



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD CIENCIAS DE LA COMPUTACIÓN

**PROGRAMA DE LA MATERIA CORRESPONDIENTE A LA
LICENCIATURA EN CIENCIAS DE LA COMPUTACIÓN.**

Coordinación: Área de Interfaces Humano- Computadoras

NOMBRE DE LA MATERIA:	Procesamiento del Lenguaje Natural
------------------------------	------------------------------------

Clave: CCO 492

Créditos: 10

Modalidad: Escolarizada

Nivel de Ubicación: Formativo

Tipo de Materia: Optativa

PRE-REQUISITOS:	Nivel Básico
------------------------	--------------

MATERIA CONSECUENTE:	Ninguna
-----------------------------	---------

TIEMPO TOTAL ASIGNADO:	80 horas
-------------------------------	----------

PRIMAVERA - OTOÑO

HRS. TEÓRICAS/SEM: 5 horas		HRS. PRÁCTICAS/SEM:	
-----------------------------------	--	----------------------------	--

VERANO

HRS. TEÓRICAS/SEM: 10 horas		HRS. PRÁCTICAS/SEM:	
------------------------------------	--	----------------------------	--

AUTOR(ES) DEL PROGRAMA:	
--------------------------------	--

REVISADO POR:	Héctor Jiménez Salazar
----------------------	------------------------

APROBADO POR:	
----------------------	--

AUTORIZADO POR:	
------------------------	--

FECHA DE ELABORACIÓN/REVISIÓN:	Junio 30, 2003
---------------------------------------	----------------

VIGENCIA:	
------------------	--

JUSTIFICACIÓN:

En el mundo cada vez hay más información textual disponible en forma electrónica. Pero además

son pocos o deficientes los recursos para explotarla. El procesamiento del lenguaje natural (PLN) habrá de desarrollarse a la par de la demanda. Necesidad que se traduce en la preparación de recursos humanos con una visión básica frente al copioso caudal de información no estructurada.

OBJETIVO GENERAL DE LA MATERIA:

El estudiante conocerá la amplitud de las aplicaciones del PLN y su impacto en la tecnología actual, asimismo practicará con algunos recursos lingüísticos y técnicas básicas del PLN para llevar a cabo la integración de sistemas simples.

CONTRIBUCIÓN DE LA ASIGNATURA AL PERFIL DE EGRESO:

Aunque este es un curso básico, en él se presenta una serie de aplicaciones tecnológicas hacia un tipo de datos poco común: texto con un alto contenido informativo. Este curso orienta frente a las demandas presentes y futuras de procesamiento de información.

CONTENIDO TEMÁTICO

UNIDAD: I			TÍTULO: INTRODUCCION		
OBJETIVO ESPECÍFICO:					
El estudiante identificará los problemas relacionados con el PLN.					
CONTENIDO DE LA UNIDAD	Tiempo de impartición (hrs.).		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
	HT	HP			
1.1 Qué es PLN, áreas, problemática.	4		Lectura (subrayando y resumiendo). Práctica de estos conocimientos mediante la respuesta a preguntas sobre problemas en ese contexto.	En general se sigue el método de lectura de comprensión.	1. Notas en WEB 2. Lista de ejercicios en WEB. 3. Acetatos en WEB
1.2 Soluciones y eficiencia	1				
1.3 Simbolismo vs empirismo	5				
HORAS TOTALES:	10				

UNIDAD: II			TÍTULO: ANALISIS MORFOSINTACTICO		
OBJETIVO ESPECÍFICO:					
Diseñar algún componente morfosintáctico simple con base a recursos lingüísticos existentes.					

CONTENIDO DE LA UNIDAD	Tiempo de impartición (hrs.).		Actividades de Aprendizaje	Técnicas	Recursos Necesarios	
	HT	HP				
	2 1	Conceptos básicos (morfemas, vocablos, reglas-excepciones, ambigüedad, definitud, género, número, silabación, n-gramas, partes del discurso, clífticos, conjugaciones, truncamiento y lematización)				10
2 2	Etiquetamiento con partes del discurso (uso de un diccionario, probabilidad apriori, probabilidad condicional de n-gramas, reglas inducidas)	5				
2 3	Lematización (Algoritmo de silabación, uso de diccionario y reglas, lematización de verbos, lematización de nominales)	10				
2 4	Análisis sintáctico (generación y reconocimiento usando programación lógica, backtracking, esquina izquierda, parsing generalizado)	10				
HORAS TOTALES		35				

UNIDAD: III	TÍTULO: ANÁLISIS LÉXICO-SEMÁNTICO
OBJETIVO ESPECÍFICO:	
Aplicar los métodos más conocidos en problemas semánticos del manejo de texto.	

