

### 1.- DATOS DE LA ASIGNATURA

<b>Nombre de la asignatura:</b> PROCESAMIENTO DE LENGUAJE NATURAL
<b>Carrera:</b> LICENCIATURA EN INFORMÁTICA.
<b>Clave de la asignatura:</b>
<b>Horas teoría-horas práctica- créditos</b> 1-3-5

### 2.- HISTORIA DEL PROGRAMA

<b>Lugar y fecha de elaboración o revisión</b>	<b>Participantes</b>	<b>Observaciones (cambios y justificación)</b>
Noviembre de 2006	M.C. Yesenia Nohemí González Meneses. M.C. Blanca Estela Pedroza Méndez.	

### 3.- UBICACIÓN DE LA ASIGNATURA

a). Relación con otras asignaturas del plan de estudio

Anteriores		Posteriores	
Asignaturas	Temas	Asignaturas	Temas
<ul style="list-style-type: none"> <li>• Software de Sistemas</li> <li>• Programación II</li> </ul>	Ensambladores Compiladores. Intérpretes. Todos		

**b). Aportación de la asignatura al perfil del egresado**

El procesamiento del lenguaje natural, es una parte de la computación inteligente que le permite al alumno egresado de la licenciatura en informática proponer diversas soluciones en cuanto al manejo de grandes volúmenes de información almacenada en textos, tanto impresos como escritos a mano, lo cual permite la generación de herramientas automatizadas para el manejo de dicho texto. Esto es en cuanto al procesamiento de información escrita, que es básicamente, el enfoque de este curso.

**4.- OBJETIVO(S) GENERAL(ES) DEL CURSO**

El alumno conocerá las principales técnicas computacionales para el análisis de la gramática del lenguaje español y la aplicación de estas a los sistemas inteligentes.

**5.- TEMARIO**

Unidad	Temas	Subtemas
1.	Introducción.	1.1 ¿Qué es el procesamiento de lenguaje natural? 1.2 Tipos de Gramáticas. 1.3 Gramáticas libres de contexto 1.4 Proceso de compilación. 1.5 Lenguajes regulares.
2	Análisis léxico	2.1 ¿Qué es un analizador léxico? 2.2 ¿Cuáles son las funciones de un analizador léxico?. 2.3 Ejemplo de analizadores léxico. 2.4 Ejemplo de un algoritmo para el analizador léxico. 2.5 Funciones del Análisis léxico en el Procesamiento del Lenguaje Natural.
3	Análisis	3.1 ¿Qué es un analizador

	sintáctico.	sintáctico? 3.2 ¿Cuáles son las funciones de un analizador sintáctico? 3.3 Ejemplos de analizadores sintácticos 3.4 Algoritmos para el analizador sintáctico. 3.5 Funciones del Análisis semántico en el Procesamiento del Lenguaje Natural
4	Análisis semántico.	4.1 ¿Qué es un analizador semántico? 4.2 ¿Cuáles son las funciones de un analizador semántico? 4.3 Funciones del Análisis semántico en el Procesamiento del Lenguaje Natural 4.4 Ambigüedad y Desambigüedad de las palabras en el lenguaje natural.
5	Aplicaciones del Procesamiento del lenguaje natural.	5.1 Lenguaje escrito. 5.2 ¿Qué es la etiquetación de las palabras? 5.3 Aplicaciones del lenguaje escrito. 5.3.1. OCR (Reconocedores de caracteres ópticos.? 5.3.2 Extracción de información. 5.3.3 Recuperación de información 5.2 Lenguaje hablado. 5.2.1 Aplicaciones del lenguaje hablado.

## 6.- APRENDIZAJES REQUERIDOS

- Conceptos sobre Software de Sistemas.
- Diseño y desarrollo de Software de Sistemas.
- Experiencia en el uso de un lenguaje de programación.

## 7.- SUGERENCIAS DIDÁCTICAS

Se propone realizar la exposición de los temas por parte del maestro y posteriormente, los alumnos deben realizar ejercicios prácticos en cuanto a diseño y desarrollo de gramáticas, para posteriormente implementar los programas que hagan el análisis léxico, sintáctico y semántico en

algún lenguaje de programación e incorporen diferentes herramientas para el procesamiento de lenguaje natural.

## 8.- SUGERENCIAS DE EVALUACIÓN

Exposiciones 20%  
 Programas 30%  
 Exámenes 50%

## 9.- UNIDADES DE APRENDIZAJE

### UNIDAD 1.- INTRODUCCION

Objetivo Educativo	Actividades de Aprendizaje	Fuentes de Información
Introducir al alumno a los conceptos básicos de compiladores, gramáticas, lenguajes regulares y procesamiento de lenguaje natural	1.1. Investigar el estado del arte del PLN. 1.2. Exposición de lo investigado 1.3. Creación y definición de un lenguaje regular posible a evaluar. 1.4. Creación y definición de una gramática para trabajar como proyecto final.	1,2,5

### UNIDAD 2.- ANÁLISIS SINTÁCTICO.

Objetivo Educativo	Actividades de Aprendizaje	Fuentes de Información
Comprender que es y como funciona un analizador léxico implementando su algoritmo	2.1 Búsqueda y obtención de un analizador léxico libre. 2.2 Búsqueda y obtención de un generador automático de analizadores léxicos. 2.3 Programación del algoritmo de un analizador léxico.	2,3, 11

### UNIDAD 3.- ANÁLISIS LÉXICO.

Objetivo Educativo	Actividades de Aprendizaje	Fuentes de Información
Comprender que es y como funciona un	3.1 Búsqueda y obtención de un analizador sintáctico libre.	2,3,11

analizador sintáctico implementando su algoritmo, comprender la relación con el PLN	3.2 Búsqueda y obtención de un generador automático de analizadores sintácticos. 3.3 Programación del algoritmo de un analizador léxico.	
---	---	--

