



Guía Docente

Datos Identificativos					2012/13
Asignatura (*)	Linguaxes naturais	Código	614451220		
Titulación	Mestrado Universitario en Enxeñaría de Sistemas Informáticos				
Descritores					
Ciclo	Período	Curso	Tipo	Créditos	
Mestrado Oficial	1º cuatrimestre	Segundo	Optativa	3	
Idioma	Castelán				
Prerrequisitos					
Departamento	Computación				
Coordinación	Vilares Ferro, Jesus	Correo electrónico	jesus.vilares@udc.es		
Profesorado	Alonso Pardo, Miguel angel Gómez Rodríguez, Carlos Vilares Ferro, Jesus	Correo electrónico	miguel.alonso@udc.es carlos.gomez@udc.es jesus.vilares@udc.es		
Web	campusvirtual.udc.es				



Descrición xeral	<p>DESCRIPCIÓN:</p> <p>Esta asignatura aborda el Procesamiento del Lenguaje Natural (NLP, Natural Language Processing), la rama de las ciencias computacionales encargada del diseño e implementación de los elementos software y hardware necesarios para el tratamiento computacional del lenguaje natural, entendiéndose como tal todo lenguaje humano, en contraposición a los lenguajes formales propios del ámbito lógico, matemático, o computacional. El objetivo último que se persigue, y que todavía está muy lejos de conseguirse, es el de la comprensión del lenguaje humano por parte de la computadora.</p> <p>Brevemente, los objetivos generales de la asignatura son:</p> <ul style="list-style-type: none">- Diferenciar entre lenguajes naturales y lenguajes artificiales.- Comprender la complejidad del lenguaje humano.- Familiarizarse con la terminología lingüística.- Conocer los algoritmos, técnicas y métodos más utilizados actualmente para el tratamiento automático de los fenómenos léxicos, sintácticos y semánticos del lenguaje humano.- Asumir que el procesamiento del lenguaje natural no se puede automatizar completamente, pero que se pueden desarrollar soluciones satisfactorias en la práctica.- Identificar los aspectos del lenguaje sobre los que debemos trabajar más para obtener sistemas de procesamiento del lenguaje natural útiles.- Reflexionar sobre los avances realizados en el campo y los errores cometidos a lo largo de las décadas. <p>últimas décadas.</p> <p>CONTEXTUALIZACIÓN:</p> <p>La asignatura Lenguajes Naturales se imparte bajo diversas denominaciones como asignatura optativa de segundo ciclo de II en diversos planes de estudio, entre ellos los de las universidades de Alicante (Ingeniería del Lenguaje Natural, 7,5 créditos LRU), Granada (Interfaces Software en Lenguaje Natural, 6 créditos LRU), País Vasco (Procesamiento del Lenguaje Natural, 6 créditos LRU), Politécnica de Madrid (Lenguaje Natural, 6 créditos LRU), Politécnica de Valencia (Lenguaje Natural, 4,5 créditos LRU) y Vigo (Lenguajes Naturales, 6 créditos LRU). Como caso particular, la Universidad Alfonso X El Sabio incluye dos asignaturas optativas de primer ciclo, denominadas respectivamente Lingüística Computacional I (4,5 créditos LRU) y Lingüística Computacional II (6 créditos LRU) que desarrollan un temario similar.</p> <p>En el caso de la Facultad de Informática de la Universidade da Coruña, se ha optado por una asignatura optativa del Máster en Informática, de 3 créditos ECTS (2 teóricos y 1 práctico). Esta posición se justifica por la fuerte interrelación que mantiene con otras asignaturas del plan de estudios, ya que la asignatura conjuga aspectos de la inteligencia artificial, la teoría de autómatas y lenguajes formales, y la construcción de compiladores. Este hecho permite al alumno relacionar conceptos, métodos y técnicas que se enseñan en asignaturas dispares. Por otra parte, en la asignatura se combinan los aspectos teóricos, centrados en el estudio de ciertos formalismos, con los aspectos prácticos derivados de la realización de algoritmos y programas de ordenador que transforman esos fundamentos teóricos en aplicaciones reales.</p>
-------------------------	--

Competencias da titulación

Código	Competencias da titulación
A1	Análise estatístico.
A5	Capacidade para entender e avaliar especificacións internas e externas.
A7	Dirección, planificación e xestión de proxectos.
A8	Deseño e arquitectura de Sistemas de Información.
A10	Xestión do cambio e do coñecemento.
B1	Capacidade de análise e síntese.
B3	Capacidade de xestión da información.



