

UCLM-ESI
PROYECTO FIN DE CARRERA:

TÍTULO	Sistema para el almacenamiento búsqueda y recuperación de documentos científicos
ALUMNO	PABLO MORALEDA CALVO
DIRECTOR	FRANCISCO RUIZ GONZÁLEZ
NÚMERO	17-01
FECHA	2001, Julio
CALIFICACIÓN	MATRÍCULA DE HONOR

RESUMEN:

El acceso a los datos es fundamental, y a veces crítico, en los procesos de negocio. Por esta razón, es de máxima importancia disponer de sistemas que automaticen el almacenamiento y la recuperación de la información de una manera eficiente y sencilla.

Dependiendo del tipo de información que manejan, existen dos tipos de sistemas que realizan estas tareas: sistemas para gestión de información estructurada (Sistemas Gestores de Bases de Datos) y sistemas para gestión de información no-estructurada (Bibliotecas Digitales, Sistemas de Información Bibliográfica, Sistemas de Información Documental...). Los sistemas basados en información no-estructurada almacenan los nodos de información (documentos de texto) de manera que se facilite la recuperación de los mismos mediante consultas sencillas, normalmente utilizando aproximaciones al lenguaje natural. Además, con la aparición de Internet y la facilidad de acceso a la información de forma distribuida, se amplían las posibilidades de construir sistemas que gestionen la información no-estructurada utilizando el soporte web para el acceso a los documentos de forma remota.

El presente proyecto pretende construir una aplicación con soporte para Web que permita almacenar, buscar y recuperar documentos por grupos de usuarios. Por sus características especiales, está especialmente orientada a grupos de investigación.

Abstract:

Access to data has always been fundamental and sometimes critical in business processes. For this reason, it 's extremely important the availability of systems to automate store and retrieval information in a simple and efficient way.

Depending the kind of information they use, there are two types of systems that carry out these tasks: structured information management systems (Database Management Systems, DBMS) and non-structured information management systems (Digital Libraries, Bibliographic Information Systems...). Systems based upon non-structured information store information nodes (text documents) so that retrieval them can be easily carried out through a simple query, usually, approximations to natural language. Furthermore, with the outbreak of the Internet and the easy access to information in a distributed way, the possibilities of building systems that manage non-structured information in the web to have access to remote documents widen.

The purpose of this project is to build an application in the web that allows to store, search and retrieval scientific documents by users groups. Because of its especial features, it 's especially oriented to researches groups.

ÍNDICE:

1. INTRODUCCIÓN	Página 1
1.1. INTRODUCCIÓN.....	Página 1
1.2. ENTORNO.....	Página 4
1.3. OBJETIVOS.....	Página 4
1.4. ENFOQUE.....	Página 5
1.6. ESTRUCTURA DEL TRABAJO.....	Página 5
2. ANTECEDENTES Y ESTADO DEL ARTE	Página 7
2.1. SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN.....	Página 7
2.1.1. Estructura de un sistema de recuperación de información.....	Página 8
2.1.2. Modelos de recuperación de información.....	Página 9
2.1.3. Documentos.....	Página 12
2.1.3.1. <i>Documentos Subrogados</i>	Página 13
2.1.3.2. <i>Metadatos</i>	Página 14
2.1.3.3. <i>Lenguajes de Marcas</i>	Página 14
2.1.4. Lenguajes y Tipos de Consultas.....	Página 15
2.1.4.1. <i>Consultas Booleanas</i>	Página 16
2.1.4.2. <i>Consultas Vectoriales</i>	Página 18
2.1.4.3. <i>Consultas en Lenguaje Natural</i>	Página 20
2.1.5. Indexación de Documentos.....	Página 20
2.1.5.1. <i>Ficheros Invertidos</i>	Página 21
2.1.5.2. <i>Árboles y Arrays de Sufijos</i>	Página 23
2.1.5.3. <i>Ficheros de Signatura</i>	Página 25
2.1.6. Visualización de Resultados.....	Página 26
2.2: BUSCADORES DE RECURSOS EN EL WEB.....	Página 27
2.2.1. Motores de Búsqueda.....	Página 28
2.3. BASES DE DATOS DOCUMENTALES COMERCIALES.....	Página 32
3. OBJETIVOS DEL PROYECTO	Página 35
4. HIPÓTESIS DE TRABAJO	Página 36
5. METODOLOGÍA Y RESULTADOS	Página 37
5.1. MÉTODO DE TRABAJO.....	Página 37
5.1.1. Introducción al Lenguaje Unificado de Modelado.....	Página 37
5.1.1.1. <i>Elementos</i>	Página 38
5.1.1.1.1. Elementos Estructurales.....	Página 38
5.1.1.1.2. Elementos de Comportamiento.....	Página 40
5.1.1.1.3. Elementos de Agrupación.....	Página 41
5.1.1.1.4. Elementos de Anotación.....	Página 41
5.1.1.2. <i>Relaciones</i>	Página 41
5.1.1.3. <i>Diagramas</i>	Página 42
5.1.1.3.1. Diagramas Estructurales.....	Página 43
5.1.1.3.2. Diagramas de Comportamiento.....	Página 43
5.1.2. Introducción al Proceso Unificado de Software.....	Página 45
5.1.2.1. <i>Dirigido por Casos de Uso</i>	Página 45
5.1.2.1.1. La captura de los casos de uso.....	Página 46
5.1.2.1.2. Análisis, Diseño e Implementación para realizar los casos de uso.....	Página 46
5.1.2.1.3. Prueba de los casos de uso.....	Página 48
5.1.2.2. <i>Centrado en la Arquitectura</i>	Página 49
5.1.2.3. <i>Iterativo e Incremental</i>	Página 49
5.1.2.4. <i>La vida del Proceso Unificado</i>	Página 51

