

# ASIGNATURA: TECNOLOGÍAS DE LENGUAJE NATURAL

## 1. DATOS DE LA ASIGNATURA

Nombre de la asignatura: TECNOLOGÍAS DEL LENGUAJE NATURAL
Línea de trabajo: Desarrollo y aplicación de tecnologías inteligentes.
Horas teoría-horas prácticas-horas trabajo adicional-horas totales-créditos
32 - 32 -64 - 128 - 8

## 2. HISTORIAL DE LA ASIGNATURA

Lugar y fecha de elaboración o revisión	Participantes	Observaciones (cambios y justificación)
Apizaco Tlax., Febrero Del 2005	M.C. Blanca Estela Pedroza Méndez	

## 3. PRE-REQUISITOS Y CORREQUISITOS

Requisitos:

Introducción a la Inteligencia Artificial.

## 3 OBJETIVO DE LA ASIGNATURA

El alumno deberá conocer diversas técnicas para el análisis de diferentes tipos de lenguajes, empezando primero por algunas técnicas para el análisis de los lenguajes de alto nivel, a partir del análisis de sus gramáticas. Posteriormente el análisis se enfoca al reconocimiento del lenguaje natural escrito, mediante el reconocimiento de caracteres, palabras y frases. Dicho análisis se basa un poco en el lenguaje inglés, por lo que el alumno deberá de ser capaz de relacionar dichas técnicas con el lenguaje español para lo cual se requiere que el alumno se empape un poco del manejo de las reglas de la gramática del español.

## 5. APORTACIÓN AL PERFIL DEL GRADUADO

El procesamiento del lenguaje natural, es una parte de la computación inteligente que le permite al alumno egresado de la maestría en ciencias de la computación 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, pero además, la materia introduce un poco al alumno al conocimiento del procesamiento de lenguaje hablado.

## 6. CONTENIDO TEMÁTICO POR TEMAS Y SUBTEMAS

UNIDAD	TEMAS	SUBTEMAS
1	Introducción  Tiempo: 8 horas teóricas.	1.1 Proceso de compilación 1.2 Gramáticas libres de contexto 1.3 Expresiones regulares y lenguajes 1.4 Autómatas finitos deterministas y no deterministas. 1.4.1 Conversión de un A.F.N a un A.F.D
2	Análisis Léxico, Sintáctico y Semántico  Tiempo: 8 horas teóricas.	2.1 Función del analizador léxico 2.2 Algoritmos del analizador léxico 2.3 Función del analizador sintáctico 2.4 Algoritmos del analizador sintáctico 2.5 Análisis semántico
3	Introducción al procesamiento del Lenguaje Natural  Tiempo: 6 horas teóricas.	3.1 El estudio del lenguaje 3.2 Niveles del análisis del lenguaje 3.3 Reconocimiento del lenguaje escrito 3.3.1 Caracteres impresos 3.3.2 Caracteres manuscritos 3.3.3 Reconocimiento de palabras
4	Comprensión del lenguaje  Tiempo: 10 horas teóricas.	4.1 Introducción 4.2 Algoritmo del analizador de diagramas 4.3 Modelos estadísticos 4.3.1 Cadenas de Markov 4.3.2 Entropía cruzada 4.3.3 Modelos ocultos de Markov

## 7. METODOLOGÍA DE DESARROLLO DEL CURSO

Se propone realizar la exposición de los temas por parte del maestro y posteriormente, los alumnos deben realizar la lectura y exposición de diversos artículos de investigación relacionados con cada uno de los temas. Además, el curso contiene varios algoritmos que pueden ser desarrollados en algún lenguaje de programación por parte de los alumnos, una vez que el profesor los expuso en clase con algunos ejemplos prácticos.

## 8. SUGERENCIAS DE EVALUACIÓN

Exposiciones 20%  
Programas 30%  
Exámenes 50%

## 9. BIBLIOGRAFÍA Y SOFTWARE DE APOYO

### *Unidad 1.*

Dean Kelly  
TEORIA DE AUTÓMATAS Y LENGUAJES  
Editorial Prentice Hall

### *Unidad 2.*

Alfred V. Aho, Orbi Sethi, Jeffrey D. Ullman  
Compiladores, Principios, Técnicas y Herramientas  
Ed. Edisson Wesley

Andrew W. Appel.  
Modern Compiler, Implementation in Java.  
Cambridge University Press.

### *Unidad 3.*

Allen James  
Natural Understanding  
Menlo Park.

Ronald A Cole  
Survey of the state of the art in human language technology  
CSLU/NSF 1996.

### *Lecturas adicionales:*

Dayne Freitag, Andrew McCallum  
**Information Extraction with HMM Structures Learned by Stochastic Optimization.** Association for Artificial Intelligence 2001.

L.R Rabiner.  
**A tutorial on hidden Markov Models and selected aplicaciones in speech recognition.** Proceedings of the IEEE, february 1989.

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> Annual International Conference of the IEEE in Medicine and Biology Society 1997

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.

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.

## 10. PRÁCTICAS PROPUESTAS

Se sugiere que las prácticas propuestas sean realizadas por equipos para estar en concordancia con la finalidad de fomentar la discusión de ideas que plantea el curso. En este sentido, se proponen las siguientes prácticas por unidad:

Unidad	Práctica
1 Introducción. 8 horas prácticas.	1.1 Diseñar un programa que simule un autómata finito determinista y un autómata finito no determinista. 1.2 Realizar un programa para el algoritmo de conversión de un AFN a un AFD
1. Análisis Léxico, Sintáctico y Semántico. 8 horas prácticas.	2.1 Diseñar un programa que emule un pequeño compilador, para alguna gramática definida por el usuario.
3. Introducción al procesamiento del Lenguaje Natural 8 horas prácticas.	3.1 Realizar un programa para el manejo de bigramas y trigramas de letras, para el manejo de estadísticas de la aparición de diferentes sílabas en un texto. 3.2 Modificar el programa del apartado 3.1 , para palabras.
4. Comprensión del lenguaje 8 horas prácticas.	4.1 Realizar un programa para el algoritmo del analizador de diagramas. 4.2 Aplicar las estadísticas vistas en esta unidad a los programas que se diseñaron en la unidad 3.

**Catedrático Responsable:**

***M.C. Blanca Estela Pedroza Méndez.***