



ASIGNATURA:	PROCESAMIENTO DEL LENGUAJE NATURAL	MODALIDAD:	Cuatrimestral
DEPARTAMENTO:	ING. EN SIST. DE INFORMACION	HORAS SEM.:	4 horas
AREA:	ELECTIVA	HORAS/AÑO:	64 horas
BLOQUE	TECNOLOGÍA APLICADA	HS RELOJ/AÑO	48
		NIVEL:	3°
		AÑO DE DICTADO	2019

OBJETIVOS

- Profundizar los conocimientos relacionados a la Sintaxis y Semántica del Lenguaje, Probabilidad, Álgebra Lineal, Algoritmos y Estructuras de Datos y Paradigmas de Programación.
- Exponer las técnicas y herramientas modernas utilizadas para el desarrollo de sistemas robustos y prácticos que puedan comunicarse con los usuarios en lenguaje natural. Así como introducir los modelos, teorías, y técnicas de los enfoques computacionales para el tratamiento computacional del lenguaje natural.
- Comprender la aplicación de procesos, algoritmos y modelos propuestos por la Ingeniería del Lenguaje Natural. Así como las características de diferentes modelos del Lenguaje Natural, sus desafíos y limitaciones actuales.
- Desarrollar habilidades para el análisis crítico de metodología de investigación aplicada a un problema concreto basado en las competencias del Ingeniero en Sistemas de Información, utilizando herramientas de desarrollo de software de uso libre y gratuito.
- Adquirir nociones sobre investigación aplicada a la industria (I+D+i), así como su implementación y comunicación efectiva a través de un caso real.
- Desarrollar habilidades para implementar campos avanzados de Procesamiento de Lenguaje Natural, tales como: *Clasificación de artículos, Generación de lenguaje escrito, Análisis de sentimientos de texto y voz, generación de Respuestas a través de Preguntas, Descripción automática de imágenes con lenguaje natural, Resúmenes Automáticos, Bots de Texto, Procesamiento de Lenguaje Natural Robusto*, entre otros.

PROGRAMA ANALITICO

1. Introducción al Procesamiento del Lenguaje Natural. Estado del Arte y Aplicaciones actuales

Introducción formar a la asignatura. Definiciones. Problemas frecuentes. Diferentes tipos de categorías. Estado del Arte y aplicación del mismo a la industria desde la mirada del Ingeniero en Sistemas de Información. Limitaciones.

2. Análisis morfológico y etiquetación morfosintáctica

Definiciones de Lenguaje Natural Español. Etiquetado morfosintáctico. Etiquetado basado en reglas. Etiquetado estocástico. Etiquetado basado en transformación. Etiquetado POST (Part of Speech Tagging).



3. Unidades lingüísticas y representaciones vectoriales de palabras

Palabra como unidad básica. Repaso Álgebra Vectorial. Multipalabras o unidades léxicas. Acortamiento. Composición. Clases de palabras. Lematización. One hot encoding. Bag of Words (BoW). CBOW. Skip-gram. Autoencoders. Transformadores. Otros modelos, estado del arte.

4. Modelos Secuencia a Secuencia

Seq2seq network. Procesamiento a nivel de caracteres. Mecanismos de Atención. Procesamiento a nivel de palabra. Embeddings. Introducción traducción de máquinas. Métricas para evaluación de modelos de Procesamiento del Lenguaje Natural.

5. Clasificación de Textos. Minado de Texto.

Modelos computacionales de clasificación y etiquetado de textos. Revisión de corpus. Práctica y laboratorio intensivo en Python. Preparación de textos para análisis y etiquetado. Descubrimiento de patrones en texto no estructurado. Análisis automático de texto.

6. Modelos de Lenguaje Condicionales

Breve repaso a probabilidad y estadística. Gramática estocástica: N-gramas. Modelo de consulta del lenguaje probabilístico. Limitaciones y ventajas del modelado a través de N-gramas

7. Reconocimiento de Entidades Nombradas (NER)

Detección y extracción de Entidades. Expresiones Regulares. Limitaciones. Algoritmos y técnicas para su detección. Desambiguación de Entidades.. Métodos de Evaluación: valor-F. Ejemplos de importancia en la industria.

8. Modelado de Preguntas y Respuestas

Definición del concepto de “Recuperación de la Información” (R-I). Método de Filtrado, Encaminado, Routeo. Definición del concepto “Preguntas y Respuestas” (P-R) y “Búsqueda de Respuestas” (B-R). Dominios libres y dominios específicos. Fuentes de información. Tipos de usuarios. Precisión, Completitud y Relevancia de la respuesta. Taxonomía de preguntas. P-R. Repaso unidades anteriores para implementación de sistemas de P-R.

9. Generación automática de resúmenes

Recuperación de la Información (R-I). Minería de textos. Repaso Estadística y Procesamiento Lenguaje Natural. Seguimiento de tópicos. Agrupamiento. Vinculación conceptual. Generación automática de texto de dominio restringido. Competencias internacionales de Generación Automática de Resúmenes. Condiciones y Formas de evaluación: ROUGE, BLUE score, Evaluación Humana.

10. Reconocimiento de voz. Texto a voz

Sistemas y algoritmos de código abierto para reconocimiento de voz. Estado del arte y limitaciones del reconocimiento de voz. Palabras aisladas, Palabras conectadas, Palabras continuas. Problema de múltiples oradores en simultáneo. Procesamiento de Texto, generación de la prosodia, generación de voz sintética. Grafemas. Fonemas. Silabificación. Errores de cuadrados medios y Coeficiente de correlación como métodos de evaluación de algoritmos y modelos.