5.1.2.4.1. El Producto	Página 51
5.1.2.4.2. Fases dentro de un ciclo	Página 51
5.2. ANÁLISIS Y DISEÑO DE LA HERRAMIENTA	Página 52
5.2.1. Especificación de requisitos candidatos	Página 52
5.2.2. Modelos de Casos de Uso	Página 54
5.2.2.1. <i>Modelo de casos de uso para el subsistema cliente</i>	Página 54
5.2.2.1.1. Descripción general	Página 55
5.2.2.1.2. Identificación y descripción de cada caso de uso	Página 56
5.2.2.2. <i>Modelo de casos de uso para el subsistema servidor</i>	Página 58
5.2.2.2.1. Diagrama de casos de uso	Página 58
5.2.2.2.2. Descripción general	Página 60
5.2.3. Estructura del sistema. Relación entre subsistemas y componentes	Página 62
5.2.4. Diagrama de tres niveles	Página 63
5.2.5. Diagrama de clases	Página 66
5.2.6. Modelo de seguridad del sistema	Página 67
5.2.7. Diseño de la base de datos del sistema	Página 69
5.2.7.1. <i>Modelo conceptual de datos</i>	Página 69
5.2.7.2. <i>Modelo de datos lógico</i>	Página 72
5.3. IMPLEMENTACIÓN	Página 73
5.3.1. Entorno de programación de la aplicación	Página 73
5.3.2. Almacenamiento e indexación de documentos	Página 74
5.3.3. Presentación de la información	Página 75
5.3.4. Gestión y administración del sistema	Página 76
5.3.4.1. <i>Esquema físico de la BD de Gestión</i>	Página 77
5.4. ENTORNO TECNOLÓGICO	Página 79
5.4.1. Aplicaciones WEB	Página 79
5.4.1.1. <i>Arquitectura Windows DNA</i>	Página 82
5.4.2. Active Server Pages	Página 82
5.4.2.1. <i>Funcionamiento de las páginas ASP</i>	Página 84
5.4.2.1.1. Introducción a los objetos de ASP	Página 86
5.4.2.1.1.1. <i>El objeto Request</i>	Página 87
5.4.2.1.1.2. <i>El objeto Response</i>	Página 88
5.4.2.1.1.3. <i>El objeto Server</i>	Página 90
5.4.2.1.1.4. <i>El objeto Application</i>	Página 92
5.4.2.1.1.5. <i>El objeto Session</i>	Página 93
5.4.2.1.1.6. <i>Directivas ASP</i>	Página 95
5.4.3. Modelo de Objetos Componentes (COM)	Página 95
5.4.4. Tecnología XML	Página 97
5.4.4.1. <i>Introducción a los lenguajes de marcas</i>	Página 97
5.4.4.2. <i>Estándar Generalized Markup Language (SGML)</i>	Página 98
5.4.4.3. <i>eXtended Markup Language (XML)</i>	Página 98
5.4.4.3.1. Estructura y contenido de un documento XML	Página 99
5.4.4.3.1.1. <i>Elementos</i>	Página 100
5.4.4.3.1.2. <i>Prólogo</i>	Página 102
5.4.4.3.1.3. Atributos	Página 102
5.4.4.3.2. Integración de XML con HTML	Página 103
5.4.4.4. <i>Document Object Model (DOM)</i>	Página 104
5.4.4.5. <i>XML Data Reduced Schemas</i>	Página 107
5.4.4.6. <i>Extended Stylesheet Language (XSL)</i>	Página 110
5.4.5. Consultas de texto mediante el Servicio de Indización	Página 112
5.4.5.1. <i>Introducción al Servicio de Indización de Microsoft</i>	Página 112
5.4.5.1.1. Catálogos	Página 114
5.4.5.1.2. Indexación	Página 115
5.4.5.1.2.1. <i>Filtros de documentos</i>	Página 115
5.4.5.1.2.2. <i>Caché de propiedades</i>	Página 116

