



MAESTRIA EN CIENCIAS DE LA COMPUTACION

Área: Base de Datos y Recuperación de Información

Programa de Asignatura: Procesamiento de Lenguaje Natural

Código: MCOM 22228

Tipo: Optativa

Créditos: 9

Fecha: Noviembre 2012



1. DATOS GENERALES

Nombre del Programa Educativo:	Maestría en Ciencias de la Computación
Modalidad Académica:	Escolarizada
Nombre de la Asignatura:	Procesamiento de lenguaje natural
Ubicación:	Segundo o Tercer semestre (Optativa)

2. REVISIONES Y ACTUALIZACIONES

Autores:	Dr. David Eduardo Pinto Avendaño
Fecha de diseño:	Noviembre 2012
Fecha de la última actualización:	Abril 2019
Revisores:	Dra. María Josefa Somodevilla García Dra. Concepción Pérez de Célis Dra. Mireya Tovar Vidal Dr, David Eduardo Pinto Dra. Darnes Vilariño Ayala
Sinopsis de la revisión y/o actualización:	Se agregó las unidades 5 y 6.



3. OBJETIVOS

GENERAL

El estudiante conocerá y aplicará métodos actuales en diversos problemas relacionados con el tratamiento automático del lenguaje natural.

ESPECIFICOS

- 1.- El estudiante conocerá conceptos básicos de lingüística.
- 2.- El estudiante revisará los conceptos básicos de matemáticas para el tratamiento estadístico de corpora.
- 3.- El estudiante entenderá los diversos problemas y retos asociados con el tratamiento automático del lenguaje natural.
- 4.- El estudiante aplicará los conceptos aprendidos en la resolución de problemas del procesamiento del lenguaje natural.



4. CONTENIDO

Unidad	Contenido Temático
I. INTRODUCCIÓN	1.1 Motivación 1.2 Conceptos básicos de lingüística 1.2 Conceptos básicos de matemáticas 1.3 Aplicaciones del procesamiento del lenguaje natural
2. ANÁLISIS LÉXICO, MORFOLÓGICO Y SINTÁCTICO	2.1 Lingüística de corpus 2.2 Colocaciones 2.3 Inferencia estadística 2.4 Adquisición léxica 2.5 Etiquetamiento de partes del discurso
3. CLASIFICADORES	3.1 El clasificador de Naive Bayes 3.2 El clasificador Gaussiano 3.3 Máquinas de soporte vectorial 3.4 Modelos de Markov 3.5 Conditional Random Fields
4. APLICACIONES DEL PROCESAMIENTO DEL LENGUAJE	4.1 Desambiguación del sentido de las palabras 4.2 Recuperación de información 4.3 Parsing estadístico 4.4 Traducción 4.5 Agrupamiento 4.6 Clasificación
5. TÓPICOS DE RECUPERACIÓN DE INFORMACIÓN	5.1 El modelo de espacio vectorial 5.2 Similitud vectorial 5.3 Pesado de términos 5.4 Modelo de Distribución de términos 5.5 Distribución de Poisson 5.6 Frecuencia inversa de un documento
6. CATEGORIZACIÓN DE TEXTOS	6.1 Árboles de decisión 6.2 Modelo de máxima entropía 6.3 Perceptrones 6.4 Clasificación de vecinos más cercanos



Bibliografía	
Básica	Complementaria
1.- Daniel Jurafsky and James H. Martin Speech and Language Processing. Prentice Hall, New Jersey, second edition, 2008. 2.- Eneko Agirre and Philip Edmonds. Word Sense Disambiguation: Algorithms and Applications, Springer, 2007.	

5. CRITERIOS DE EVALUACIÓN

Criterios	Porcentaje
• Exámenes	40%
• Participación en clase	
• Tareas	
• Exposiciones	
• Simulaciones	
• Trabajo de investigación y/o de intervención	
• Prácticas de laboratorio	40%
• Visitas guiadas	
• Reporte de actividades académicas y culturales	
• Mapas conceptuales	
• Portafolio	
• Proyecto final	20%
• Otros	
Total	100%