



Guía Docente

Datos Identificativos					2014/15
Asignatura (*)	Linguaxes Naturais	Código	614111625		
Titulación					
Descritores					
Ciclo	Período	Curso	Tipo	Créditos	
1º e 2º Ciclo	1º cuatrimestre	Todos	Optativa	4	
Idioma	CastelánGalego				
Prerrequisitos					
Departamento	Computación				
Coordinación	Vilares Ferro, Jesus	Correo electrónico	jesus.vilares@udc.es		
Profesorado	Vilares Ferro, Jesus	Correo electrónico	jesus.vilares@udc.es		
Web	campusvirtual.udc.es				



Descrición xeral	<p>DESCRICIÓN:</p> <p>Esta materia aborda o Procesamento da Linguaxe Natural (NLP, Natural Language Processing), a rama das ciencias computacionais encargada do deseño e implementación dos elementos software e hardware necesarios para o tratamento computacional da linguaxe natural, entendendo como tal toda linguaxe humana, en contraposición ás linguaxes formais propias do ámbito léxico, matemático, ou computacional. O obxectivo último que se persegue, e que aínda está moi lonxe de se conseguir, é o da comprensión da linguaxe humana por parte da computadora.</p> <p>Brevemente, os obxectivos xerais da materia son:</p> <ul style="list-style-type: none"> - Diferenciar entre linguaxes naturais e linguaxes artificiais. - Comprender a complexidade da linguaxe humana. - Familiarizarse coa terminoloxía lingüística. - Coñecer os algoritmos, técnicas e métodos máis utilizados actualmente para o tratamento automático dos fenómenos léxicos, sintáticos e semánticos da linguaxe humana. - Asumir que o procesamento da linguaxe natural non se pode automatizar completamente, pero que se poden desenvolver solucións satisfactorias na práctica. - Identificar os aspectos da linguaxe sobre os que debemos traballar máis para obter sistemas de procesamento da linguaxe natural útiles. - Reflexionar sobre os avances realizados no campo e os erros cometidos ao longo das últimas décadas. <p>*CONTEXTUALIZACIÓN:</p> <p>A materia Linguaxes Naturais impártese baixo diversas denominacións como materia optativa de segundo ciclo de EI en diversos plans de estudo, entre eles os das universidades de Alacante (Enxeñaría da Linguaxe Natural, 7.5 créditos LRU), Granada (Interfaces Software en Linguaxe Natural, 6 créditos LRU), País Vasco (Procesamento da Linguaxe Natural, 6 créditos LRU), Politécnica de Madrid (Linguaxe Natural, 6 créditos LRU), Politécnica de Valencia (Linguaxe Natural, 4.5 créditos LRU) e Vigo (Linguaxes Naturais, 6 créditos LRU). Como caso particular, a Universidade Alfonso X O Sabio inclúe dúas materias optativas de primeiro ciclo, denominadas respectivamente Lingüística Computacional I (4,5 créditos LRU) e Lingüística Computacional II (6 créditos LRU) que desenvolven un temario similar.</p> <p>No caso da Facultade de Informática da Universidade da Coruña, optouse por unha materia optativa de segundo ciclo de EI de 4 créditos ECTS (2.5 teóricos e 1.5 prácticos). Esta posición xustifícase pola forte interrelación que mantén con outras materias do plan de estudos, xa que a materia conxuga aspectos da intelixencia artificial, a teoría de autómatas e linguaxes formais, e a construción de compiladores. Este feito permite ao alumno relacionar conceptos, métodos e técnicas que se ensinan en materias dispares. Por outra banda, na materia combínanse os aspectos teóricos, centrados no estudo de certos formalismos, cos aspectos prácticos derivados da realización de algoritmos e programas de computador que transforman eses fundamentos teóricos en aplicacións reais.</p>
-------------------------	---

Competencias da titulación	
Código	Competencias da titulación

Resultados da aprendizaxe			
Competencias de materia (Resultados de aprendizaxe)		Competencias da titulación	
Comprender a descrición lingüística dos fenómenos presentes na linguaxe natural.	A1	B1 B3 B4	
Distinguir entre o nivel léxico, sintático e semántico da linguaxe.		B3	
Recoñecer os fenómenos da linguaxe que son tratables e aqueles que non o son.		B3	