5.4.5.1.2.3. <i>Listas de palabras</i>	Página 116
5.4.5.2. <i>Búsquedas en el catálogo de indización</i>	Página 117
5.4.6. Active Data Object (ADO).....	Página 120
5.4.6.1. <i>El objeto Connection</i>	Página 121
5.4.6.2. <i>El objeto Command</i>	Página 122
5.4.6.3. <i>El objeto Recorset</i>	Página 123
5.4.6.4. <i>El objeto Record</i>	Página 123
5.4.6.5. <i>El objeto Stream</i>	Página 123
5.4.6.6. <i>La colección Fields</i>	Página 124
5.4.6.7. <i>La colección Parameters</i>	Página 124
5.4.6.8. <i>La colección Errors</i>	Página 124
5.5. MANUAL DE USUARIO.....	Página 124
5.5.1. Consideraciones generales.....	Página 125
5.5.2. Subsistema cliente.....	Página 125
5.5.2.1. <i>Estructura del interfaz de usuario</i>	Página 126
5.5.2.2. <i>Identificarse en el sistema</i>	Página 127
5.5.2.3. <i>Desconexión</i>	Página 128
5.5.2.4. <i>Buscar documentos</i>	Página 129
5.5.2.5. <i>Recuperar documento</i>	Página 131
5.5.6.2. <i>Almacenar documento</i>	Página 133
5.5.3. Subsistema servidor.....	Página 135
5.5.3.1. <i>Gestión de usuarios</i>	Página 138
5.5.3.1.1. <i>Añadir un nuevo usuario</i>	Página 139
5.5.3.1.2. <i>Modificar usuario</i>	Página 139
5.5.3.1.3. <i>Eliminar Usuario</i>	Página 140
5.5.3.1.4. <i>Buscar Usuario</i>	Página 140
5.5.3.2. <i>Gestión de definiciones de tipos de documentos</i>	Página 141
5.5.3.2.1. <i>Visualizar definiciones de tipos de documentos</i>	Página 142
5.5.3.2.2. <i>Añadir nuevo tipo de documento al sistema</i>	Página 143
5.5.3.2.3. <i>Eliminar definición de tipo existente</i>	Página 144
5.5.3.3. <i>Gestión de la seguridad del sistema</i>	Página 144
5.5.3.3.1. <i>Añadir nuevo nivel de usuario</i>	Página 145
5.5.3.3.2. <i>Eliminar nivel de usuario</i>	Página 146
5.5.3.3.3. <i>Añadir nuevo nivel de documento</i>	Página 147
5.5.3.3.4. <i>Eliminar nivel de documento</i>	Página 147
5.5.3.3.5. <i>Añadir nuevo privilegio</i>	Página 147
5.5.3.3.6. <i>Eliminar privilegio</i>	Página 149
5.5.3.3.7. <i>Visualizar y modificar restricciones de privilegios</i>	Página 149
5.5.3.3.8. <i>Modificar asignaciones de privilegios</i>	Página 150
5.5.3.4. <i>Gestión de estilos de visualización</i>	Página 151
5.5.3.4.1. <i>Añadir nuevo estilo</i>	Página 152
5.5.3.4.2. <i>Eliminar Estilo</i>	Página 152
5.5.3.4.3. <i>Visualizar Estilo</i>	Página 152
6. CONCLUSIONES Y PROPUESTAS	Página 153
6.1. Conclusiones.....	Página 153
6.2. Propuestas.....	Página 153
5.3. Líneas Futuras.....	Página 154
7. BIBLIOGRAFÍA	Página 155
ANEXO A. DIAGRAMAS ADICIONALES UML	Página 158
A.1. CAPTURA DE REQUISITOS COMO CASOS DE USO.....	Página 158
A.1.1 Diagramas de estados para cada caso de uso.....	Página 158
A.2. ANÁLISIS.....	Página 160

A.2.1 Diagramas de colaboración de la realización de cada caso de uso	Página 160
A.3. DISEÑO	Página 162
A.3.1 Diagramas de secuencia de la realización de cada caso de uso.....	Página 162
ANEXO B. ACRÓNIMOS.....	Página 165
ANEXO C. EJEMPLOS DE CÓDIGO FUENTE.....	Página 167
C.1. CÓDIGO FUENTE DE LA PÁGINA ASP <i>Rbuscar.asp</i>	Página 168
C.2. CÓDIGO FUENTE DE LA CLASE <i>Schemas</i>	Página 176
C.3. CÓDIGO FUENTE DE LA CLASE <i>Elementos</i>	Página 178
C.4. CÓDIGO FUENTE DE LA CLASE <i>Atributos</i>	Página 180