



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

ASIGNATURA

103000654 - Ingeniería Lingüística

PLAN DE ESTUDIOS

10AN - Master Universitario en Ingeniería Informática

CURSO ACADÉMICO Y SEMESTRE

2020/21 - Primer semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	2
5. Descripción de la asignatura y temario.....	3
6. Cronograma.....	5
7. Actividades y criterios de evaluación.....	7
8. Recursos didácticos.....	10
9. Otra información.....	10

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	103000654 - Ingeniería Lingüística
<b>No de créditos</b>	4.5 ECTS
<b>Carácter</b>	Optativa
<b>Curso</b>	Segundo curso
<b>Semestre</b>	Tercer semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10AN - Master Universitario en Ingeniería Informática
<b>Centro responsable de la titulación</b>	10 - Escuela Técnica Superior de Ingenieros Informaticos
<b>Curso académico</b>	2020-21

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Jesus Cardeñosa Lera (Coordinador/a)	L-3310	jesus.cardenosa@upm.es	M - 19:00 - 21:00 Puedo atender tutorías cualquier día y hora previa cita por email.
Igor Boguslavskiy Margolin	2201	igor.boguslavskiy@upm.es	X - 15:00 - 17:00 Puede atender cualquier día y hora previa cita por email

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

### 2.3. Profesorado externo

Nombre	Correo electrónico	Centro de procedencia
Carolina Gallardo Pérez	cgallardo@etsisi.upm.es	ETSISInformáticos

## 3. Conocimientos previos recomendados

---

### 3.1. Asignaturas previas que se recomienda haber cursado

- Sistemas Inteligentes

### 3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Linguística computacional
- bases de datos
- Inteligencia artificial

## 4. Competencias y resultados de aprendizaje

---

### 4.1. 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.

## 4.2. Resultados del aprendizaje

RA59 - Conocer y diseñar los componentes del sistema UNL

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

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

RA57 - Conocer y diseñar sistemas de Traducción automática

RA58 - Conocer y diseñar sistemas de extracción de Información

## 5. Descripción de la asignatura y temario

---

### 5.1. 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.

### 5.2. Temario de la asignatura

#### 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

BORRADOR

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad presencial en aula	Actividad presencial en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>Tema 1.1: Recursos léxicos. ¿Qué son?</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Búsqueda de recursos léxicos y gramaticales</b> Duración: 04:00 OT: Otras actividades formativas	
2	<b>Tema 1.2: Lexicones; Tesauros</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Manejo de aplicaciones de creación y gestión de Tesauros</b> Duración: 04:00 AC: Actividad del tipo Acciones Cooperativas	
3	<b>Tema 1.3: Corpus. Tipos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Búsqueda de recursos (Corpus)</b> Duración: 06:00 OT: Otras actividades formativas	
4	<b>Tema 1.4: Diseño, construcción y evaluación de corpus</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio	<b>Evaluación de corpus</b> Duración: 06:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	<b>Tema 2.1: Traducción automática. Estado del Arte</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica I</b> PI: Técnica del tipo Presentación Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
6	<b>Tema 2.2: Sistemas comerciales de traducción automática</b> Duración: 04:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica I</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
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: 02:00 OT: Otras actividades formativas	<b>Realización de la práctica I</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
8	<b>Tema 3.1: Recuperación de información. Tipos de sistemas (Booleano; Probabilístico; VSM).</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica II y entrega de la práctica I</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 04:00
9	<b>3.2. Extracción de Información (NER; eventos; relaciones)</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Realización de la práctica II</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00

10	<b>3.3. Evaluación. Métricas</b> Duración: 02:00 AC: Actividad del tipo Acciones Cooperativas		<b>Prueba de métricas</b> Duración: 04:00 OT: Otras actividades formativas	<b>Realización de práctica III y entrega de práctica II</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 04:00
11	<b>3.4. El sistema FLEX</b> Duración: 02:00 AC: Actividad del tipo Acciones Cooperativas			<b>Realización de práctica III</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
12	<b>4.1: Interlinguas. ¿Qué son? tipos.</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica III y entrega de práctica III</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 04:00
13	<b>Tema 4.2: Sistemas basados en interlinguas. Arquitecturas. Componentes</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Realización de práctica IV</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
14	<b>Tema 4.3: 4.3 Desarrollo de componentes</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Diseño y desarrollo de componentes</b> Duración: 08:00 PR: Actividad del tipo Clase de Problemas	<b>Realización de práctica IV</b> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final Presencial Duración: 04:00
15	<b>Tema 4.4: Procesos y visualización</b> Duración: 02:00 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> TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 04:00
16	<b>Resumen, evaluación y entrega de trabajos</b> Duración: 02:00 AC: Actividad del tipo Acciones Cooperativas		<b>Confeción de documentación técnica</b> Duración: 04:00 OT: Otras actividades formativas	<b>Realización de práctica IV y entrega de práctica IV</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 04:00
17	<b>Presentaciones de trabajos individuales</b> Duración: 10:00 OT: Otras actividades formativas			

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

\* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.



## 7. Actividades y criterios de evaluación

### 7.1. Actividades de evaluación de la asignatura

#### 7.1.1. Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Realización de práctica I	PI: Técnica del tipo Presentación Individual	No Presencial	04:00	25%	5 / 10	
6	Realización de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
7	Realización de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
9	Realización de la práctica II	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
11	Realización de práctica III	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
13	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
14	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	Presencial	04:00	25%	5 / 10	
15	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	

#### 7.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Realización de práctica I	PI: Técnica del tipo Presentación Individual	No Presencial	04:00	25%	5 / 10	
6	Realización de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
7	Realización de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
8	Realización de la práctica II y entrega de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
9	Realización de la práctica II	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
10	Realización de práctica III y entrega de práctica II	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
11	Realización de práctica III	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
12	Realización de práctica III y entrega de práctica III	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	CE12
13	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
14	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	Presencial	04:00	25%	5 / 10	
15	Realización de práctica IV	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	

16	Realización de práctica IV y entrega de práctica IV	TI: Técnica del tipo Trabajo Individual	No Presencial	04:00	25%	5 / 10	
----	---	---	---------------	-------	-----	--------	--

### 7.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

## 7.2. 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.

## 8. Recursos didácticos

### 8.1. Recursos didácticos de la asignatura

Nombre	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

## 9. Otra información

### 9.1. Otra información sobre la asignatura

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. <i>Linguisticae Investigationes</i> 30(1):3-26
Daniel Gildea and Daniel Jurafsky (2002). Automatic labeling of semantic roles. <i>Comput. Linguist.</i> 28, 3 (September 2002), 245-288.

BORRADOR