

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA



PROGRAMA DE ESTUDIO

ANÁLISIS Y PROCESAMIENTO INTELIGENTE DE TEXTOS

0575

8°, 9°

06

Asignatura

Clave

Semestre

Créditos

Ingeniería Eléctrica

Ingeniería en Computación

Ingeniería en Computación

División

Departamento

Carrera en que se imparte

Asignatura:

Obligatoria

Optativa

Horas:

Teóricas

Prácticas

Total (horas):

Semana

16 Semanas

Aprobado:

Consejo Técnico de la Facultad

Consejo Académico del Área de las Ciencias

Físico Matemáticas y de las Ingenierías

Fecha:

25 de febrero, 17 de marzo y 16 de junio de 2005

11 de agosto de 2005

Modalidad: Curso

Asignatura obligatoria antecedente: Ninguna.

Asignatura obligatoria consecuente: Ninguna.

Objetivo(s) del curso:

El alumno conocerá las principales aplicaciones del procesamiento de lenguaje natural para el análisis y procesamiento inteligente de textos, lo que le permitirá proponer soluciones tecnológicas a las necesidades de información de cualquier organismo.

Temario

NÚM.	NOMBRE	HORAS
1.	Introducción al análisis y procesamiento inteligente de textos	3.0
2.	Recuperación de información (RI)	10.0
3.	Extracción de información (EI)	11.0
4.	Resumen automático de textos (RAT)	12.0
5.	Minería de textos (MT)	12.0
		48.0
	Prácticas de laboratorio	0.0
	Total	48.0



1 Introducción al análisis y procesamiento inteligente de textos

Objetivo: El alumno conocerá la necesidad del análisis automático e inteligente de textos como datos no estructurados.

Contenido:

- 1.1 Información estructurada y no estructurada
- 1.2 Necesidades de información de las organizaciones
- 1.3 El PLN para el análisis y procesamiento inteligente de textos
- 1.4 Tecnologías del habla
- 1.5 Business Intelligence

2 Recuperación de información (RI)

Objetivo: El alumno adquirirá los conocimientos básicos sobre la recuperación de información (ri) y los componentes de un sistema.

Contenido:

- 2.1 Introducción
 - 2.1.1 Antecedentes
 - 2.1.2 Concepto
 - 2.1.3 El PLN en la RI
 - 2.1.4 Aplicaciones
- 2.2 Sistemas de RI
 - 2.2.1 Preproceso de documentos
 - 2.2.1.1 Palabras funcionales
 - 2.2.1.2 Lematización
 - 2.2.1.3 Términos clave
 - 2.2.1.4 Relaciones entre términos
 - 2.2.2 Técnicas de indexación
 - 2.2.2.1 Indexación lingüística
 - 2.2.3 Recuperación
 - 2.2.3.1 Procesamiento de peticiones
 - 2.2.3.1.1 Expansión de peticiones
 - 2.2.3.2 Técnicas de recuperación
 - 2.2.4 Recuperación basada en la semántica
 - 2.2.5 Medidas de Evaluación

3 Extracción de información (EI)

Objetivo: El alumno aprenderá los conceptos principales relacionados con la extracción de información (EI), sus técnicas y métodos.

Contenido:

- 3.1 Introducción
 - 3.1.1 Antecedentes
 - 3.1.1.1 Las MUC's



- 3.1.2 Concepto
- 3.1.3 Diferencias entre RI y EI
- 3.1.4 Aplicaciones
- 3.2 Técnicas de EI
 - 3.2.1 Expresiones regulares
 - 3.2.2 Autómatas
 - 3.2.3 Gramáticas independientes del contexto
- 3.3 Extracción de entidades
 - 3.3.1 Reconocimiento de entidades
 - 3.3.2 Identificación y clasificación de nombres
- 3.4 Extracción de eventos
 - 3.4.1 Reconocimiento de eventos
- 3.5 Extracción de relaciones
- 3.6 Algoritmos para aprendizaje de máquinas

4 Resumen automático de textos (RAT)

Objetivo: El alumno comprenderá el análisis y procesamiento inteligente de textos mediante el acercamiento al resumen automático de textos (RAT), así como la importancia del PLN en la solución de requerimientos de información específicos.

