

MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN (UNaM-UNNE)

Curso de Posgrado “Procesamiento del Lenguaje Natural”

Destinado a:

*Alumnos de la Maestría en Tecnologías de la Información.
Profesionales con título de Grado vinculados con actividades informáticas.*

Fecha y hora de Dictado de las Instancias presenciales:

- Viernes 10 de agosto 16:00 a 21:00 hs
- Sábado 11 de agosto de 8:00 a 13:00 hs.

Lugar de Dictado:

Aula Magna – Facultad de Ciencias Exactas, Químicas y Naturales. Félix de Azara 1552 – Posadas

Carga Horaria:

20 hs. (10 hs presenciales y las restantes a distancia)

Aranceles:

Gratuito Alumnos de la Maestría en Tecnologías de la Información.

\$1000 Externos a la Maestría en Tecnologías de la Información.

Inscripción:

<http://www.fceqyn.unam.edu.ar/mti/index.php/cursos-de-posgrado>

Consultas:

cursos.mti.unam@gmail.com

Docente a Cargo:

Dr. WAISSMAN VILANOVA Julio (*Universidad de Sonora – México*)

*Doctor en Sistemas (especialidad Sistemas automáticos) Institut National Polytechnique de Toulouse
(Francia) 2000.*

Recientemente a trabajado en estrecha colaboración en el modelado y control de campos de helióstatos de torre central con el Instituto de Energías Renovables de la UNAM (Universidad Nacional Autónoma de México), así como en la estimación de la precipitación durante el monzón de verano con el Departamento de Meteorología de la Universidad de Arizona. Actualmente su interés se centra en el modelado y análisis de conducta de estudiantes en sistemas tutoriales inteligentes. De 2002 a 2017 formó parte del Sistema Nacional de Investigadores.

Programa de Dictado

- **Unidad 1.** Introducción al procesamiento del lenguaje natural.
- **Unidad 2.** Normalización de documentos: Tokenización, expresiones regulares, palabras vacías, lematización.
- **Unidad 3.** Clasificación de documentos: Extracción de características por bolsa de palabras, TF-IDF y word2vect. Métodos de clasificación de documentos. Calidad de la clasificación de documentos.
- **Unidad 4.** Modelos de lenguaje: Modelo de lenguaje por N-gramas. Perplejidad. Modelos ocultos de Markov para identificación de entidades. POST (Part of speech tagging).
- **Unidad 5.** Modelado de tópicos: Representación densa de palabras en documentos. Mediciones de similitud entre documentos en base a contexto. Análisis probabilístico de semántica latente (PLSA). Asignación latente de Dirichlet (LDA)