



**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER  
BIBLIOTECA EDUARDO COTE LAMUS**



## **RESUMEN - TESIS DE GRADO**

**AUTORES** DIEGO ARMANDO HERNANDEZ PEREZ

**FACULTAD** INGENIERIAS

**PLAN DE ESTUDIOS** INGENIERIA DE SISTEMAS

**DIRECTOR** Mg LUIS IGNACIO LIZCANO BUENO

**TÍTULO DE LA TESIS** DISEÑO E IMPLEMENTACION DE UN LEMATIZADOR  
SINTACTICO EN CASTELLANO PARA UN SISTEMA DE RECUPERACION DE  
INFORMACION

### **RESUMEN**

Este proyecto consiste en el diseño e implementación de un Lematizador en castellano, que permita extracción de raíces de palabras en un documento o una consulta en la formación de términos en el modelo interno del sistema, ofreciendo los servicios de conjugación de verbos y silabeo y lematización de términos y que esté disponible en Internet.

El algoritmo de extracción de raíces de los términos se encuentra orientado a obtener un único término. El resultado del algoritmo debe ser una misma forma canónica para las diferentes variaciones morfológicas de una palabra, que no tiene porque ser necesariamente, la raíz lingüística.

El sistema se ha desarrollado en base a la integración de las siguientes tecnologías web: Java Server Pages (JSP), Javascript y PostgreSQL. Disponible en la dirección : <http://200.75.48.172:8080/LESP>.

### **CARACTERISTICAS**

PAGINAS 124 PLANOS        ILUSTRACIONES        CD ROM   1

**DISEÑO E IMPLEMENTACION DE UN LEMATIZADOR SINTACTICO EN  
CASTELLANO PARA UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN**

**DIEGO ARMANDO HERNÁNDEZ PÉREZ**

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER  
FACULTAD DE INGENIERIAS  
PLAN DE ESTUDIOS DE INGENIERIA DE SISTEMAS  
SAN JOSÉ DE CÚCUTA**

**2003**

**DISEÑO E IMPLEMENTACION DE UN LEMATIZADOR SINTACTICO EN  
CASTELLANO PARA UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN**

**DIEGO ARMANDO HERNÁNDEZ PÉREZ**

**Proyecto de grado presentado como requisito para optar al título de Ingeniero de  
Sistemas**

**Director:**

**LUIS IGNACIO LIZCANO B.**

**Candidato a Phd en Informatica de la UCM**

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER**

**FACULTAD DE INGENIERIAS**

**PLAN DE ESTUDIOS DE INGENIERIA DE SISTEMAS**

**SAN JOSÉ DE CÚCUTA**

**2003**



UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERIA



ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

FECHA: Cúcuta, 27 de noviembre de 2003  
HORA: 14:00  
LUGAR: Auditorio Biblioteca "Eduardo Cote Lamus"  
Plan de estudio: INGENIERÍA DE SISTEMAS

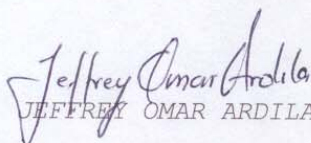
Título de la tesis: "DISEÑO E IMPLEMENTACIÓN DE UN LEMATIZADOR SINTÁCTICO EN CASTELLANO PARA UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN"


Jurados: MERY YANETH SARMIENTO  
MYRIAM DEISY GARCÍA  
JEFFREY OMAR ARDILA GALVIS

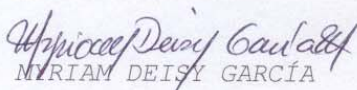
Director: LUIS IGNACIO LIZCANO BUENO

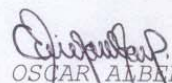
Nombre de los estudiantes	Código	Calificación Letra	Número
DIEGO ARMANDO HERNÁNDEZ PÉREZ	151040	Cuatro, cinco	4,5

M E R I T O R I A

  
JEFFREY OMAR ARDILA GALVIS

  
MERY YANETH SARMIENTO

  
MYRIAM DEISY GARCÍA

  
Vo.Bo. OSCAR ALBERTO GALLARDO PÉREZ  
Coordinador Comité Curricular  
Ingeniería de Sistemas

Jeannette C.

Avenida Gran Colombia No. 12e-96 B Colsag Tel. 5753515 - 5776655 Fax (97) 5771988  
CUCUTA - COLOMBIA

A Dios, por brindarme la sabiduría necesaria para alcanzar esta meta y acompañarme en cada momento y permitirme contar con el apoyo de mi familia.

A mi madre Anayibe, a quien que tanto quiero y le debo todo mi ser. Por haber creído en mi, brindándome su amor, su comprensión y su apoyo.