CONTENIDO DE LA UNIDAD	Tiempo de impartición (hrs.).		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
	HT	HP			
	3 1 Agrupamiento con poco conocimiento	5			
3 2 Clasificación de palabras extraídas de textos.	5				
3 3 Desambiguación del sentido de una palabra.	5				
HORAS TOTALES:	15				

UNIDAD: IV		TÍTULO: EJEMPLOS DE SISTEMAS DE PLN			
OBJETIVO ESPECÍFICO:					
Conocer la integración de diferentes componentes en sistemas variados para el PLN					
CONTENIDO DE LA UNIDAD	Tiempo de impartición (hrs.).		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
	HT	HP			

4 1	Traducción automática de textos	5	Lectura (subrayando y resumiendo) de las definiciones y resultados importantes. Observación de la aplicación de los conocimientos en la solución de algunos ejercicios. Resolución de ejercicios y discusión sobre estos (diseño de algoritmos)	En general se sigue el método de lectura de comprensión frente a la solución de ejercicios. Y explicaciones sobre algunos algoritmos clásicos.	1. Notas en WEB 4. Lista de ejercicios en WEB. 5. Recursos lingüísticos en WEB.
4 2	Corrección ortográfica	5			
4 3	Extracción de información	5			
HORAS TOTALES:		15			

PRACTICAS

UNIDAD	NOMBRE DE LA PRACTICA	OBJETIVO	HORAS
2	Construcción de un silabador usando reglas (este ejercicio puede variar de cursos a curso, otro puede ser construcción de un parser de esquina izquierda).	Que el estudiante practique con algunos paradigmas del tratamiento del lenguaje natural orientados a la morfosintaxis.	8 horas (en equipos de dos o tres.)
3	Agrupamiento de palabras según su contexto: palabras relacionadas dentro de un dominio. (este ejercicio puede variar de cursos a curso, otro puede ser desambiguación del sentido de una palabra)	Que el estudiante practique con algunos paradigmas del tratamiento del lenguaje natural orientados a la semántica.	8 horas (en equipos de dos o tres.)

CRITERIOS DE EVALUACIÓN

EXÁMENES PARCIALES DEPARTAMENTALES

Parcial	Contenido a evaluar	Periodos
I	U1 a U2: 2.1-2.2	6-8a. semana
II	U2 2.3-2.4 a U3: 3.1-3.3	12a. semana

	%
Exámenes:	40 (20 c/u)
Proyecto Final:	60
Tareas:	
Trabajos de Investigación:	
Prácticas de Laboratorio:	
TOTAL:	

REQUISITOS DE ACREDITACIÓN:

1. Se pretende que el estudiante exente el curso haciendo solamente dos exámenes y la práctica.
2. Los exámenes se promedian si son aprobatorios.
3. El examen final (de todo el curso) se aplica sólo cuando no fue (o fueron) aprobado un parcial.

FOMENTO DE VALORES:

1. Cumplimiento de entrega en trabajo.
2. Entrega de trabajos "en limpio".
3. Honestidad en la realización (resultados propios).

BIBLIOGRAFÍA:

- (B) Jiménez S. "Procesamiento del lenguaje natural", notas del curso FCC-BUAP, www.aleteya.cs.bubap.mx/~hjimenez, 2003.
- (C) Hausser, Roland: "Fundamentals of computational linguistics", Springer, 1999.
- (C) MacEnery & Wilson: "Corpus linguistics", 1996.
- (C) Franz: "Ambiguity resolution in natural language processing", Springer, 1996.
- B: Básico
C: Complementario

TITULAR (RESPONSABLE) DE LA MATERIA:

FECHA DE ELABORACIÓN Y AUTOR(ES) DEL PROGRAMA:

Junio 26, 2003.