



**TEMAS SELECTOS DE
INGENIERÍA EN COMPUTACIÓN I
(APLICACIONES DE
TECNOLOGÍAS DEL LENGUAJE))**

2596

10

6

Asignatura	Clave	Semestre	Creditos
INGENIERÍA ELÉCTRICA	COMPUTACIÓN	INGENIERÍA EN COMPUTACIÓN	
División	Departamento	Carrera(s) en que se imparte	

Asignatura:

Obligatoria

Horas/semana:

Teóricas

3.0

Prácticas

0.0

Total

3.0

Horas/semestre:

Teóricas

48.0

Prácticas

0.0

Total

48.0

Optativa

Modalidad: Curso teórico

Seriación obligatoria antecedente: Ninguna

Seriación obligatoria consecuente: Ninguna

Descripción del curso:

Curso teórico-práctico sobre una aplicación de tecnologías del lenguaje.

Objetivo(s) del curso:

El alumno demostrará la experiencia necesaria, conocimientos y habilidades de la ingeniería en computación en temas selectos e innovadores de tecnologías del lenguaje, enfocando su participación en un proyecto, llegando a actividades concretas que pueden ser utilizadas para su titulación, ya sea a través de tesis o de artículo.

Temario

NÚM.	NOMBRE	HORAS
1	Introducción a las tecnologías del lenguaje	3.0
2	Gestión y anotación de corpus	9.0
3	Análisis de concordancias y lista de palabras	6.0
4	Extracción terminológica	9.0
5	Extracción de contextos definitorios	9.0
6	Motor de búsqueda de términos	12.0
		48.0
	Actividades prácticas	48.0
	Total	48.0

TEMARIO DESGLOSADO

1 INTRODUCCIÓN A LAS TECNOLOGÍAS DEL LENGUAJE

Objetivo: El alumno se familiarizará con los problemas de tecnologías del lenguaje, en general, y con el proyecto sésamo, de búsqueda de términos a partir de la descripción del concepto, en el cual trabajará a lo largo del semestre, para tener un producto entregado, que puede ser un desarrollo o un artículo.

Contenido:

- 1.1 Introducción al curso.
- 1.2 Las tecnologías del lenguaje.
- 1.3 El proyecto sésamo.

2 GESTIÓN Y ANOTACIÓN DE CORPUS

Objetivo: El alumno profundizará su conocimiento sobre los corpus lingüísticos como el recurso principal de tecnologías del lenguaje, los sistemas de gestión existentes y la información básica para su aprovechamiento en los proyectos de tecnologías del lenguaje.

Contenido:

- 2.1 Los corpus lingüísticos como recursos de tecnologías del lenguaje.
- 2.2 Sistemas de gestión de corpus.
- 2.3 Anotación y etiquetado de corpus.
- 2.4 Arquitectura del gestor de corpus Geco.

3 ANÁLISIS DE CONCORDANCIAS Y LISTAS DE PALABRAS

Objetivo: El alumno conocerá las principales herramientas de análisis cuantitativo y cualitativo de los corpus textuales y diseñará las mismas como parte del gestor de corpus.

Contenido:

- 3.1 Toquenización y segmentación de corpus.
- 3.2 Generadores de concordancias.
- 3.3 Visualización de datos.

4 EXTRACCIÓN TERMINOLÓGICA

Objetivo: El alumno dominará las técnicas de extracción de términos de corpus de especialidad, y diseñará un multiextractor con base en los extractores existentes.

Contenido:

- 4.1 Bases de la extracción terminológica.
- 4.2 Métodos de extracción terminológica.
- 4.3 Extractores terminológicos.
- 4.4 Evaluación de los extractores.
- 4.5 Implementación de un multiextractor.

5 EXTRACCIÓN DE CONTEXTOS DEFINITORIOS

Objetivo: El alumno dominará las técnicas de extracción de contextos definitorios en corpus lingüísticos, su anotación y desarrollará un extractor.

Contenido:

- 5.1 Fundamentos de los contextos definitorios.
- 5.2 Extracción de contextos definitorios.
- 5.3 Anotación en XML y JSON de los contextos definitorios.
- 5.4 Implementación del Ecode.

6 MOTOR DE BÚSQUEDA DE TÉRMINOS

Objetivo: El alumno se familiarizará con la búsqueda léxica a partir de la descripción del concepto en lenguaje natural, utilizando grafos de asociación léxica.

Contenido:

- 6.1 La búsqueda inversa de términos.
- 6.2 Ejercicio de conceptualización.
- 6.3 Grafos de asociación léxica.
- 6.4 Diseño y evaluación del motor de búsqueda léxica.
- 6.5 Diseño del sistema Describeme.

Bibliografía básica	Temas para los que se recomienda
Sierra, Gerardo, <i>Introducción a los corpus lingüísticos</i> , Universidad Nacional Autónoma de México, 2017.	1, 2 ,3, 4, 5 ,6
Sierra, Gerardo, Julián Solórzano y Arturo Curiel, “GECO, un gestor colaborativo basado en web”, <i>Linguamatica 9 (2)</i> , 2017.	1, 2 ,3, 4, 5 ,6

Bibliografía complementaria	Temas para los que se recomienda
Sierra, Gerardo, Jorge Reyes, Gemma Bel Enguix, Helena Gómez Adorno, “Búsqueda onomasiológica basada en normas de asociación de palabras”, <i>Tecnologías del lenguaje humano</i> , Universidad de Medellín.	
Sierra, Gerardo, “Extracción automática de contextos definitorios en textos especializados”, <i>Linguamatica 2</i> , 2009.	

Materiales del curso	
Classroom	

Evaluación	
------------	--

Proyecto	50 %
Exámenes	0 %
Prácticas	50 %

Información del profesor

Nombre completo:

Gerardo Sierra Martínez

Correo electrónico institucional:

gsierram@iingen.unam.mx

Horario de la clase:

Martes y jueves de 11:340 a 13 horas

Semblanza corta del profesor.

Investigador titular del Instituto de Ingeniería, Jefe del Grupo de Ingeniería Lingüística, Sistema Nacional de Investigadores nivel III.