Coñecer os algoritmos, técnicas e métodos máis empregados actualmente no procesamento da linguaxe natural.	A1 A5	B1 B4 B9 B15	C3 C6 C8
Comprender a representación formal de diversos fenómenos léxicos, sintáticos e semánticos da linguaxe humana.	A3 A5	B2 B3 B9 B15	C3
Deseñar algoritmos e estruturas de datos para o tratamento de diversos fenómenos das linguaxes humanas.	A3 A5	B2 B3 B8 B9 B10 B15	C3 C6
Implementar as solucións propostas.	A3 A5	B2 B3 B6 B9 B10 B15	C3 C6
Usar as técnicas e métodos do procesamento da linguaxe natural para resolver problemas reais de comunicación home-máquina.	A1 A3 A5 A8	B2 B3 B4 B9 B10	C3 C6 C8
Asumir a complexidade da linguaxe humana e as limitacións dos ordenadores para tratala.	A5	B3	C6
Aceptar solucións parciais a un problema ante a imposibilidade de obter solucións completas.	A5	B3 B9 B15	C6
Valorar o esforzo que require realizar avances nun campo complexo.		B6 B15	C6 C7 C8
Rechazar a compartimentación do saber en áreas cerradas (por exemplo, ciencias en contraposición a humanidades)	A5	B3 B5 B8	C6 C7 C8
Colaborar no desenvolvemento de recursos e software libre.			C3 C5 C8

Contidos	
Temas	Subtemas
INTRODUCCIÓN Ó PROCESAMENTO DA LINGUAXE NATURAL	Niveles de análise. Ambigüedade.
ANÁLISE LÉXICA	Segmentación de textos. Morfoloxía flexiva e derivativa. Modelización de grandes dicionarios. Autómatas finitos acíclicos deterministas numerados. Tradutores de estado finito e morfoloxía de dous niveis.



ETIQUETACIÓN	<p>Modelos de Markov ocultos.</p> <p>Execución eficiente dos modelos de Markov ocultos.</p> <p>Técnicas de suavizado.</p> <p>Tratamento de palabras descoñecidas.</p> <p>Aprendizaxe de etiquetas baseado en transformacións e dirixido polo erro.</p>
ANÁLISE SINTÁCTICA: GRAMÁTICAS INDEPENDENTES DO CONTEXTO	<p>Esquemas de análise sintáctica.</p> <p>Análise ascendente.</p> <p>O algoritmo de Earley.</p> <p>Autómatas a pila e programación dinámica.</p> <p>Análise sintáctica LR xeralizada.</p> <p>Representación compartida das árbores de análise sintáctica.</p> <p>Análise sintáctica probabilística.</p>
ANÁLISE SINTÁCTICA: GRAMÁTICAS SUAVEMENTE DEPENDENTES DO CONTEXTO	<p>Gramáticas de adxunción de árbores.</p> <p>Análise sintáctica de gramáticas de adxunción de árbores.</p> <p>Autómatas para as gramáticas de adxunción de árbores.</p> <p>Representación compartida das árbores de derivación.</p> <p>Gramáticas de adxunción de árbores probabilísticas.</p>
ANÁLISE SEMÁNTICA	<p>Estruturas de rasgos e formalismos baseados en unificación.</p> <p>Relacións léxicas: WordNet e EuroWordNet.</p>
RECUPERACIÓN DE INFORMACIÓN (RI)	<p>Conceptos básicos</p> <p>Modelos de recuperación: booleano, vectorial e probabilístico</p> <p>Os procesos de indexación e recuperación</p> <p>Avaliación en RI</p> <p>RI sobre web. Un caso práctico: Google</p> <p>Aplicacións do procesamento da linguaxe natural a RI: a variación lingüística</p>
EXTRACCIÓN DE INFORMACIÓN (EI)	<p>Conceptos básicos</p> <p>Arquitectura dun sistema de EI</p> <p>Tarefas de EI</p> <p>Avaliación en EI</p> <p>Exemplos de sistemas de EI: FASTUS e outros</p>
PROCURA DE RESPOSTAS (PR)	<p>Conceptos básicos</p> <p>PR vs. RI/EI</p> <p>Arquitectura dun sistema de PR</p> <p>Procesamento da pregunta</p> <p>Recuperación e selección de documentos/pasaxes</p> <p>Extracción da resposta</p> <p>Avaliación en PR</p>
TRADUCIÓN AUTOMÁTICA (TA)	<p>Conceptos básicos e problemática</p> <p>Técnicas "clásicas"</p> <p>Técnicas estadísticas</p> <p>Aplicacións en RI interlingüe</p>

