

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

103000346 - Ingeniería Lingüística

PLAN DE ESTUDIOS

10AJ - Master Universitario En Inteligencia Artificial

CURSO ACADÉMICO Y SEMESTRE

2023/24 - 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.....	8
9. Otra información.....	9

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	103000346 - Ingeniería Lingüística
No de créditos	5 ECTS
Carácter	Optativa
Curso	Primer curso
Semestre	Primer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	10AJ - Master Universitario en Inteligencia Artificial
Centro responsable de la titulación	10 - Escuela Técnica Superior De Ingenieros Informaticos
Curso académico	2023-24

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Igor Boguslavskiy Margolin	2201	igor.boguslavskiy@upm.es	V - 15:00 - 21:00 Cualquier día y hora previa cita
Mariano Rico Almodovar (Coordinador/a)	2110	mariano.rico@upm.es	Sin horario. Cualquier día y hora previa cita

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

3. Conocimientos previos recomendados

3.1. Asignaturas previas que se recomienda haber cursado

El plan de estudios Master Universitario en Inteligencia Artificial no tiene definidas asignaturas previas recomendadas para esta asignatura.

3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Lingüística computacional
- Procesamiento del lenguaje natural
- Inteligencia artificial
- Aprendizaje automático

4. Competencias y resultados de aprendizaje

4.1. Competencias

CEIA1 - Capacidad de integrar tecnologías y sistemas propios de la Inteligencia Artificial, con carácter generalista, y en contextos más amplios y multidisciplinares

CEIA2 - Capacidad de conectar la tecnología puntera en Inteligencia Artificial con las necesidades de los clientes

CEIA9 - Comprensión del mercado, sus hábitos y necesidades de productos o servicios en el ámbito de la Inteligencia Artificial.

CG11 - Integración del conocimiento a partir de disciplinas diferentes, así como el manejo de la complejidad.

CG12 - Comprensión amplia de las técnicas y métodos aplicables en una especialización concreta, así como de sus límites.

CG6 - Gestión de la información.

CG9 - Aplicación de los métodos de resolución de problemas más recientes o innovadores y que puedan implicar el uso de otras disciplinas.

CG11 - Adquirir conocimientos científicos avanzados del campo de la informática que le permitan generar nuevas ideas dentro de una línea de investigación.

CG14 - Capacidad de leer y comprender publicaciones dentro de su ámbito de estudio/investigación, así como su catalogación y valor científico.

4.2. Resultados del aprendizaje

RA65 - Conocer los componentes que integran cualquier sistema de Ingeniería lingüística

RA66 - Conocer y diseñar sistemas de representación de contenidos

RA67 - Diseño y organización de sistemas de soporte a recursos léxicos

RA68 - Conocer y diseñar sistemas de recuperación de información

RA69 - Conocer y diseñar sistemas de extracción de información

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

La asignatura pretende mostrar al alumno las actividades de la industria de la Ingeniería Lingüística, así como los aspectos más relevantes de la investigación actual.

Se incluyen las formas de representación de los contenidos de un texto escrito de manera independiente de la lengua, así como grandes retos de la industria moderna de la Ingeniería lingüística como: (1) sistemas avanzados de almacenamiento, recuperación y tratamiento masivo de textos/documentos, (2) traducción automática mediante reglas/gramáticas y los sistemas neuronales, y (3) sistemas conversacionales. Se ilustrará al alumno con demos de sistemas reales. El objetivo de esta asignatura es mostrar un área específica al alumno que requiere conocimientos dispersos, como lingüística, computación y estadística avanzada sin los cuales cualquier trabajo de investigación se encuentra con serios problemas.

5.2. Temario de la asignatura

1. Introducción
2. Modelos de representación de contenidos
 - 2.1. Modelos morfológicos, sintácticos y semánticos
 - 2.2. Interlinguas
3. Recursos lingüísticos. Creación, Organización e Infraestructura
 - 3.1. Tipos de recursos
 - 3.2. Lexicons, Tesoros, Diccionarios, Corpus
 - 3.3. Anotación de recursos
 - 3.4. Sistemas de creación y gestión de recursos
4. Traducción automática
 - 4.1. Sistemas clásicos. Uso de GF (Grammatical Framework)
 - 4.2. Sistemas neuronales. Uso de modelos pre-entrenados y plataformas como OpenNMT
5. Almacenamiento y extracción de información
 - 5.1. Information retrieval con Lucene
 - 5.2. Uso de Apache Solr
6. Tratamiento masivo de textos
 - 6.1. Crawling masivo. Tratamiento de ficheros WARC
 - 6.2. Creación de resúmenes mediante técnicas neuronales
7. Sistemas conversacionales
 - 7.1. Sistemas clásicos de detección de intenciones
 - 7.2. Sistemas neuronales

