



Teaching Guide				
Identifying Data			2016/17	
Subject (*)	Linguaxes Naturais		Code	614111625
Study programme	Enxeñeiro en Informática			
Descriptors				
Cycle	Period	Year	Type	Credits
First and Second Cycle	1st four-month period	All	Optativa	4
Language	SpanishGalician			
Teaching method	Face-to-face			
Prerequisites				
Department	Computación			
Coordinador	Vilares Ferro, Jesus	E-mail	jesus.vilares@udc.es	
Lecturers	Vilares Ferro, Jesus	E-mail	jesus.vilares@udc.es	
Web	moodle.udc.es			



General description	DESCRICIÓN: 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 lóxico, 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. Brevemente, os obxectivos xerais da materia son: - 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. * CONTEXTUALIZACIÓN: 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. 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.
---------------------	--

Study programme competences	
Code	Study programme competences
A1	Aprender de maneira autónoma novos coñecementos e técnicas avanzadas axeitadas para a investigación, o deseño e o desenvolvemento de sistemas e servizos informáticos.
A3	Concibir e planificar o desenvolvemento de aplicacións informáticas complexas ou con requisitos especiais.
A5	Saber especificar, deseñar e implementar sistemas intelixentes cando as solucións convencionais non resultaren satisfactorias.
A8	Concibir, despregar, organizar e xestionar un servizo informático complexo.
B1	Aprender a aprender.
B2	Resolver problemas de forma efectiva.
B3	Aplicar un pensamento crítico, lóxico e creativo.
B4	Aprendizaxe autónoma.



B5	Traballar de forma colaborativa.
B6	Comportarse con ética e responsabilidade social como cidadán e como profesional.
B8	Traballar en equipos de carácter interdisciplinar.
B9	Capacidade para tomar decisións.
B10	Capacidade de xestión da informática (captación e análises da información).
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.

Learning outcomes			
Learning outcomes	Study programme competences		
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	B1 B4	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

Contents	
Topic	Sub-topic
INTRODUCCIÓN Ó PROCESAMENTO DA LINGUAXE NATURAL	Niveles de análise. Ambigüidade.
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	Modelos de Markov ocultos. Execución eficiente dos modelos de Markov ocultos. Técnicas de suavizado. Tratamento de palabras descoñecidas. Aprendizaxe de etiquetas baseado en transformacións e dirixido polo erro.
ANÁLISE SINTÁCTICA: GRAMÁTICAS INDEPENDENTES DO CONTEXTO	Esquemas de análise sintáctica. Análise ascendente. O algoritmo de Earley. Autómatas a pila e programación dinámica. Análise sintáctica LR xeralizada. Representación compartida das árbores de análise sintáctica. Análise sintáctica probabilística.
ANÁLISE SINTÁCTICA: GRAMÁTICAS SUAVEMENTE DEPENDENTES DO CONTEXTO	Gramáticas de adxunción de árbores. Análise sintáctica de gramáticas de adxunción de árbores. Autómatas para as gramáticas de adxunción de árbores. Representación compartida das árbores de derivación. Gramáticas de adxunción de árbores probabilísticas.
ANÁLISE SEMÁNTICA	Estruturas de rasgos e formalismos baseados en unificación. Relacións léxicas: WordNet e EuroWordNet.
RECUPERACIÓN DE INFORMACIÓN (RI)	Conceptos básicos Modelos de recuperación: booleano, vectorial e probabilístico Os procesos de indexación e recuperación Avaliación en RI RI sobre web. Un caso práctico: Google Aplicacións do procesamento da linguaxe natural a RI: a variación lingüística
EXTRACCIÓN DE INFORMACIÓN (EI)	Conceptos básicos Arquitectura dun sistema de EI Tarefas de EI Avaliación en EI Exemplos de sistemas de EI: FASTUS e outros



PROCURA DE RESPOSTAS (PR)	Conceptos básicos PR vs. RI/EI Arquitectura dun sistema de PR Procesamento da pregunta Recuperación e selección de documentos/pasaxes Extracción da resposta Avaliación en PR
TRADUCCIÓN AUTOMÁTICA (TA)	Conceptos básicos e problemática Técnicas "clásicas" Técnicas estadísticas Aplicacións en RI interlingüe