B4	Capacidade de resolución de problemas.
B5	Toma de decisións.
B6	Traballo en equipo.
B8	Razoamento crítico.
B10	Aprendizaxe autónoma.
B11	Adaptación a novas situacións.
B12	Creatividade.
B14	Iniciativa e espírito emprendedor.
B15	Motivación pola calidade.
C3	Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.
C5	Entender a importancia da cultura emprendedora e coñecer os medios ao alcance das persoas emprendedoras.
C6	Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.
C7	Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.
C8	Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.

Resultados da aprendizaxe			
Competencias de materia (Resultados de aprendizaxe)	Competencias da titulación		
Comprender la descripción lingüística de los fenómenos presentes en el lenguaje natural	AP5 AP10	BP1 BP8 BP10 BP12	
Distinguir entre el nivel léxico, sintáctico y semántico del lenguaje.		BP8 BP12	
Reconocer los fenómenos del lenguaje que son tratables y aquellos que no lo son.		BP8 BP12	
Conocer los algoritmos, técnicas y métodos más utilizados actualmente en el procesamiento del lenguaje natural.	AP1 AP5 AP10	BP1 BP10	CM3 CM6 CM8
Comprender la representación formal de diversos fenómenos léxicos, sintácticos y semánticos del lenguaje humano.	AP5 AP8	BP4 BP5 BP8 BP12 BP15	CM3
Diseñar algoritmos y estructuras de datos para el tratamiento de diversos fenómenos de los lenguajes humanos.	AP5 AP8 AP10	BP3 BP4 BP5 BP6 BP8 BP12 BP15	CM3 CM6
Implementar las soluciones propuestas.	AP5 AP8 AP10	BP3 BP4 BP5 BP6 BP8 BP12 BP15	CM3 CM6



Usar las técnicas y métodos del procesamiento del lenguaje natural para resolver problemas reales de comunicación hombre-máquina.	AP1 AP5 AP8 AP10	BP3 BP4 BP5 BP8 BP10 BP12	CM3 CM6 CM8
Asumir la complejidad del lenguaje humano y las limitaciones de los ordenadores para tratarlo.	AP7	BP8 BP11	CM6
Aceptar soluciones parciales a un problema ante la imposibilidad de obtener soluciones completas.	AP7 AP8	BP5 BP8 BP12 BP15	CM6
Valorar el esfuerzo que requiere realizar avances en un campo complejo.		BP15	CM6 CM7 CM8
Rechazar la compartimentación del saber en áreas cerradas (por ejemplo, ciencias contraposición a humanidades)	AP7	BP6 BP8 BP12 BP14	CM6 CM7 CM8
Colaborar en el desarrollo de recursos y software libre.			CM3 CM5 CM8

Contidos	
Temas	Subtemas
INTRODUCCIÓN AL PROCESAMIENTO DEL LENGUAJE NATURAL	Niveles de análisis. Ambigüedad.
ANÁLISIS LÉXICO	Segmentación de textos. Morfología flexiva y derivativa. Modelización de grandes diccionarios. Autómatas finitos acíclicos deterministas numerados. Traductores de estado finito y morfología de dos niveles.
ETIQUETACIÓN	Modelos de Markov ocultos. Ejecución eficiente de los modelos de Markov ocultos. Técnicas de suavizado. Tratamiento de palabras desconocidas. Aprendizaje de etiquetas basado en transformaciones y dirigido por el error.
ANÁLISIS SINTÁCTICO: GRAMÁTICAS INDEPENDIENTES DEL CONTEXTO	Esquemas de análisis sintáctico. Análisis ascendente. El algoritmo de Earley. Autómatas a pila y programación dinámica. Análisis sintáctico LR generalizado. Representación compartida de los árboles de análisis sintáctico. Análisis sintáctico probabilístico.
ANÁLISIS SINTÁCTICO: GRAMÁTICAS SUAVEMENTE DEPENDIENTES DEL CONTEXTO	Gramáticas de adjunción de árboles. Análisis sintáctico de gramáticas de adjunción de árboles. Autómatas para las gramáticas de adjunción de árboles. Representación compartida de los árboles de derivación. Gramáticas de adjunción de árboles probabilísticas.



