

# Facultad de Ingeniería Comisión Académica de Posgrado

---

## Formulario de Aprobación Curso de Posgrado

### Asignatura: Análisis de Textos

---

**Profesor de la asignatura <sup>1</sup>:** Dra. Ing. Aiala Rosá, Grado 3, InCo

**Profesor Responsable Local <sup>1</sup>:**

**Otros docentes de la Facultad:** Dra. Ing. Dina Wonsever, Grado 5, InCo  
Msc. Juan José Prada, Grado 3, InCo

**Docentes fuera de Facultad:**

**Programa(s):** Diploma Especialización en Ciencia de datos (en trámite)  
Diploma Especialización en Sistemas de Información y Tecnologías de Gestión de Datos

**Instituto ó Unidad:** Instituto de Computación

**Departamento ó Area:** Grupo PLN

<sup>1</sup> Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

---

**Horas Presenciales:** 30 hs (clases teóricas y prácticas)

**Nº de Créditos:** 4

**Público objetivo y Cupos:**

Estudiantes de Posgrado de Informática y áreas afines.

Cupo: 30 estudiantes.

---

**Objetivos:**

El objetivo del curso es introducir los conceptos, los métodos y las aplicaciones principales del área Procesamiento de Lenguaje Natural, haciendo énfasis en problemas y enfoques actuales, basados principalmente en técnicas de Aprendizaje Automático. Se presentará un panorama completo de los desafíos que involucra el área y se experimentará con las herramientas y los recursos habituales, incluyendo alguna práctica orientada a problemas de dimensiones reales de gran desarrollo en la actualidad, como Análisis de Sentimientos y Extracción de Información en Redes Sociales.

---

**Conocimientos previos exigidos:** Programación

**Conocimientos previos recomendados:** Lenguajes Formales, Aprendizaje Automático, Probabilidad y Estadística, Álgebra.

---

**Metodología de enseñanza:**

(comprende una descripción de las horas dedicadas por el estudiante a la asignatura y su distribución en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

- Horas clase (teórico): 12
  - Horas clase (práctico): 3
  - Horas clase (laboratorio): 10
  - Horas consulta: 3
  - Horas evaluación: 2
    - Subtotal horas presenciales: 30
  - Horas estudio: 10
-

# Facultad de Ingeniería

## Comisión Académica de Posgrado

---

- Horas resolución ejercicios/prácticos: -
  - Horas proyecto final/monografía: 20
    - Total de horas de dedicación del estudiante: 60
- 

### Forma de evaluación:

- Informe sobre el proyecto final (grupal).
  - Prueba escrita individual presencial.
- 

### Temario:

- Introducción al Procesamiento de Lenguaje Natural
    - Conceptos generales, niveles clásicos de análisis.
    - Aplicaciones más comunes del área: traducción automática, resumen automático, extracción de información, análisis de textos de prensa y de redes sociales.
    - Herramientas y recursos usuales: analizadores lingüísticos, corpus.
    - Evaluación.
  - Conceptos generales de gramática del español:
    - Definiciones básicas: oración y sintagma, categorías gramaticales (léxicas y sintagmáticas), funciones sintácticas y semánticas.
    - Análisis gramatical básico de oraciones.
  - Tratamiento previo y procesamiento usual de los textos:
    - Obtención de textos.
    - Pre-procesamiento: limpieza, tokenización, lematización, stemming, pos-tagging.
    - Análisis sintáctico y semántico: recursos disponibles, aplicaciones y problemas.
    - Recursos léxicos.
    - Corpus anotados.
  - Métodos clásicos y actuales para el PLN:
    - Métodos simbólicos (reglas).
    - Métodos de aprendizaje automático basados en atributos.
    - Modelos de datos de desarrollo reciente: representaciones distribuidas de palabras y aprendizaje profundo.
  - Problemas actuales del área:
    - Análisis de sentimientos.
    - Extracción de información.
- 

### Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, 2nd edition. Daniel Jurafsky, and James H. Martin. Prentice-Hall. ISBN-10: 0131873210. 2009.

Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, 3rd edition draft. Daniel Jurafsky, and James H. Martin.  
<https://web.stanford.edu/~jurafsky/slp3/>. 2017.

### Bibliografía complementaria

Gramática del español para maestros y profesores del Uruguay. Ángela Di Tullio y Marisa Malcuori. ANEP, Prolee. ISBN 978-9974-688-79-7. 2012.

Machine Learning. Tom Mitchell. McGraw-Hill. ISBN 0070428077. 1997.

Learning scikit-learn: Machine Learning in Python. Raúl Garreta y Guillermo Moncecchi. Packt Publishing. ISBN-10: 1783281936. 2013.

---

**Facultad de Ingeniería  
Comisión Académica de Posgrado**

---

**Datos del curso**

---

**Fecha de inicio y finalización:** 28/10/2019 al 18/11/2019.

**Horario y Salón:** lunes, martes y jueves, 18 a 21 horas, salón a definir.

---