Planificación

Metodoloxías / probas	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Proba de resposta breve	3	4	7
Sesión maxistral	17	17	34
Obradoiro	8	8	16
Prácticas de laboratorio	10	20	30



Mesa redonda	4	4	8
Atención personalizada	5	0	5

*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 realizarase unha proba escrita con problemas de similar dificultade aos plantexados durante o curso.
Sesión maxistral	Nas clases presenciais de teoría, o profesor realizará unha breve descrición dos contidos temáticos e dos obxectivos básicos perseguidos, co fin de dotar ao alumno dunha visión global da materia. Ademais tratará de establecer interrelacións con outros conceptos previamente adquiridos, de forma que se poida establecer unha liña temporal, e exporá a bibliografía recomendada. Seguidamente pasará a desenvolver os contidos teóricos, utilizando como método a clase maxistral.
Obradoiro	Nas clases presenciais de problemas, co fin de afianzar os conceptos teóricos, presentaranse supostos prácticos, que nun principio serán resoltos polo profesor para orientar aos alumnos. Como actividades non presenciais, expóranse exercicios adicionais que o alumno deberá resolver e comentar/corrixir co profesor durante as horas de clases prácticas. Trátase de fomentar a participación dos alumnos e promover, na medida do posible, o diálogo aberto e a valoración de solucións.
Prácticas de laboratorio	As clases de prácticas de laboratorio obrigan á implementación de solucións para un problema dado. Imporase unha periodicidade na súa entrega para fomentar o estudo continuo. O enunciado das prácticas, que se proporcionará con antelación abonda, detallará o problema e as especificacións, que o alumno deberá respectar estritamente. Este enunciado analizarase cos alumnos. Posteriormente, a labor do profesor será a de supervisar as sesións de prácticas, solucionando dúbidas, corrixindo erros de interpretación, suxerindo lecturas, etc.
Mesa redonda	Fomentarase o espírito crítico dos alumnos e o desenvolvemento da súa capacidade para argumentar mediante a realización de debates e discusións en forma de titorías colectivas.

Atención personalizada	
Metodoloxías	Descrición
Prácticas de laboratorio Obradoiro Sesión maxistral	A labor do profesor será, tanto nas clases teóricas como prácticas, a de supervisar o traballo del alumno, solucionando dúbidas, corrixindo erros de interpretación, suxerindo lecturas, etc., non só como grupo, senon tamén como individuo.

Avaliación		
Metodoloxías	Descrición	Cualificación
Proba de resposta breve	Proba escrita con problemas de dificultade semellante aos expostos durante o curso. Avaliaranse o dominio de coñecementos teóricos e a súa aplicación en resolución de problemas. O exame final é optativo, polo que o alumno pode optar por incrementar o peso da parte práctica ata o 80% da nota da materia.	40
Mesa redonda	Valolarase a participación nas clases e titorías colectivas.	3
Prácticas de laboratorio	Entregaranse unha ou varias prácticas a realizar individualmente ou en grupo, segundo o alcance das mesmas. Avaliaranse a posta en práctica dos coñecementos da materia, a metodoloxía de deseño, o funcionamento e a innovación. Independentemente da nota obtida no resto dos apartados de avaliación, o alumno deberá superar esta parte práctica para superar a materia. O exame final é*optativo, polo que o alumno pode optar por incrementar o peso da parte práctica ata o 80% da nota da materia.	40



Obradoiro	Valoralase a participación nas clases e titorías colectivas.	7
Sesión maxistral	Valoralase a participación nas clases e titorías colectivas.	10
Outros		

Observacións avaliación

No caso de novos alumnos, ao non haber xa nin clases teóricas nin horario de laboratorios asignados á materia, a cualificación da materia basearase únicamente na nota obtida no exame, do tipo de resposta breve.

Fontes de información

Bibliografía básica

- Christopher D. Manning y Hinrich Schütze (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

Teoría de Autómatas e Linguaxes Formais/614111301

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