



Universidad de Puerto Rico
Recinto Universitario de Mayagüez
Colegio de Artes y Ciencias
Departamento de Estudios Hispánicos
Programa de Bachillerato en Ciencias en Biolingüística



PRONTUARIO OFICIAL

**PROCESAMIENTO DE LENGUAS NATURALES
LING/ICOM/CIIC 5180/6180**

Horas crédito: 3	Horas contacto: 3 horas de conferencia
Requisitos previos: LING 4010 y (COMP 3010 o CIIC 3015 o INGE 3016) y (ESMA 3016 o ESMA 4001 o ININ 4010)	Requisitos concurrentes: N/A
Descripción del curso (español): Introducción al estudio de métodos de búsqueda y de aprendizaje en el diseño y análisis de representaciones y algoritmos computacionales para el procesamiento de lenguas naturales. Implementación de algoritmos de aprendizaje lineales y no-lineales, con atención particular a redes neuronales. Uso de modelos lingüísticos en el etiquetaje y clasificación de secuencias. Aplicación de conceptos de teoría sintáctica y de lenguas formales en el análisis estructural de lenguas naturales. Representación, denotación y computación del significado mediante formalismos composicionales. Revisión de herramientas en línea y de tecnologías de lenguaje humano. Discusión de las relaciones interdisciplinarias entre el procesamiento de lenguas naturales, la lingüística computacional y la inteligencia artificial.	
Descripción del curso (inglés): Introduction to the study of search and learning methods in designing and analyzing computational representations and algorithms for natural language processing. Implementation of linear and non-linear learning algorithms, with special attention to neural networks. Use of language models in sequence labeling and classification. Application of concepts in syntax and formal language theory to natural language parsing. Meaning representation, denotation and computation by means of compositional formalisms. Survey of on-line tools and human language technologies. Discussion of the interdisciplinary relationships between natural language processing, computational linguistics and artificial intelligence.	
Objetivos: Al finalizar el curso, el estudiantado estará capacitado para: <ol style="list-style-type: none">1. diseñar y analizar algoritmos computacionales para el procesamiento de lenguas naturales.2. construir implementaciones en aspectos específicos del procesamiento de lenguas naturales.3. implementar tecnologías de lenguaje humano para producir representaciones estructurales de lenguas naturales.4. utilizar herramientas tecnológicas disponibles para la creación de aplicaciones de procesamiento de lenguas naturales.5. etiquetar un corpus siguiendo diferentes modelos de procesamiento de lenguas naturales.	

6. identificar constituyentes, marcos de subcategorización y relaciones gramaticales de dependencia en estructuras sintácticas.
7. usar analizadores sintácticos de diferentes técnicas y algoritmos para representar estructuras sintácticas.
8. utilizar métodos formales para la representación y composición de estructuras semánticas.
9. identificar características de los algoritmos de segmentación y de resolución de referencialidad y coherencia en el procesamiento computacional del discurso.
10. reconocer y evaluar implicaciones éticas en la creación y utilización de tecnologías de lenguaje humano.

Bosquejo de contenido:

<i>Temas a cubrir</i>	<i>Horas contacto</i>
Introducción: temas y disciplinas relacionadas con el PLN	2 horas
Aprendizaje lineal	4 horas
Aprendizaje mediante redes neuronales	4 horas
Aplicaciones lingüísticas de clasificación	3 horas
Aprendizaje no supervisado	3 horas
Modelos lingüísticos	3 horas
Etiquetaje de secuencias	3 horas
Aplicaciones de etiquetaje de secuencias	3 horas
Teoría de lenguas formales	3 horas
Representaciones sintácticas y ambigüedad	2 horas
Análisis estructural de lenguas independientes del contexto	4 horas
Semántica composicional y cálculo lambda	4 horas
Semántica de predicados y argumentos	3 horas
Aplicaciones: traducción automática	3 horas
Conclusiones	1 hora
<i>Total de horas: (deben ser equivalentes a las horas crédito del curso)</i>	45 horas

Estrategias instruccionales:

- conferencia
 discusión
 cómputos
 laboratorio
 seminario con presentación formal
 seminario sin presentación formal
 taller
 taller de arte
 práctica
 viaje
 tesis
 problemas especiales
 tutoría
 investigación
 otros, especifique:

Recursos mínimos disponibles:

computadora y sistema de proyección digital

Estrategias de evaluación y su peso relativo:

	Por ciento
<input checked="" type="checkbox"/> pruebas escritas	50
<input checked="" type="checkbox"/> informes orales	10
<input checked="" type="checkbox"/> proyecto	25
<input checked="" type="checkbox"/> asignaciones o pruebas cortas	15
TOTAL: 100%	100

Sistema de calificación:

- cuantificable (de letra)
 no cuantificable

Curva estándar**100-90 A; 89-80 B; 79-70 C; 69-60 D; 59-50 F****Bibliografía:**

TEXTO: A discreción del profesor. Se recomienda Eisenstein (2019).

Berwick, Robert & Stabler, Edward & (2019). *Minimalist Parsing*. Oxford: OUP.

Bird, Steven; Klein, Ewan & Loper, Edward (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.

Clark, Alexander; Fox, Chris & Lappin, Shalom, eds. (2013). *The Handbook of Computational Linguistics and Natural Language Processing*. Oxford: Blackwell.

Dickinson, Markus; Brew, Chris & Meurers, Detmar (2013). *Language and Computers*. Oxford: Wiley-Blackwell.

Eisenstein, Jacob (2019). *Introduction to Natural Language Processing*. Cambridge, MA: The MIT Press.

Grune, Dick & Jacobs, Criel (2008). *Parsing Techniques*. Springer.

Jurafsky, Daniel & Martin, James. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 2da. edición. New Jersey: Prentice Hall.

