



RESUMEN – TESIS DE GRADO

AUTORES PEDRO ANTONIO CHACÓN GARNICA

FACULTAD INGENIERIA

PLAN DE ESTUDIOS SISTEMAS

DIRECTOR LUIS IGNACIO LIZCANO BUENO

TITULO DE LA TESIS ANÁLISIS, DISEÑO E IMPLEMENTACION DE UN
INDIZADOR AUTOMÁTICO, PARA UN SISTEMA DE RECUPERACIÓN DE LA
INFORMACIÓN.

RESUMEN

La recuperación de la información es una disciplina de creciente interés, que se encarga del manejo de grandes volúmenes de información, cuando existe la necesidad de consulta por parte del usuario, el Sistema se encarga de entregar a este los artículos que más correspondan a su consulta. Dentro de esta disciplina existen conceptos como, indización automática, la cual es el motor del Sistema de Recuperación de Información, y consiste en una serie de procesos que se encargan de seleccionar los índices que permiten caracterizar el contenido de un documento, este trabajo trata sobre la Indización automática que hace parte del Sistema de Recuperación de Información.

CARACTERÍSTICA

PAGINAS 104 PLANOS X ILUSTRACIONES 22 CD-ROM 1

**ANÁLISIS, DISEÑO E IMPLEMENTACION DE UN INDIZADOR
AUTOMÁTICO, PARA UN SISTEMA DE RECUPERACIÓN DE LA
INFORMACIÓN**

PEDRO ANTONIO CHACON GARNICA

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER
FACULTAD DE INGENIERIAS
PLAN DE ESTUDIOS DE INGENIERIA DE SISTEMAS
SAN JOSE DE CUCUTA
2003**

**ANÁLISIS, DISEÑO E IMPLEMENTACION DE UN INDIZADOR
AUTOMÁTICO, PARA UN SISTEMA DE RECUPERACIÓN DE LA
INFORMACIÓN**

PEDRO ANTONIO CHACON GARNICA

**Anteproyecto de grado presentado como requisito para optar al título de Ingeniería de
Sistemas.**

**Director:
LUIS IGNACIO LIZCANO
Licenciado en Matemáticas**

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER
FACULTAD DE INGENIERIAS
PLAN DE ESTUDIOS DE INGENIERIA DE SISTEMAS
SAN JOSE DE CUCUTA
2003**



ACTA DE SUSTENTACION DE UN TRABAJO DE GRADO

FECHA : 8 DE AGOSTO DE 2003 HORA : 10:00 a. m.

LUGAR : AULA 4 - TERCER PISO - EDIFICIO CREAD

PLAN DE ESTUDIOS: INGENIERIA DE SISTEMAS

TITULO DE LA TESIS: ANALISIS, DISEÑO E IMPLEMENTACION DE UN INDIZADOR AUTOMATICO PARA UN SISTEMA DE RECUPERACION DE LA INFORMACION.

JURADOS : MERY YANETH SARMIENTO SAAVEDRA
MYRIAM DEISY GARCIA
JEFFREY OMAR ARDILA

DIRECTOR : MAGISTER LUIS IGNACIO LIZCANO BUENO

NOMBRE DE LOS ESTUDIANTES:	CODIGO	CALIFICACION	
		NUMERO	LETRA
PEDRO ANTONIO CHACON GARNICA	151908	4,3	CUATRO, TRES

APROBADA

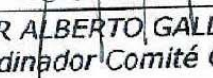
FIRMA DE LOS JURADOS


MERY YANETH SARMIENTO


MYRIAM DEISY GARCIA


JEFFREY OMAR ARDILA

Vo.Bo.


OSCAR ALBERTO GALLARDO PEREZ
Coordinador Comité Curricular

Betty M.

Dedico este triunfo en forma especial a Daniela y Pedro mis Hijos, Claudia mi esposa, Julieta Garnica mi madre, Doris Zaraza mi suegra; quienes han estado con migo en cada difcil momento al cual he tenido que enfrentar “Muchas Gracias”.

AGRADECIMIENTOS

Este proyecto se inicio a partir de una propuesta hecha por el Magister Luis Ignacio Lizcano Bueno dentro de la materia Recuperación de la Información a la cual tuve oportunidad de asistir. Estoy muy agradecido a Luis Ignacio por su gran ayuda durante todo el proceso de iniciación y culminación del proyecto “Gracias Profe”. Además, debo mencionar la generosa y desinteresada ayuda de mi compañero y amigo Hilmer Chona quien hizo parte del grupo de investigación que lidero el proyecto Sistema de Recuperación de la Información.

CONTENIDO

	pág.
INTRODUCCIÓN	15
1. PLANTEAMIENTO DEL PROBLEMA	16
1.1 TITULO DEL PROYECTO	16
1.2 DESCRIPCIÓN Y PLANTEAMIENTO DEL PROBLEMA	16
1.3 OBJETIVOS	16
1.3.1 Objetivo general	16
1.3.2 Objetivos específicos	16
1.4 JUSTIFICACIÓN DEL PROBLEMA	17
1.5 ALCANCES Y LIMITACIONES	18
1.5.1 Alcances	18
1.5.2 Limitaciones	19
2. MARCO TEÓRICO	20
2.1 INTRODUCCIÓN	20

2.2 ARQUITECTURA DE UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN (SRI)	20
2.2.1 Análisis documental	21
2.3 LA INDIZACIÓN	22
2.3.1 Indización libre	22
2.3.2 Indización controlada	22
2.3.3 Identificación automática de las entradas, la palabra como objeto	22
2.3.4 La función de las palabras	23
2.4 LA COLECCIÓN	23
2.5 LENGUAJE DE PROGRAMACIÓN	24
2.6 MODELO VECTORIAL	24
2.7 FORMATEADO DE TEXTO	26
2.7.1 Análisis léxico del texto	27
2.7.2 Eliminación de palabras vacías	38
2.7.3 Aplicación de lematización	38
2.7.4 Construcción de tesauros	38

2.8 ALMACENAMIENTO, INDEXADO Y BÚSQUEDA DE INFORMACIÓN	39
3. ARQUITECTURA DEL SISTEMA EN LA METODOLOGÍA ORIENTADA A OBJETOS	40
3.1 INTRODUCCIÓN (VISIÓN GENERAL)	40
3.2 ANÁLISIS ORIENTADO A OBJETOS	40
3.2.1 Definición del problema	40
3.2.2 Modelado de objetos	41
3.2.3 Modelado dinámico	46
3.2.4 Modelado funcional	54
3.3 DISEÑO DEL SISTEMA	58
3.3.1 Arquitectura del sistema	58
3.3.2 Concurrencia	59
3.3.3 Almacenes de datos	60
3.4 DISEÑO DE OBJETOS	60
3.4.1 Diseño de algoritmos para los métodos de las clases	62
4. ESPECIFICACIÓN DEL SISTEMA	69

4.1 LA COLECCIÓN	69
4.2 SUGERENCIAS PARA LA UTILIZACIÓN DEL SISTEMA	69
4.2.1 Sugerencias generales	69
4.2.2 Como ingresar al sistema de indización automática	70
5. CONCLUSIONES	71
6. RECOMENDACIONES	73
BIBLIOGRAFÍA	74
ANEXOS	76