

**ANX-PR/CL/001-02**  
**GUÍA DE APRENDIZAJE**

**ASIGNATURA**

Ingeniería lingüística

**CURSO ACADÉMICO - SEMESTRE**

2015-16 - Primer semestre

## Datos Descriptivos

---

<b>Nombre de la Asignatura</b>	Ingenieria lingüística
<b>Titulación</b>	10AN - Master Universitario en Ingeniería Informatica
<b>Centro responsable de la titulación</b>	E.T.S. de Ingenieros Informaticos
<b>Semestre/s de impartición</b>	Tercer semestre
<b>Carácter</b>	Optativa
<b>Código UPM</b>	103000654
<b>Nombre en inglés</b>	Linguistic Engineering

## Datos Generales

---

<b>Créditos</b>	4.5	<b>Curso</b>	2
<b>Curso Académico</b>	2015-16	<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano	<b>Otros idiomas de impartición</b>	

## Requisitos Previos Obligatorios

---

### Asignaturas Superadas

El plan de estudios Master Universitario en Ingeniería Informatica no tiene definidas asignaturas previas superadas para esta asignatura.

### Otros Requisitos

El plan de estudios Master Universitario en Ingeniería Informatica no tiene definidos otros requisitos para esta asignatura.

## Conocimientos Previos

---

### Asignaturas Previas Recomendadas

Sistemas inteligentes

### Otros Conocimientos Previos Recomendados

Linguística computacional

bases de datos

Inteligencia artificial

## Competencias

---

CE12 - Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

CE19 - Capacidad para desarrollar e implantar una solución informática en un entorno empresarial

CG1 - Capacidad de organizar y planificar

CG12 - Capacidad de trabajar de forma independiente en su campo profesional

CG2 - Capacidad de gestionar la información

CG6 - Capacidad de pensamiento creativo con el objetivo de desarrollar enfoques y métodos nuevos y originales

CG9 - Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente

## Resultados de Aprendizaje

---

RA56 - Conocer los procedimientos y herramientas para la creación de recursos lingüísticos

RA59 - Conocer y diseñar los componentes del sistema UNL

RA55 - Conocer los principales sistemas que representan las aplicaciones de la Ingeniería Lingüística

RA57 - Conocer y diseñar sistemas de Traducción automática

RA58 - Conocer y diseñar sistemas de extracción de Información

## Profesorado

---

### Profesorado

Nombre	Despacho	e-mail	Tutorías
Cardenosa Lera, Jesus ( <b>Coordinador/a</b> )	L-3310	jesus.cardenosa@upm.es	M - 19:00 - 21:00 Puedo atender tutorías cualquier día y hora previa cita por email.
Boguslavskiy, Igor	2201	igor.boguslavskiy@upm.es	X - 15:00 - 17:00 Puede atender cualquier día y hora previa cita por email

**Nota.-** Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

## Descripción de la Asignatura

---

La asignatura de Ingeniería Lingüística. está orientada a dar a conocer al estudiante los distintos tipos de sistemas que existen, sus componentes básicos y las tecnologías y recursos que los soportan. Es una asignatura muy orientada al diseño de sistemas y también a su análisis para poder proceder al diseño de nuevas aplicaciones y solución de necesidades en entornos corporativos, que conlleven el manejo de documentos, su búsqueda y la extracción de información principalmente.

## Temario

---

1. Creación de recursos lingüísticos
  - 1.1. Recursos léxicos ¿Qué son?
  - 1.2. Lexicones. Tesoros
  - 1.3. Corpus. Tipos
  - 1.4. Diseño, construcción y evaluación de corpus
2. Traducción automática
  - 2.1. Estado del Arte
  - 2.2. Sistemas comerciales
  - 2.3. Evaluación
3. Sistemas de extracción de información
  - 3.1. Recuperación de información. Tipos de sistemas (Booleano, Probabilístico, VSM)
  - 3.2. Extracción de información (NER, eventos, relaciones)
  - 3.3. Evaluación. Métricas
  - 3.4. El sistema FLEX
4. El sistema UNL
  - 4.1. Interlinguas. ¿Qué son? Tipos
  - 4.2. Sistemas basados en interlinguas. Arquitecturas. Componentes
  - 4.3. Desarrollo de componentes
  - 4.4. Procesos y Visualización