A mi abuela Ermelina, a quien que tanto quiero y quien creyendo en mi con un profundo e incomparable amor, ha dado más de lo que ha podido para ayudarme a cumplir todas mis metas.

A mis hermanas Martha y Diana quienes son fuente interminable de inspiración y compromiso para luchar en la vida.

A mi padre Numa, por todas sus enseñanzas, por enseñarme el valor del trabajo honesto y por darme todo lo que me ha dado.

A mi flaquita Denis, quien con su presencia cambio mi vida de la forma que nadie más hubiese podido. Por su compañía, su apoyo incondicional y sus consejos.

A mis compañeros y amigos, quienes con sus consejos y su ayuda contribuyeron para hacer realidad esta etapa de mi vida.

***Diego***

## **AGRADECIMIENTOS**

El autor expresa sus Agradecimientos a:

Al Magíster Luis Ignacio Lizcano Bueno, por su valiosa colaboración como tutor y director del proyecto, a sus valiosos esfuerzos por expandir sus conocimientos y nuestras ideas a otras instituciones.

A los Ingenieros Fredy Vargas, Leonardo Ibarra, Fabián Roperó, Alvaro Martínez y Jesus Roperó, mis amigos, por toda su colaboración y apoyo incondicional

A la Universidad Francisco de Paula Santander, por la formación y los recursos ofrecidos para llevar a cabo el desarrollo de nuestro trabajo.

Todas aquellas personas que de una u otra forma colaboraron para la realización de este proyecto.

## CONTENIDO

	<b>Pág.</b>
INTRODUCCION	16
1. PLANTEAMIENTO DEL PROBLEMA	17
1.1 TITULO DEL PROYECTO	17
1.2 DESCRIPCION Y PLANTEAMIENTO DEL PROBLEMA	17
1.3 ANTECEDENTES	17
1.4 JUSTIFICACION	18
1.5 OBJETIVOS	18
1.5.1 Objetivo General	18
1.5.2 Objetivos Específicos	19
1.6 ALCANCES Y LIMITACIONES	19
1.6.1 Alcances	19
1.6.2 Limitaciones	19
2. MARCO TEÓRICO	21

2.1 INTRODUCCION	21
2.2 SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN	21
2.3 ARQUITECTURA DE UN SISTEMA DE IR	22
2.4 LEMATIZACIÓN	23
2.5 LEMATIZADORES	23
2.5.1 Lematizador S-Stemmer	23
2.5.2 Lematizador de Lovins	24
2.5.3 Lematizador de Porter	24
2.5.4 Lematizador de Krovetz	25
2.5.5 Lematizador Nice	25
2.6 MORFOLOGIA DEL CASTELLANO	25
2.6.1 Morfología Flexiva	26
2.6.2 Morfología léxica	26
2.7 CARACTERÍSTICAS MORFOLÓGICAS DEL CASTELLANO	27
2.7.1 Derivaciones de género y número	28



2.7.2	Conjugación verbal.	28
2.7.3	Formas enclíticas	32
2.7.4	Nombres derivados de verbos	32
2.7.5	Adverbios derivados de adjetivos	32
2.7.6	Superlativos y diminutivos	32
2.7.7	Vocales acentuadas	32
2.7.8	La sílaba	32
3.	ARQUITECTURA DEL SISTEMA EN LA METODOLOGIA ORIENTADA A OBJETOS	36
3.1	INTRODUCCION ( VISION GENERAL )	36
3.2	ANALISIS ORIENTADO A OBJETOS	36
3.2.1	Definición del problema	36
3.2.2	Modelado de objetos	37
3.2.3	Modelado dinámico	41
3.2.4	Modelado funcional	72
3.3	DISEÑO DEL SISTEMA	85

3.3.1 Arquitectura del Sistema	85
3.3.2 Concurrencia	87
3.3.3 Almacenes de Datos	87
3.4 DISEÑO DE OBJETOS	88
3.4.1 Diseño de algoritmos para los métodos de las clases	88
4. ESPECIFICACIÓN DEL SISTEMA	102
4.1 ESPECIFICACIONES TECNICAS	102
4.2 ESPECIFICACIONES GENERALES DE LOS PROCESOS	103
4.2.1 Lematizador	103
4.2.2 Silabeador	104
4.2.3 Conjugador	104
4.3 SUGERENCIAS PARA LA UTILIZACION DEL SISTEMA	105
4.3.1 Sugerencias generales	105
4.4 COMO ACCEDER AL SISTEMA LESP	106
5. CONCLUSIONES	107

6. RECOMENDACIONES	109
BIBLIOGRAFÍA	110
ANEXOS	111