ANÁLISIS SEMÁNTICO	Estructuras de rasgos y formalismos basados en unificación. Relaciones léxicas: WordNet y EuroWordNet.
RECUPERACIÓN DE INFORMACIÓN (RI)	Conceptos básicos Modelos de recuperación: booleano, vectorial y probabilístico Los procesos de indexación y recuperación Evaluación en RI RI sobre web. Un caso práctico: Google Aplicaciones del procesamiento del lenguaje natural a RI: la variación lingüística
EXTRACCIÓN DE INFORMACIÓN (EI)	Conceptos básicos Arquitectura de un sistema de EI Tareas de EI Evaluación en EI Ejemplos de sistemas de EI: FASTUS y otros
BÚSQUEDA DE RESPUESTAS (BR)	Conceptos básicos BR vs. RI/EI Arquitectura de un sistema de BR Procesamiento de la pregunta Recuperación y selección de documentos/pasajes Extracción de la respuesta Evaluación en BR
TRADUCCIÓN AUTOMÁTICA (TA)	Conceptos básicos y problemática Técnicas "clásicas" Técnicas estadísticas Aplicaciones en RI interlingüe

Planificación			
Metodoloxías / probas	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Proba de resposta breve	3	4	7
Sesión maxistral	13	13	26
Obradoiro	6	6	12
Prácticas de laboratorio	7	14	21
Mesa redonda	3	3	6
Atención personalizada	3	0	3

*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

Metodoloxías	
Metodoloxías	Descrición
Proba de resposta breve	A final de curso se realizará una prueba escrita con problemas de similar dificultad a los planteados durante el curso.
Sesión maxistral	En las clases presenciais de teoría, el profesor realizará una breve descripción de los contenidos temáticos y de los objetivos básicos perseguidos, con el fin de dotar al alumno de una visión global de la materia. Además tratará de establecer interrelaciones con otros conceptos previamente adquiridos, de forma que se pueda establecer una línea temporal, y expondrá la bibliografía recomendada. Seguidamente pasará a desarrollar los contenidos teóricos, utilizando como método la clase magistral.



Obradoiro	En las clases presenciales de problemas, con el fin de afianzar los conceptos teóricos se presentarán supuestos prácticos, que en un principio serán resueltos por el profesor para orientar a los alumnos. Como actividades no presenciales, se plantearán ejercicios adicionales que el alumno deberá resolver y comentar/corregir con el profesor durante las horas de clases prácticas. Se trata de fomentar la participación de los alumnos y promover, en la medida de lo posible, el diálogo abierto y la valoración de soluciones.
Prácticas de laboratorio	Las clases de prácticas de laboratorio obligan a la implementación de soluciones para un problema dado. Se impondrá una periodicidad en su entrega para fomentar el estudio continuo. El enunciado de las prácticas, que se proporcionará con la suficiente antelación, detallará el problema y las especificaciones, que el alumno deberá respetar estrictamente. Este enunciado se analizará con los alumnos. Posteriormente, la labor del profesor será la de supervisar las sesiones de prácticas, solucionando dudas, corrigiendo errores de interpretación, sugiriendo lecturas, etc.
Mesa redonda	Se fomentará el espíritu crítico de los alumnos y el desarrollo de su capacidad para argumentar mediante la realización de debates y discusiones en forma de tutorías colectivas.

Atención personalizada

Metodoloxías	Descrición
Sesión maxistral Obradoiro Prácticas de laboratorio	La labor del profesor será, tanto en las clases teóricas como prácticas, la de supervisar el trabajo del alumno, solucionando dudas, corrigiendo errores de interpretación, sugiriendo lecturas, etc., no sólo como grupo, sino también como individuo.

Avaliación

Metodoloxías	Descrición	Cualificación
Sesión maxistral	Se valorará la participación activa en las clases y tutorías colectivas.	10
Obradoiro	Se valorará la participación activa en las clases y tutorías colectivas.	7
Prácticas de laboratorio	Se entregarán una o varias prácticas a realizar individualmente o en grupo, según el alcance de las mismas. Se evaluarán la puesta en práctica de los conocimientos de la materia, la metodología de diseño, el funcionamiento y el estilo de programación. Independientemente de la nota obtenida en el resto de los apartados de evaluación, el alumno deberá superar esta parte práctica para superar la asignatura. El examen final es optativo, por lo que el alumno puede optar por incrementar el peso de la parte práctica hasta el 80% de la nota de la asignatura.	40
Mesa redonda	Se valorará la participación activa en las clases y tutorías colectivas.	3
Proba de resposta breve	Prueba escrita con problemas de similar dificultad a los planteados durante el curso. Se evaluarán el dominio de conocimientos teóricos y su aplicación en resolución de problemas. El examen final es optativo, por lo que el alumno puede optar por incrementar el peso de la parte práctica hasta el 80% de la nota de la asignatura.	40
Outros		

Observacións avaliación



Consideraciones generales.