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	Tema 1 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 02:00 OT: Otras actividades formativas	
2	Tema 2.1 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
3	Tema 2.2 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
4	Tema 3.1 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
5	Tema 3.2 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
6	Tema 3.3 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
7	Tema 3.4 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
8	Tema 4.1 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
9	Tema 4.2 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Tutoría Duración: 06:00 OT: Otras actividades formativas	
10	Tema 5 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	Entrega de la práctica I TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 00:00
11	Tema 6 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	

12	Tema 7 Duración: 02:00 LM: Actividad del tipo Lección Magistral		Tutoría Duración: 06:00 OT: Otras actividades formativas	
13	Preparación prácticas estudiantes Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
14	Preparación prácticas estudiantes Duración: 02:00 LM: Actividad del tipo Lección Magistral		Estudio y consulta de material bibliográfico Duración: 01:00 OT: Otras actividades formativas	
15	Preparación prácticas estudiantes Duración: 02:00 LM: Actividad del tipo Lección Magistral		Tutorías Duración: 06:00 OT: Otras actividades formativas	
16	Preparación prácticas estudiantes Duración: 02:00 LM: Actividad del tipo Lección Magistral			Entrega de la práctica II TI: Técnica del tipo Trabajo Individual Evaluación continua y sólo prueba final No presencial Duración: 00:00
17	Evaluaciones Duración: 04:00 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.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
10	Entrega de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	50%	5 / 10	CG11 CGI1 CGI4 CG6 CG12
16	Entrega de la práctica II	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	50%	5 / 10	CG11 CGI1 CGI4 CG6 CEIA2

7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
10	Entrega de la práctica I	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	50%	5 / 10	CG11 CGI1 CGI4 CG6 CG12
16	Entrega de la práctica II	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	50%	5 / 10	CG11 CGI1 CGI4 CG6 CEIA2

7.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

7.2. Criterios de evaluación

La asignatura de Ingeniería Lingüística tiene una parte teórica muy orientada a su aplicación y una parte práctica. 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 dos **prácticas obligatorias** en grupos de dos o tres personas. Hay que aprobar las dos prácticas con al menos cinco puntos sobre 10. La nota total es la media de las de las prácticas. Si alguna práctica estuviese suspensa, el grupo que la ha realizado debería defenderla mediante entrevista oral con el profesor correspondiente. Si algún alumno aprobado quiere subir nota podrá hacerlo mediante examen oral del trabajo. En la convocatoria extraordinaria (julio), ya sea por no haberse presentado a la anterior o por haber suspendido, se deberá hacer una presentación oral a título individual sobre las prácticas entregadas. No podrá haber en julio examen oral sin entrega de prácticas.

Se establece que el método es de evaluación continua, pues el alumno entrega las sucesivas prácticas a medida que las va realizando. Solamente en el caso de que un alumno por decisión propia se presente exclusivamente a la prueba final, deberá hacerlo tras entregar las prácticas.

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Bibliografía	Bibliografía	Amplia bibliografía sobre el tema, marcando algunos artículos y libros específicos obtenibles con facilidad

Aplicaciones de mercado	Equipamiento	acceso a descarga de algunas aplicaciones
Recursos léxicos	Recursos web	Recursos de diferentes tipos obtenibles por descargas gratuitas en web.

9. Otra información

9.1. Otra información sobre la asignatura

Se recomienda esta Bibliografía no exhaustiva que indica los fundamentos de las diferentes tecnologías mencionadas en las clases regulares:

- Quillian M.R., ¿Semantic Memory. Semantic Information Processing?. M. Minsky (Ed.), MIT press, 1968

- 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).

Nirenburg, S., V. Raskin. 2004. Ontological Semantics. The MIT Press. Cambridge, Massachusetts. London, England

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.

=====

La asignatura se relaciona con el ODS4: Calidad en cuanto que los materiales y las enseñanzas vertidas tienen un fundamento basado en materiales e iniciativas amparadas por normativas internacionales de las Naciones Unidas. Asimismo esta asignatura específica se relaciona con el ODS10 de reducción de desigualdades puesto que se enseñan unas técnicas que las propias Naciones Unidas declararon como de garantía de acceso a recursos educacionales a todos los pueblos independientemente de su idioma evitando así áreas lingüísticas de influencia que segregaran el acceso al conocimiento según la lengua en la que estuvieran escritas.