Kurdi, Mohammed (2016) *Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax*. Londres: Wiley.

Kurdi, Mohammed (2017) *Natural Language Processing and Computational Linguistics: Semantics, Discourse and Applications*. Londres: Wiley.

Levelt, Willem J.M. (2008). *An Introduction to the Theory of Formal Languages and Automata*. Amsterdam: John Benjamins.

Linz, Peter (2012). *An Introduction to Formal Languages and Automata*. 5ta. edición. Sudbury, Massachusetts: Jones & Bartlett Learning.

Manning, Christopher & Schütze, Heinrich (1999/2003). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: The MIT Press.

Nugues, Pierre M. (2014) *Language Processing with Perl and Prolog*. 2da. edición. Berlín: Springer.

Ojeda, Almerindo. (2013) *A Computational Introduction to Linguistics*. Stanford: CSLI Publications.

Pustejovsky, James & Stubbs, Amber (2013). *Natural Language Annotation for Machine Learning*. Sebastopol, CA: O'Reilly Media.

Ruslan, Mitkov (2005). *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.

Sportiche, Dominique; Koopman, Hilda & Stabler, Edward (2014) *An Introduction to Syntactic Analysis and Theory*. Oxford: Wiley-Blackwell

RECURSOS EN INTERNET

Child Language Data Exchange System <http://childes.psy.cmu.edu/>

Corpus multi-medios de producción lingüística infantil.

FrameNet <https://framenet.icsi.berkeley.edu/fndrupal/>
Base de datos léxica legible por máquinas y humanos.

International Phonetic Association <http://www.langsci.ucl.ac.uk/ipa/>
Plataforma de recursos sobre el Alfabeto Fonético Internacional.

The Lambda Calculator <http://dylanbumford.com/LambdaCalculator/>
Programa interactivo para practicar cálculo lambda de tipos.

Linguistic Data Consortium <https://www ldc.upenn.edu/>
Repositorio y punto de distribución de recursos lingüísticos.

Natural Language Toolkit <http://www.nltk.org/>
Plataforma para construir programas en Python para trabajar con datos en lenguas naturales.

OT-Help <http://people.umass.edu/othelp/>
Herramientas para el estudio de tipología lingüística en versiones en paralelo y en serie de la Teoría de la Optimalidad y Gramática Armónica.

Penn TreeBank Project <http://www.cis.upenn.edu/~treebank/>
Banco de árboles sintácticos a partir de un corpus natural anotado.

Praat: Doing Phonetics by Computer <http://www.fon.hum.uva.nl/praat/>
Página electrónica del programa de análisis fonético-acústico de Boersma y Weenink.

Rutgers Optimality Archive <http://roa.rutgers.edu/>
Portal de la colección de artículos sobre la Teoría de la Optimalidad.

Speech and Language Processing <http://www.prenhall.com/jurafsky-martin/>
Portal del texto de Jurafsky & Martin con recursos para el instructor.

The Stanford Parser <http://nlp.stanford.edu/software/lex-parser.shtml>
Analizador sintáctico estadístico desarrollado por el Stanford NLP Group.

WordNet <http://wordnet.princeton.edu/>
Base de datos léxica del inglés.

Acomodo Razonable:

El estudiantado puede orientarse y solicitar los servicios del Programa de Acomodo Razonable, que está ubicado en el Departamento de Consejería y Servicios Psicológicos (Decanato de Estudiantes, oficina DE21, en los bajos de Asistencia Económica). Para aclarar cualquier duda u obtener más información, puede comunicarse al teléfono directo [787-265-3864](tel:787-265-3864) o a través del cuadro [787-832-4040](tel:787-832-4040), extensiones 2040, 3372 o 3864.

Integridad Académica:

La Universidad de Puerto Rico promueve los más altos estándares de integridad académica y científica. El Artículo 6.2 del Reglamento General de Estudiantes de la UPR (Certificación Núm. 13, 2009-2010, de la Junta de Síndicos) establece que “la deshonestidad académica incluye, pero no se limita a:

acciones fraudulentas, la obtención de notas o grados académicos valiéndose de falsas o fraudulentas simulaciones, copiar total o parcialmente la labor académica de otra persona, plagiar total o parcialmente el trabajo de otra persona, copiar total o parcialmente las respuestas de otra persona a las preguntas de un examen, haciendo o consiguiendo que otro tome en su nombre cualquier prueba o examen oral o escrito, así como la ayuda o facilitación para que otra persona incurra en la referida conducta”. Cualquiera de estas acciones estará sujeta a sanciones disciplinarias en conformidad con el procedimiento disciplinario establecido en el Reglamento General de Estudiantes de la UPR vigente.

Hostigamiento Sexual: La certificación 130-2014-2015, indica:

El hostigamiento sexual en el empleo y en el ambiente de estudio es una práctica ilegal y discriminatoria, ajena a los mejores intereses de la Universidad de Puerto Rico. Toda persona que entienda ha sido objeto de actuaciones constitutiva de hostigamiento sexual en la Universidad de Puerto Rico podrá quejarse para que se investigue, de ser necesario, y se tome la correspondiente acción por parte de las autoridades universitarias. Si quien reclama fuera estudiante, deberá referir su queja a la Oficina de la Procuradora Estudiantil o al Decanato de Estudiantes.

La certificación 06-43 del Senado Académico indica:

“Las guías académicas para el ofrecimiento de cursos en línea”, define: Cursos presenciales son aquellos que tienen menos de un 25% de las horas contacto regular del curso a través de la Internet. Así, un curso de 3 créditos, será considerado “presencial” si, de las 45 horas de contacto regular, 11 o menos son a través de la Internet.

Incluye anejos:

Si
No