La evaluación formativa del alumno se realizará a través de las prácticas de la asignatura y las actividades relativas a la resolución de problemas. En lo que respecta a las prácticas de laboratorio, éstas se evaluarán de forma continuada al finalizar cada una de ellas. En el enunciado de cada práctica se detallan los criterios específicos de valoración de la misma.

Por otra parte, las actividades de aprendizaje no presenciales se utilizarán para realizar una valoración de la evolución y participación del alumno en el cuatrimestre.

Por último, se realiza una evaluación sumativa del alumno mediante un examen final al término del cuatrimestre. Este será eminentemente práctico para que el alumno pueda demostrar que ha adquirido los conocimientos necesarios y se ha entrenado lo suficiente como para poseer las habilidades precisas para resolver supuestos prácticos. La puntuación asignada a cada una de las preguntas del examen irá consignada en la prueba.

El estudio de la asignatura no puede plantearse como una actividad de estudio memorístico de los algoritmos presentados en clase y de lectura de la bibliografía, sino que deberá tomar una orientación más práctica tal y como se propone en las actividades no presenciales.

El examen final es optativo, por lo que el alumno puede optar por incrementar el peso de la parte práctica hasta el 80% de la nota de la asignatura.

Fontes de información

Bibliografía básica

- Christopher D. Manning y Hinrich Schütza (1999). Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge (Massachusetts, EE.UU.)/Londres (Reino Unido)
- Robert Dale, Hermann Moisi y Harold Somers (eds.) (2000). Handbook of Natural Language Processing. Marcel Dekker, Inc., Nueva York/Basilea
- Christopher D. Manning, Prabhakar Raghavan, y Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge
- A. Arampatzis, Th. P. van der Weide, P. van Bommel y C.H.A. Koster (2000). Linguistically-motivated Information Retrieval. En Vol. 69 de Encyclopedia of Library and Information Science, pág. 201-222. Marcel Dekker
- Peter Jackson y Isabelle Moulinier (2007). Natural language processing for online applications : text retrieval, extraction and categorization (2nd ed.). John Benjamins, Amsterdam/Philadelphia
- James Allen (1995). Natural Language Understanding (2a ed.). The Benjamin/Cummings Publishing Company, Inc., Redwood City, CA, EE.UU
- Marius Pasca (2003). Open-domain question answering from large text collections. CSLI Publications, Standford
- W. Bruce Croft, Donald Metzler y Trevor Strohman (2009). Search Engines: Information Retrieval in Practice. Pearson Education, Upper Saddle River, NJ, USA
- Daniel Jurafsky y James H. Martin (2009). Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (2nd ed.). Pearson Prentice Hall, Upper Saddle River, New Jersey, EE.UU



Bibliografía complementaria	<ul style="list-style-type: none">- Pierre M. Nugues (2006). An introduction to Language Processing with Perl and Prolog. Springer , Berlin - Heidelberg- New York- W. John Hutchings y Harold L. Somers (1992). An Introduction to Machine Translation. Academic Press, Londres/San Diego- Gregory Grefenstette (ed.) (1998). Cross-language information retrieval. Kluwer Academic Publishers, Boston- José Francisco Quesada Moreno y José Gabriel De Amores Carredano (2000). Diseño e implementación de sistemas de traducción automática. Secretariado de Publicaciones de la Universidad de Sevilla- Piek Vossen (ed.) (1998). EuroWordNet. A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Holanda- Eric Wehrli (1997). L?analyse syntaxique des langues naturelles. Masson, París- Ricardo Baeza-Yates y Berthier Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley y ACM Press, Harlow, Inglaterra- Steven Bird, Ewan Klein y Edward Loper (2009). Natural Language Processing with Python. O'Reilly Media, Sebastopol, USA- Klaas Sikkel (1997). Parsing Schemata ? A Framework for Specification and Analysis of Parsing Algorithms, Texts in Theoretical Computer Science ? An EATCS Series. Springer-Verlag, Berlín/Heidelberg/Nueva York- Fernando C. N. Pereira y Stuart M. Shieber (1987). Prolog and Natural Language Analysis, CSLI Lecture Notes 10. Center for the Study of Language and Information, Stanford, CA, EE.UU.- José Luis Vicedo González (2003). Recuperación de Información de alta precisión: Los sistemas de búsqueda de respuestas. Sociedad Española para el Procesamiento del Lenguaje Natural- Jerry R. Hobbs (1993). The generic information extraction system. En Proceedings of the 5th Conference on Message understanding (MUC-5), pág. 87-91 . Morgan Kauffman Publishers, San Francisco, USA
------------------------------------	--

Recomendacións

Materias que se recomenda ter cursado previamente

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións

(*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías