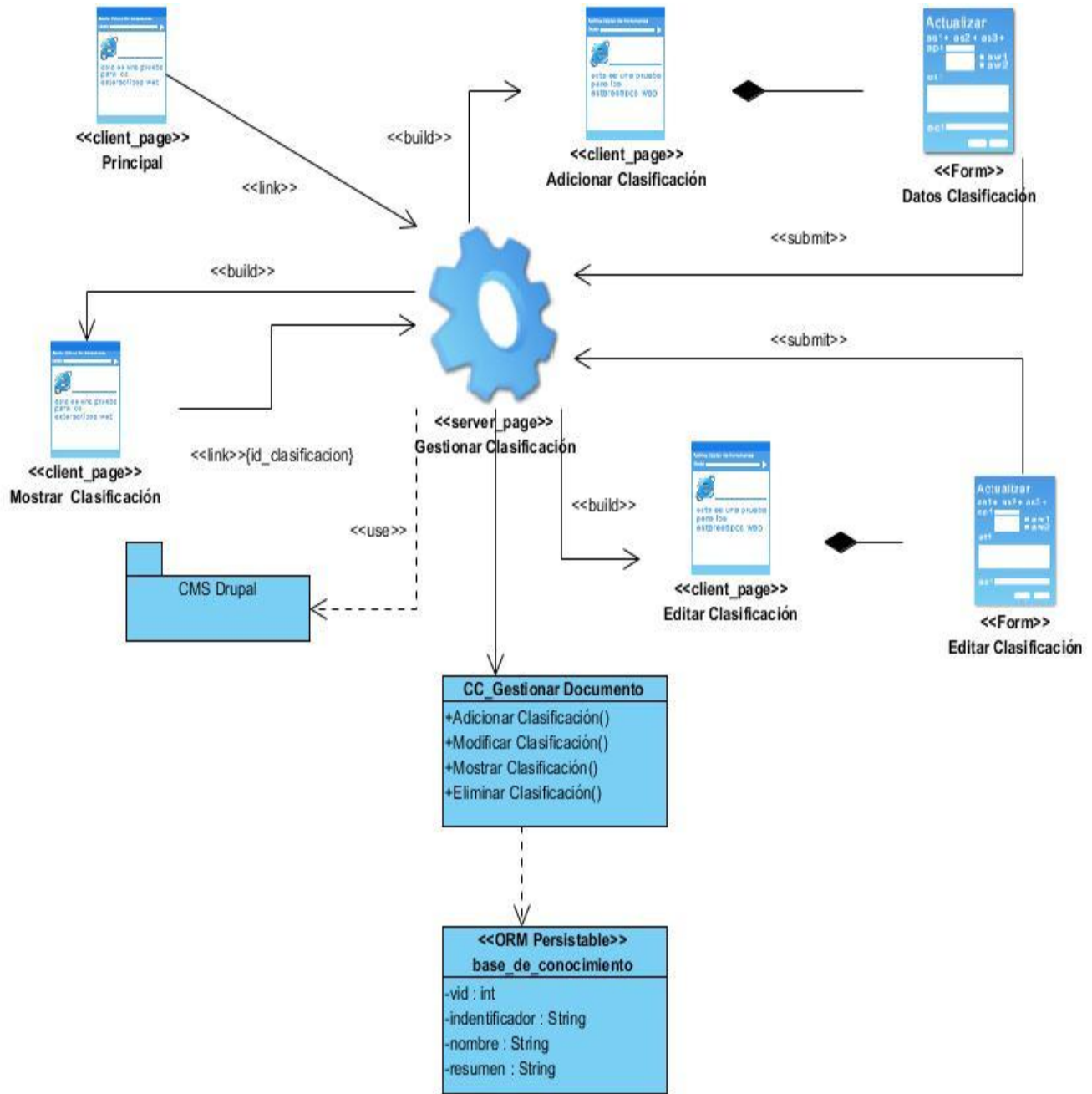


Figura 21: Diagrama de clases del diseño Gestionar clasificación.

Nombre: relación_doc_clas		
Descripción: Tabla para almacenar la relación de las clasificaciones y los documentos		
Atributo	Tipo	Descripción
tid	integer	Identificador de la posición de la tabla
Ident_clas	varchar	Identificador de la clasificación
Ident_doc	varchar	Identificador del documento

Tabla 20: Descripción de la tabla de base de datos relación_doc_clas.

Nombre: selección_tipo_contenido		
Descripción: Tabla para almacenar los tipos de contenidos a seleccionar por el usuario para la clasificación		
Atributo	Tipo	Descripción
tid	integer	Identificador de la posición de la tabla
nombre	varchar	Nombre del tipo de contenido
type	varchar	El tipo de contenido

Tabla 21: Descripción de la tabla de base de datos selección_tipo_contenido.