## Cronograma

**Horas totales:** 114 horas

**Horas presenciales:** 66 horas (54.3%)

**Peso total de actividades de evaluación continua:**  
300%

**Peso total de actividades de evaluación sólo prueba final:**  
300%

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades Evaluación
Semana 1	<b>Tema 1.1: Recursos léxicos. ¿Qué son?</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			
Semana 2	<b>Tema 1.2: Lexicones; Tesoros</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			
Semana 3	<b>Tema 1.3: Corpus. Tipos</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			
Semana 4	<b>Tema 1.4: Diseño, construcción y evaluación de corpus</b> Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio	<b>Evaluación de corpus</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 5	<b>Tema 2.1: Traducción automática. Estado del Arte</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica I</b> Duración: 04:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Actividad no presencial
Semana 6	<b>Tema 2.2: Sistemas comerciales de traducción automática</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica I</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 7	<b>Tema 2.3: Evaluación de sistemas de traducción automática</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Prueba de métricas</b> Duración: 01:00 OT: Otras actividades formativas	<b>Realización de la práctica I</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 8	<b>Tema 3.1: Recuperación de información. Tipos de sistemas (Booleano; Probabilístico; VSM).</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica II y entrega de la práctica I</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 9	<b>3.2. Extracción de Información (NER; eventos; relaciones)</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica II</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial

Semana 10	<b>3.3. Evaluación. Métricas</b> Duración: 03:00 AC: Actividad del tipo Acciones Cooperativas		<b>Prueba de métricas</b> Duración: 02:00 OT: Otras actividades formativas	<b>Realización de práctica III y entrega de práctica II</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 11	<b>3.4. El sistema FLEX</b> Duración: 03:00 AC: Actividad del tipo Acciones Cooperativas			<b>Realización de práctica III</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 12	<b>4.1: Interlinguas. ¿Qué son? tipos.</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica III y entrega de práctica III</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 13	<b>Tema 4.2: Sistemas basados en interlinguas. Arquitecturas. Componentes</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica IV</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 14	<b>Tema 4.3: 4.3 Desarrollo de componentes</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral		<b>Diseño y desarrollo de componentes</b> Duración: 04:00 PR: Actividad del tipo Clase de Problemas	<b>Realización de práctica IV</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 15	<b>Tema 4.4: Procesos y visualización</b> Duración: 03:00 AC: Actividad del tipo Acciones Cooperativas	<b>Visualización de grafos semánticos</b> Duración: 04:00 OT: Otras actividades formativas		<b>Realización de práctica IV</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 16	<b>Resumen, evaluación y entrega de trabajos</b> Duración: 03:00 AC: Actividad del tipo Acciones Cooperativas		<b>Confección de documentación técnica</b> Duración: 04:00 OT: Otras actividades formativas	<b>Realización de práctica IV y entrega de práctica IV</b> Duración: 04:00 TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Actividad no presencial
Semana 17	<b>Presentaciones de trabajos individuales</b> Duración: 02:00 OT: Otras actividades formativas			

**Nota.-** El cronograma sigue una planificación teórica de la asignatura que puede sufrir modificaciones durante el curso.

**Nota 2.-** Para poder calcular correctamente la dedicación de un alumno, la duración de las actividades que se repiten en el tiempo (por ejemplo, subgrupos de prácticas") únicamente se indican la primera vez que se definen.

## Actividades de Evaluación

Semana	Descripción	Duración	Tipo evaluación	Técnica evaluativa	Presencial	Peso	Nota mínima	Competencias evaluadas
5	Realización de práctica I	04:00	Evaluación continua y sólo prueba final	TG: Técnica del tipo Trabajo en Grupo	No	25%	5 / 10	CE18, CB7, CG9, CG8, CG1
6	Realización de la práctica I	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CG8, CG9, CE19, CB7, CG1
7	Realización de la práctica I	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB7, CG9, CE18, CG8, CG1
8	Realización de la práctica II y entrega de la práctica I	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CG8, CG9, CE18, CE12, CB7, CB9
9	Realización de la práctica II	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB9, CG8, CG9, CE12, CE19, CB7, CG2
10	Realización de práctica III y entrega de práctica II	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB10, CG4, CG8, CE12, CE19, CG12, CB7
11	Realización de práctica III	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CG4, CB7, CB10, CG8, CE12, CE19, CG12
12	Realización de práctica III y entrega de práctica III	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB7, CE12, CE19, CB10, CG4, CG8, CG12
13	Realización de práctica IV	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CG8, CG9, CE16, CB7, CG4
14	Realización de práctica IV	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB7, CG4, CG8, CG9, CE16
15	Realización de práctica IV	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CG8, CG9, CE16, CB7, CG4
16	Realización de práctica IV y entrega de práctica IV	04:00	Evaluación continua y sólo prueba final	TI: Técnica del tipo Trabajo Individual	No	25%	5 / 10	CB7, CG4, CG8, CG9, CE16