Planning				
Methodologies / tests	Competencies	Ordinary class hours	Student?s personal work hours	Total hours
Short answer questions	A5 B2 B3 B9 C6	3	4	7
Guest lecture / keynote speech	A1 A8 B1 B3 B4 B6 B8 B15 C5	17	17	34
Workshop	B2 B3 B5 B6 C6	8	8	16
Laboratory practice	A1 A3 A5 A8 B1 B2 B3 B4 B5 B6 B8 B9 B10 B15 C3 C5 C6 C7 C8	10	20	30
Panel discussion	A3 B2 B3 B5 B6 B9 B15 C6 C7 C8	4	4	8
Personalized attention		5	0	5
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description
Short answer questions	A final de curso realizarase unha proba escrita con problemas de similar dificultade aos plantexados durante o curso.
Guest lecture / keynote speech	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.
Workshop	Nas clases presenciais de problemas, co fin de afianzar os conceptos teóricos, presentaranse supostos prácticos, que nun principio serán resolto polo profesor para orientar aos alumnos. Como actividades non presenciais, exporanse 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.
Laboratory practice	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 abunda, 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.
Panel discussion	Fomentarse 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.

Personalized attention	
Methodologies	Description



Laboratory practice Workshop Guest lecture / keynote speech	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.
--	---

Assessment			
Methodologies	Competencies	Description	Qualification
Short answer questions	A5 B2 B3 B9 C6	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
Panel discussion	A3 B2 B3 B5 B6 B9 B15 C6 C7 C8	Valolarase a participación nas clases e titorías colectivas.	3
Laboratory practice	A1 A3 A5 A8 B1 B2 B3 B4 B5 B6 B8 B9 B10 B15 C3 C5 C6 C7 C8	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
Workshop	B2 B3 B5 B6 C6	Valolarase a participación nas clases e titorías colectivas.	7
Guest lecture / keynote speech	A1 A8 B1 B3 B4 B6 B8 B15 C5	Valolarase a participación nas clases e titorías colectivas.	10
Others			

Assessment comments
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.

Sources of information



Basic	- James Allen (1995). Natural Language Understanding (2a ed.). The Benjamin/Cummings Publishing Company, Inc., Redwood City, CA, EE.UU - Robert Dale, Hermann Moisi y Harold Somers (eds.) (2000). Handbook of Natural Language Processing. Marcel Dekker, Inc., Nueva York/Basilea - 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 - Christopher D. Manning y Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge (Massachusetts, EE.UU.)/Londres (Reino Unido) - Christopher D. Manning, Prabhakar Raghavan, y Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - Peter Jackson y Isabelle Moulinier (2007). Natural language processing for online applications : text retrieval, extraction and categorization (2nd ed.). John Benjamins, Amsterdam/Philadelphia - W. Bruce Croft, Donald Metzler y Trevor Strohman (2009). Search Engines: Information Retrieval in Practice. Pearson Education, Upper Saddle River, NJ, USA - Marius Pasca (2003). Open-domain question answering from large text collections . CSLI Publications, Stanford - 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
Complementary	- Gregory Grefenstette (ed.) (1998). Cross-language information retrieval. Kluwer Academic Publishers, Boston - Ricardo Baeza-Yates y Berthier Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley y ACM Press, Harlow, Inglaterra - W. John Hutchings y Harold L. Somers (1992). An Introduction to Machine Translation. Academic Press, Londres/San Diego - Piek Vossen (ed.) (1998). EuroWordNet. A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Holanda - 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. - Eric Wehrli (1997). L?analyse syntaxique des langues naturelles. Masson, París - 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 - 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 - 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 - Pierre M. Nugues (2006). An introduction to Language Processing with Perl and Prolog. Springer , Berlin - Heidelberg- New York - Steven Bird, Ewan Klein y Edward Loper (2009). Natural Language Processing with Python. O'Reilly Media, Sebastopol, USA - 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

Recommendations

Subjects that it is recommended to have taken before

Teoría de Autómatas e Linguaxes Formais/614111301

Subjects that are recommended to be taken simultaneously

Subjects that continue the syllabus



Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.