11. Trabajo Práctico Cuatrimestral

Durante todo el cuatrimestre se desarrollarán Prácticas en laboratorio que soporten la teoría vista en clase y a sí mismo darán soporte al alumno a realizar un Trabajo Práctico Integrador de complejidad suficiente para que el alumno demuestre conocimiento profundo y comprensión de los temas vistos en clase.

BIBLIOGRAFIA OBLIGATORIA

1. Allen J. Natural Language Understanding, Addison-Wesley, 1987.
2. Austin J. How to do things with words. Oxford University Press, 1988.
3. Bing Liu (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press. ISBN: 9781139084789. <https://doi.org/10.1017/CBO9781139084789>
4. Candalija Reina, J.A., 1998, "Sobre la cientificidad de la gramática: el uso de corpora informatizados como método de análisis lingüístico". Estudios de Lingüística Cognitiva. Ed. J.L
5. Christopher D. Manning y Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing. The MIT Press. ISBN-10: 0262133601. ISBN-13: 78-0262133609.
6. Cristian Cardellino, Serena Villata, Laura Alonso Alemany, Elena Cabrio: Information Extraction with Active Learning: A Case Study in Legal Text. CICLing (2) 2015: 483-494
7. Dale R., Moisl H., Somers H. Handbook of Natural Language Processing. Dekker, 2000.
8. Daniel Jurafsky & James H. Martin (2009). Speech and Language Processing, 2nd Edition. Pearson, ISBN 978-0131873216. El borrador de la tercera edición en preparación está disponible
9. Fellbaum C. WordNet: An Electronic Lexical Database. MIT Press, 1998.
10. Gazdar G., Mellish C. Natural Language Processing in Prolog: an introduction to computational linguistics, Addison-Wesley 1989.
11. Grishman, R., 1986, Computational Linguistics: An Introduction, Cambridge University Press, Cambridge.
12. Indurkha, N. y Damerau, F. J. (2010). Handbook of natural language processing. Chapman and Hall/CRC.
13. Iwanska I., Shapiro Stuart. Natural Language Processing and Knowledge Representation, MIT Press, 2000.
14. Juan-Manuel Torres-Moreno (2014). Automatic Text Summarization. Wiley-ISTE. ISBN: 978-1-119-04407-9 <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1119044073.html>
15. Laura Manoiloff, Maria Constanza Carando, Maria Cecilia Defagó, Laura Alonso Alemany, Cecilia Ferrero, Daiana Cesaretti, Adrián Ramirez and Juan Seguí (2015). The cognitive processing
16. Manning C. Schütze H. Foundations of Statistical Natural Language Processing. MIT Press, 1999.
17. Marcu D. The theory and Practice of discourse parsing and summarization. MIT Press 2000.
18. Mariani, Joseph; Francopoulo, Gil; Paroubek, Patrick; Vernier, Frédéric (2019), «The NLP4NLP Corpus (I): 50 Years of Research in Speech and Language Processing», Frontiers in Research
19. Mitkov, R. (ed.) (2003): The Oxford Handbook of Computational Linguistics, Oxford, Oxford University Press.
20. Moreno Fernandez, F. (1990), "Lingüística informática e informática lingüística", Lingüística Española Actual, 12/1, 5-16.
21. Noah A. Smith (2011). Linguistic Structure Prediction. Morgan & Claypool Publishers. ISBN: 978-1608454051. DOI <https://doi.org/10.2200/S00361ED1V01Y201105HLT013>
22. P. Estrella and Nikos Tsourakis, "Migrating from ISO/IEC 9126 to SQUARE: A case study on the evaluation of medical speech translation systems", book chapter in "Design Development and
23. Roche E., Schabes Yves. Finite-state Language Processing. MIT Press 1997
24. Rodríguez, H. (2000) "Técnicas básicas en el tratamiento informático de la lengua". Quark. Ciencia, Medicina, Comunicación y Cultura 19: 26-34.
25. Rojo, G. (2005-2006), "Informática y Lingüística: Las lenguas en la sociedad del conocimiento", Boletín de RedIRIS, 74-75. [<http://www.rediris.es/rediris/boletin/74-75/ponencia1.pdf>]
26. Smith G. Computers and Human Language. Oxford University Press, 1991.
27. Tordera, J.C. (2012), El abecedé de la Lingüística computacional, Madrid: Arco/Libros.
28. Yoav Goldberg (2017). Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers. ISBN: 978-1627052986. DOI <https://doi.org/10.2200/S00762ED1V01Y201703HLT0>
29. Zulaica Hernandez, I. (2016), "Lingüística de Corpus", en J. Gutiérrez-Rexach (ed.) Enciclopedia de Lingüística Hispánica, Vol. 1, London: Routledge, 216-224.

BIBLIOGRAFIA COMPLEMENTARIA



- Documentación y papers a ser entregados por el profesor.
- ARES, F. (2008), El robot enamorado. Una historia de la Inteligencia Artificial, Barcelona: Ariel. [caps. 7 e 11]
- Stanford CoreNLP: <https://stanfordnlp.github.io/CoreNLP/>
- UDPipe: <https://github.com/ufal/udpipe>
- Apache OpenNLP: <https://opennlp.apache.org/>
- Natural Language Toolkit (NLTK): <http://www.nltk.org/>
- Revista Procesamiento Lenguaje Natural en Español: <http://rua.ua.es/dspace/handle/10045/1112>
- Jacob Eisenstein. [Natural Language Processing](#)
- Yoav Goldberg. [A Primer on Neural Network Models for Natural Language Processing](#)

CORRELATIVAS

Para Cursar:

Cursadas:

- Análisis de Sistemas
- Sintaxis y Semántica del Lenguaje
- Paradigmas de Programación

Para Rendir:

Aprobadas:

- Análisis de Sistemas
- Sintaxis y Semántica del Lenguaje
- Paradigmas de Programación