## Criterios de Evaluación

La asignatura tiene un marcado carácter práctico, por lo que la interacción con el profesor de manera libre (quedando con él mediante cita, incluso fuera de horas de tutoría) es muy importante. Las clases magistrales se reservan para introducción teórica de las diferentes materias seguidas de interacción a voluntad del alumno para concluir las diferentes prácticas que se plantean de cada tema. Los trabajos son en principio de tipo individual sin que ello suponga que no puedan trabajar en grupo. A medida que se vayan presentando las prácticas se van corrigiendo pasando luego el alumno si lo desea por un examen oral sobre la práctica al objeto de subir nota. La nota final es la media de las 4 prácticas que "pesan" lo mismo cada una.



La evaluación de la asignatura se hará exclusivamente a través de las prácticas. El peso de cada una de ellas es el mismo. La asistencia a clase es obligatoria en al menos un 80% de las clases.

**Práctica**

Se realizarán cuatro prácticas obligatorias individualmente. Hay que aprobar las cuatro prácticas con cinco puntos sobre 10. La nota total es la media de las prácticas. Si alguna práctica estuviese suspensa, el alumno que la ha realizado debería defenderla mediante entrevista oral con el profesor correspondiente. Si algún alumno aprobado quiere subir nota podrá hacerlo mediante presentación oral del trabajo.

## Recursos Didácticos

Descripción	Tipo	Observaciones
Apuntes de la asignatura	Bibliografía	
Aplicaciones informáticas	Equipamiento	Herramientas de manejo y gestión de recursos lingüísticos
artículos y libros	Bibliografía	colección de artículos de referencia

## Otra Información

Schank, R.C., "Conceptual Dependency: A Theory of Natural Language Understanding" Cognitive Psychology, 1972 Vol 3, 532-631
Sowa, John F., ed. (1991) Principles of Semantic Networks: Explorations in the Representation of Knowledge, Morgan Kaufmann Publishers, San Mateo, CA, 1991.
R.Mitkov (Ed.) 2003. The Oxford Handbook of Computational Linguistics. Oxford University Press, (Chapter 21. Lexical Knowledge Acquisition, Chapter 24. Corpus Linguistics).
Igor Boguslavsky, Jesús Cardeñosa, Carolina Gallardo, Luis Iraola. (2005) The UNL Initiative: An Overview. Lecture Notes in Computer Science, Volume 3406 (11 páginas). ISSN: 0302-9743
Jesús Cardeñosa, Carolina Gallardo, Luis Iraola (2005). Using an Interlingua for Document Knowledge Representation. Proceedings of Fourth Conference of the European Society for Fuzzy Logic and Technology. pp 1231-1236.. Sep 09, 2005. Barcelona. ISBN: 84-7653-872-3
Ronen Feldman y James Sanger. "The Text Mining Handbook". Cambridge University Press. 2006.
Jesús Cardeñosa, Carolina Gallardo, Luis Iraola (2006). Interlinguas: A Classical Approach for the Semantic Web. A Practical Case.MICAI 2006: Advances in Artificial Intelligence. Lecture Notes in Computer Science. pp 932 ? 942 pp. Springer Berlin / Heidelberg. ISSN: 302-9743
Boguslavsky I., Cardeñosa J., Gallardo C.; (2008). A Novel Approach to Creating Disambiguated Multilingual Dictionaries. International Journal on Applied Linguistics. Oxford Journals. Oxford University Press. Vol. 30, pp. 70-92. (ISSN: 1477-450X)
J. Davies Goker. "Information Retrieval: Searching in the 21st Century". Wiley. 2009
Bing Liu. Sentiment Analysis and Subjectivity en "Handbook of Natural Language Processing", Second Edition, (eds: N. Indurkha and F. J. Damerau). 2010.
David Nadeau, and Satoshi Sekine (2007). A survey of named entity recognition and classification. Lingvisticae Investigationes 30(1):3-26
Daniel Gildea and Daniel Jurafsky (2002). Automatic labeling of semantic roles. Comput. Linguist. 28, 3 (September 2002), 245-288.