


Información académica

Año de presentación (*)
2019

1-a-

Departamento docente que inicia el tramite:	
Departamento de Computación	
Nombre del curso:	
Procesamiento de Lenguaje Natural	
Nombre, Cargo y Título del docente responsable:	
-Franco Martín Luque, Profesor Invitado, Doctor en Cs. de la Computación	
En caso de dictarse en paralelo con una materia de grado, nombre de la misma:	
Introducción al Procesamiento de Lenguaje Natural	
Nombre y Título de los docentes que colaboran con el dictado del curso (*) (*):	
--	
Fecha propuesta para el primer dictado luego de la aprobación:	
Julio-Agosto de 2019	

Duración:

Duración total en horas	32
Duración en semanas	4

Distribución carga horaria:

Número de horas de clases teóricas	16
Número de horas de clases de problemas	--
Número de horas de trabajos de laboratorio	16
Número de horas de trabajo de campo	--
Número de horas de seminarios	--

Forma de evaluación:

Resolución de tres (3) Trabajos Prácticos:

- 1: Modelado de Lenguaje
- 2: Análisis de Sentimiento
- 3: Etiquetado de Secuencias

Lugar propuesto para el dictado (departamento, laboratorio, campo, etc.):

Laboratorio de computación

Puntaje propuesto para la carrera de doctorado:	2 puntos
---	----------

Número de alumnos:	Mínimo: 5	Máximo: 40
--------------------	-----------	------------

Audiencia a quien está dirigido el curso:
Estudiantes de doctorado en Cs. de la Computación.

Necesidades materiales del curso:
Laboratorio con proyector, pantalla y pizarrón Una computadora cada dos alumnos.

1-b-

Programa analítico del curso con Bibliografía (puede adjuntarse en hojas separadas):
--

- **Procesamiento básico de texto:** Expresiones regulares, tokenización, segmentación, normalización, lematización y stemming.
- **Modelado de lenguaje:** N-gramas, suavizado add-one y por interpolación, back-off. Evaluación con métricas de teoría la información (entropía y perplejidad). Aplicaciones: Generación de lenguaje y atribución de autoría.
- **Etiquetado de secuencias:** Etiquetado morfosintáctico (PoS tagging) y Reconocimiento de Entidades Nombradas (NER). Aprendizaje supervisado. Clasificadores: árboles de decisión, regresiones logísticas y SVMs. Modelos Ocultos de Markov (HMMs), de Máxima Entropía (MEMMs) y Conditional Random Fields (CRFs). Algoritmo de Viterbi y beam search. Ingeniería de features, evaluación y análisis de error.
- **Representación de palabras y modelos neuronales:** Representación vectorial de palabras (word embeddings): word2vec, fasttext y GloVe. Aprendizaje y evaluación. Representación de oraciones y documentos. Modelos de lenguaje neuronales: ULMFiT, OpenAI, ELMo, BERT.
- **Temas complementarios:** Análisis de sentimiento (sentiment analysis), análisis sintáctico (parsing), extracción de información (information extraction), traducción automática (machine translation), recuperación de información (information retrieval) y búsqueda de respuestas (question answering).

BIBLIOGRAFIA:

- Daniel Jurafsky and James H. Martin. Speech and Language Processing, 2nd Edition . Prentice Hall, 2nd edition, 2008.
- Christopher D. Manning and Hinrich Schtze. Foundations of statistical natural language processing. Hardcover, 1999.
- Daniel Jurafsky and James H. Martin. Speech and Language Processing, 3rd Edition Draft. <https://web.stanford.edu/~jurafsky/slp3/>. 2019.
- Jacob Eisenstein, Natural Language Processing, 1st edition, Draft. MIT Press. <https://www.cc.gatech.edu/~jeisenst/>, 2019.
- Yoav Goldberg, "A primer on neural network models for natural language processing." *Journal of Artificial Intelligence Research* 57 (2016): 345-420.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning, 2019, MIT Press.
- Proceedings de las principales conferencias y revistas del área: ACL, NAACL, EACL, TACL, EMNLP, COLING, etc.