Nombre: node		
Descripción: Tabla para almacenar los tipos de contenidos a seleccionar por el usuario para la clasificación		

Atributo	Tipo	Descripción
nid	integer	Identificador del nodo
vid	varchar	Identificador de la revisión del nodo. Permite que Drupal guarde los datos del nodo en node_revisions
type	varchar	El tipo de contenido del nodo
language	varchar	Si permite traducción
title	varchar	Título del nodo
uid	integer	Id del usuario que creó el nodo
status	integer	Si la opción de publicado está activada
created	integer	Fecha de cuando fue creado
changed	integer	Última vez que fue modificado
comment	integer	Si la opción de comentario está activado
promote	integer	Si es promovido a la página principal
tnid	integer	Id para la traducción del nodo
translate	integer	Si permite la traducción

Tabla 22: Descripción de la tabla de base de datos node.

Nombre: node_revisions
Descripción: Tabla para almacenar los tipos de contenidos a seleccionar por el usuario para la

clasificación		
Atributo	Tipo	Descripción
nid	integer	Identificador del nodo
vid	varchar	Identificador de la revisión del nodo. Permite que Drupal guarde los datos del nodo en node_revisions
uid	integer	Usuario que creó el nodo
body	varchar	Cuerpo del nodo
teaser	varchar	Toma el punto de acceso al nodo para posibles actualizaciones de forma automática en el mismo nodo.
log	varchar	Un mensaje de log asociado con la revisión creada.
timestamp	integer	El tiempo Unix en el momento que la revisión fue creada.
format	integer	El formato filtrado usado en la revisión.

Tabla 23: Descripción de la tabla de base de datos node_revisions.

Nombre: node_type		
Descripción: Tabla para almacenar los tipos de contenidos a seleccionar por el usuario para la clasificación		
Atributo	Tipo	Descripción

type	varchar	Tipo del nodo.
name	varchar	Nombre del tipo de nodo.
module	varchar	Módulo que creó al nodo.
description	varchar	Descripción del tipo de nodo.
help	varchar	Breve ayuda para el uso del nodo

Tabla 24: Descripción de la tabla de base de datos node_type.

Nombre del requisito	Descripción general	Escenarios de pruebas	Flujo del escenario
2:Modificar clasificación	Se modifica una clasificación previamente seleccionada por el usuario con los campos: 1.1 identificador 2.1 nombre 3.1 resumen	EP 2.1: Modificar una clasificación introduciendo datos válidos.	<ul style="list-style-type: none"> • Se presiona el link Gestionar Clasificación. • Se presiona el link Modificar Clasificación. • Se introducen correctamente los datos. • Se presiona el botón
		EP 2.2: Modificar una clasificación introduciendo datos no válidos.	<ul style="list-style-type: none"> • Se presiona el link Gestionar Clasificación. • Se presiona el link Modificar. • Se introducen incorrectamente los datos

		<p>EP 2.3: Modificar una clasificación dejando campos vacíos.</p>	<ul style="list-style-type: none">• Se presiona el link Gestionar Clasificación.• Se presiona el link Modificar Clasificación.• Se introducen los datos dejando campos vacíos.• Se presiona el botón Aceptar.
		<p>EP 2.4: Modificar una Clasificación con nuevos valores en los campos que ya existen en la biblioteca.</p>	<ul style="list-style-type: none">• Se presiona el link Gestionar Clasificación.• Se presiona el link Modificar Clasificación.• Se modifican los datos con nuevos que ya existen en la biblioteca.• Se presiona el botón Aceptar.
		<p>EP 2.5: Cancelar una modificación</p>	<ul style="list-style-type: none">• Se presiona el link Gestionar Clasificación.• Se presiona el link

			<p>Modificar Clasificación.</p> <ul style="list-style-type: none"> • Se modifican o no los datos. • Se presiona el botón cancelar.
--	--	--	--

Tabla 25: Descripción del caso de prueba para el requisito funcional Modificar clasificación.

No	Nombre de campo	Tipo	Válido	Inválido
1	Identificador	Textfield	Números y puntos.	Letras y caracteres especiales.
2	Nombre	Textfield	Letras y caracteres especiales.	Números
3	Resumen	Textarea	Números, letras y caracteres especiales.	NA

Tabla 26: Descripción de las variables para el requisito funcional Modificar clasificación.