Contenido:

- 4.1 Introducción
 - 4.1.1 Concepto
 - 4.1.2 Objetivo
 - 4.1.3 Retos del RAT
 - 4.1.4 Unidades de resumen
 - 4.1.5 Aplicaciones
- 4.2 Criterios de extracción
- 4.3 Etapas del RAT
 - 4.3.1 Identificación de tópico
 - 4.3.2 Identificación de unidades
 - 4.3.3 Extracción
 - 4.3.4 Generación del resumen
- 4.4 Métodos del RAT
 - 4.4.1 Estadísticos
 - 4.4.2 Selección de oraciones
 - 4.4.3 Selección de párrafos
 - 4.4.4 Discursivos
 - 4.4.5 Por plantillas
 - 4.4.6 Otros
- 4.5 Evaluación de resúmenes
 - 4.5.1 Métricas
- 4.6 Resumen de múltiples documentos
 - 4.6.1 Problemática
 - 4.6.2 Métodos



5 Minería de textos (MT)

Objetivo: El alumno descubrirá el proceso de minería de textos, así como las técnicas y herramientas para realizarlo.

Contenido:

- 5.1 Introducción
 - 5.1.1 Antecedentes
 - 5.1.2 Concepto
 - 5.1.3 Minería de datos
 - 5.1.4 Aplicaciones
- 5.2 Técnicas básicas de MT
- 5.3 Algoritmos y modelos
- 5.4 Herramientas
- 5.5 Software de MT
- 5.6 Más allá de la MT
 - 5.6.1 Minería de web

Bibliografía básica:

GELBUKH, Alexander.

Computational Linguistics and Intelligent Text Processing: Second International Conference, Cicing 2001, Mexico City, Mexico, February 2001 : Proceedings (Lecture Notes in Computer Science, 2004)
Springer-Verlag Telos (March 1, 2001)

PAZIENZA, Maria Teresa

Information Extraction in the Web Era : Natural Language Communication for Knowledge Acquisition and Intelligent Information Agents (Lecture Notes in ... e / Lecture Notes in Artificial Intelligence)
Springer; (September 17, 2003)

A. LICHNEROWICZ

Intelligent Text and Image Handling: Proceedings
Elsevier Science Ltd (December 1, 1991)

JACOBS, Paul Schafran & Paul S.

Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval
Publisher: Lea (July 1, 1992)

Bibliografía complementaria:

La sugerida por el profesor.

**Sugerencias didácticas:**

Exposición oral	<input checked="" type="checkbox"/>
Exposición audiovisual	<input checked="" type="checkbox"/>
Ejercicios dentro de clase	<input checked="" type="checkbox"/>
Ejercicios fuera del aula	<input checked="" type="checkbox"/>
Seminarios	<input type="checkbox"/>

Lecturas obligatorias	<input checked="" type="checkbox"/>
Trabajos de investigación	<input checked="" type="checkbox"/>
Prácticas de taller o laboratorio	<input checked="" type="checkbox"/>
Prácticas de campo	<input type="checkbox"/>
Otras	<input type="checkbox"/>

Forma de evaluar:

Exámenes parciales	<input type="checkbox"/>
Exámenes finales	<input checked="" type="checkbox"/>
Trabajos y tareas fuera del aula	<input checked="" type="checkbox"/>

Participación en clase	<input type="checkbox"/>
Asistencias a prácticas	<input type="checkbox"/>
Otras	<input type="checkbox"/>

Perfil profesiográfico de quienes pueden impartir la asignatura

Ingeniero en computación familiarizado con extracción y recuperación de información, generación automática de resúmenes y minería de textos.