#### UNIDAD 4.- ANÁLISIS SEMANTICO.

Objetivo Educativo	Actividades de Aprendizaje	Fuentes de Información
Comprender que es y como funciona un analizador semántico implementando su algoritmo, comprender la relación con el PLN y su complejidad	4.1. Búsqueda y obtención de un meta compilador. 4.2. Programación del algoritmo de un analizador semántico. 4.3. Investigación de algoritmos de búsqueda de ambigüedades. 4.4. Investigación de algoritmos de desambigüedades en texto.	2,3,11

#### UNIDAD 5.- APLICACIONES DEL PROCESAMIENTO DE LENGUAJE NATURAL

Objetivo Educativo	Actividades de Aprendizaje	Fuentes de Información
Comprender que es y como funciona un analizador semántico implementando su algoritmo, comprender la relación con el PLN y su complejidad	5.1. Investigación y exposición de una aplicación sobre procesamiento de lenguaje natural escrito 5.2. Investigación y exposición de una aplicación sobre procesamiento de lenguaje natural hablado. 5.3. Programación de un algoritmo para el reconocimiento de caracteres ópticos. 5.4. Análisis y Programación de algoritmos para la función de Extracción de información. 5.5 Análisis y Programación de algoritmos para la función de recuperación de información. 5.6. Investigación de aplicaciones de recuperación de información.	2,3,11

## 10. FUENTES DE INFORMACIÓN

- 1) TEORIA DE AUTÓMATAS Y LENGUAJES  
Editorial Prentice Hall  
*Unidad 2.*  
Alfred V. Aho, Orbi Sethi, Jeffrey D. Ullman
- 2) Compiladores, Principios, Técnicas y Herramientas  
Ed. Edisson Wesley  
Andrew W. Appel.
- 3) Modern Compiler, Implementation in Java.  
Cambridge University Press.  
Allen James
- 4) Natural Understanding  
Menlo Park.  
Ronald A Cole
- 5) Survey of the state of the art in human language  
technology  
CSLU/NSF 1996.

### *Lecturas adicionales:*

- 6) Dayne Freitag, Andrew McCallum  
**Information Extraction with HMM Structures Learned by Stochastic Optimization.** Association for Artificial Intelligence 2001.
- 7) L.R Rabiner.  
**A tutorial on hidden Markov Models and selected applications in speech recognition.** Proceedings of the IEEE, february 1989.
- 8) Raymond C. Vasko, Jr., Amro El-Jaroudy, J.R Boston, and Thomas E. Rudy. **Hidden Markov model topology estimation to characterize the dynamic structure of repetitive lifting data.** Proceedings of the 19<sup>th</sup> Anual International Conference of the IEEE in Medicine and Biology Society 1997
- 9) E. Riloff and J. Shepherd. 1997. **A corpus based approach for building semantic lexicons.** Proceedings of the Second Conference on Empirical Methods in Natural Language Processings, pages 127- 132.
- 10) D. Yarowsky. 1995. **Unsupervised word sense disambiguation rivaling supervised methods.** Proceedings

of the 33<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics, pages 189-196.

11)Otras fuentes

## 11. PRÁCTICAS DE LABORATORIO

### UNIDAD 1.- INTRODUCCION

Actividades de Aprendizaje
1.1. Captura y preparación de la definición de un lenguaje regular para su evaluación.
1.2. Captura y preparación de una gramática para su evaluación que servirá como parte del proyecto final.

### UNIDAD 2.- ANÁLISIS SINTÁCTICO.

Actividades de Aprendizaje
2.1. Búsqueda y obtención en Internet de un analizador léxico libre.
2.2. Búsqueda y obtención en Internet de un generador automático de analizadores léxicos.
2.3. Instalación y pruebas del software encontrado.
2.4. Análisis léxico de la gramática definida en la unidad I mediante software automatizado.
2.5. Programación del algoritmo de un analizador léxico para evaluar la gramática definida en la unidad I.

### UNIDAD 3.- ANÁLISIS LÉXICO.

Actividades de Aprendizaje
3.1. Búsqueda y obtención en Internet de un analizador sintáctico libre.
3.2. Búsqueda y obtención en Internet de un generador automático de analizadores sintácticos.
3.3. Instalación y pruebas del software encontrado.

- |  |
|--|
| 3.4. Análisis sintáctico de la gramática definida en la unidad I mediante software automatizado.               |
| 3.5. Programación del algoritmo de un analizador sintáctico para evaluar la gramática definida en la unidad I. |

**UNIDAD 4.- ANÁLISIS SEMANTICO.**

<b>Actividades de Aprendizaje</b>
4.1. Búsqueda y obtención en Internet de un analizador semántico libre.
4.2. Búsqueda y obtención en Internet de un generador -automático de analizadores semánticos.
4.3. Instalación y pruebas del software encontrado.
4.4. Análisis semántico de la gramática definida en la unidad I mediante software automatizado.
4.5. Programación del algoritmo de un analizador semántico para evaluar la gramática definida en la unidad I.
4.6. Integración de las 3 facetas de compilación en un solo software de aplicación, con la interface del editor de sentencias reconocibles por la gramática definida.

**UNIDAD 5.- APLICACIONES DEL PROCESAMIENTO DE LENGUAJE NATURAL**

<b>Actividades de Aprendizaje</b>
5.1. Programación de un algoritmo para el reconocimiento de caracteres ópticos.
5.2. Programación de un algoritmo para la función de Extracción de información.
5.3. Programación de un algoritmo para la función de recuperación de información.