Id del escenario	Escenario	Identificador	Nombre	Resumen	Respuesta del sistema	Resultado de la prueba
EP 2.1	Modificar una clasificación introduciendo datos válidos.	V(112.4)	V(DanielF)	V(DanielF123)	El sistema muestra un mensaje de información de color verde donde explica que el documento ha sido modificado satisfactoriamente.	Satisfactorio

EP 2.2	Modificar una clasificación introduciendo datos no válidos.	I(Daniel2,\$)	I (111., %*)	NA	El sistema muestra un mensaje de color rojo informando que existen datos incorrectos en el campo(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Modificar Clasificación.	Satisfactorio
--------	---	---------------	--------------	----	--	---------------

EP 2.3	Modificar una clasificación dejando campos vacíos.	I(Vacío)	I(Vacío)	I(Vacío)	El sistema muestra un mensaje de color rojo informando que existen campo(s) vacío(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Modificar Clasificación.	Satisfactorio
EP 2.4	Adicionar una Clasificación con datos que ya existe en la biblioteca.	V(111)	V(DanielF)	NA	El sistema muestra un mensaje de color rojo informando que hay datos que ya existen en la biblioteca marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Adicionar Clasificación.	Satisfactorio

EP 2.5	Cancelar una Clasificación.	NA	NA	NA	Se cancela la operación y se redirecciona a la página de inicio.	No satisfactorio

Tabla 27: Descripción del juego a probar para el requisito funcional Adicionar clasificación.

Requisito	No.	No conformidad	Aspecto correspondiente	Etapas de detección	Significativa	No significativa	Recomendación	Estado NC	Respuesta del equipo de desarrollo
Modificar Clasificación	1	El nombre del link para modificar una	Es en la interfaz general, en	1ra		x		Pendiente	Resuelta

		clasificación no es el mismo que se describe en el requisito funcional. El nombre que se muestra en el sistema es "Editar".	el bloque inferior derecho, después de hacer clic en el Gestionar Clasificación						
	2	Cuando se aplica botón cancelar se direcciona para la página prevista	Es en la interfaz de adicionar una clasificación	1ra		x		Pen die nte	Resuelta
	3	Cuando se modifica una clasificación el mensaje de confirmación no tiene punto final	Es en la interfaz de modificar clasificación	1ra		x		Pen die nte	Resuelta

Tabla 28: Descripción del registro de defectos y dificultades detectadas para el requisito funcional Modificar clasificación.

Nombre del requisito	Descripción general	Escenarios de pruebas	Flujo del escenario
3:Configurar Módulo	Se configura el módulo para que este cumpla con las funcionalidades requeridas 4.1 Direccion_servicio 5.1 Direccion_base_casos	EP 3.1: Insertar direcciones con datos válidos.	<ul style="list-style-type: none"> • Se presiona el link Configurar Módulo. • Se introducen correctamente los datos. • Se presiona el botón Aceptar.
		EP 3.2: Insertar direcciones con datos no válidos.	<ul style="list-style-type: none"> • Se presiona el link Configurar Módulo. • Se introducen incorrectamente los datos. • Se presiona el botón Aceptar.
		EP 3.3: Insertar direcciones dejando campos vacíos.	<ul style="list-style-type: none"> • Se presiona el link Configurar Módulo. • Se introducen los datos dejando campos vacíos. • Se presiona el botón Aceptar.

		EP 3.4: Cancelar	<ul style="list-style-type: none"> • Se presiona el link Configurar Módulo. • Se introducen o no los datos. • Se presiona el botón Cancelar.
--	--	------------------	---

Tabla 29: Descripción del caso de prueba para el requisito funcional Configurar módulo.

No	Nombre de campo	Tipo	Válido	Inválido
1	Direccion_servicio	Textfield	Letras, puntos, :, números,/, @	Tildes, %, &, *, (), :
2	Direccion_base_casos	Textfield	Letras, números, /	Tildes, %, &, *, (), :

Tabla 30: Descripción de las variables para el requisito funcional Configurar módulo.

Id del escenario	Escenario	Direccion_Servicio	Direccion_base_casos	Respuesta del sistema	Resultado de la prueba
EP 3.1	Insertar direcciones con datos válidos.	V(http://casode12@prueba.cu)	V(sitios2/mensaje)	El sistema muestra un mensaje de información de color verde donde explica que las direcciones se han insertado satisfactoriamente	Satisfactorio

EP 3.2	Insertar direcciones con datos no válidos.	l(espp@)	l(tildes,*)	El sistema muestra un mensaje de color rojo informando que existen datos incorrectos en el campo(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Configurar Módulo.	Satisfactorio

EP 3.3	Insertar direcciones dejando campos vacíos.	l(Vacío)	l(Vacío)s	El sistema muestra un mensaje de color rojo informando que existen campo(s) vacío(s) marcando estos en color rojo. El sistema regresa automáticamente a la interfaz de Configurar Módulo.	Satisfactorio
--------	---	----------	-----------	---	---------------

Tabla 31: Descripción del juego a probar para el requisito funcional Configurar módulo.

Requisito	No.	No conformidad	Aspecto correspondiente	Etapa de detección	Significativa	No significativa	Recomendación	Estado NC	Respuesta del equipo de desarrollo
Configurar Módulo	1	La palabra clasificación no tiene tilde	Es en la interfaz general, en el bloque inferior derecho,	1ra	x			Pendiente	Resuelta

			después de hacer clic en el Configurar Módulo.						
--	--	--	---	--	--	--	--	--	--

Tabla 32: Descripción del registro de defectos y dificultades detectadas para el requisito funcional Configurar Módulo.