

**Formulario de aprobación de curso de posgrado/educación permanente**

**Asignatura:**

Introducción al Procesamiento de Lenguaje Natural

**Modalidad:**

Posgrado

Educación permanente

**Profesor de la asignatura <sup>1</sup>:**

Dra. Ing. Aiala Rosá, gr. 4, Instituto de Computación

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

**Otros docentes de la Facultad:**

Msc. Ing. Juan José Prada, gr. 3, Instituto de Computación

Dr. Ing. Luis Chiruzzo, gr. 3, Instituto de Computación

**Docentes fuera de Facultad:**

<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)

**Programa(s) de posgrado:**

Maestría en Informática (Peduciba), Doctorado en Informática (Peduciba), Maestría en Ciencia de Datos y Aprendizaje Automático.

**Instituto o unidad:**

Instituto de Computación

**Departamento o área:**

Procesamiento de Lenguaje Natural

**Horas Presenciales:**

60

**Nº de Créditos:**

12

(de acuerdo a la definición de la UdelAR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

**Público objetivo:**

Ingenieros en Computación o formación similar.

**Cupos:**

No hay cupo.

**Objetivos:**

Presentar los principales problemas del área PLN. Dar un panorama del estado de las aplicaciones existentes en la

actualidad. Desarrollar los enfoques y técnicas comúnmente utilizados en el área y estudiar las limitaciones teóricas y materiales de los diferentes enfoques propuestos. Presentar algunas de las herramientas y los recursos más utilizados, aplicarlos en la resolución de problemas prácticos.

---

**Conocimientos previos exigidos:**

- Lógica de predicados.
- Probabilidad y estadística.
- Lenguajes formales.
- Programación.

---

**Conocimientos previos recomendados:**

---

**Metodología de enseñanza:**

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

**Descripción de la metodología:**

1. Clases

Los temas serán expuestos en clases teórico/prácticas a ser dictadas por los docentes.

2. Presentación y cuestionamiento de artículos

Los estudiantes deberán realizar una presentación de un artículo y preguntas sobre otro artículo, presentado por otros estudiantes.

3. Laboratorio

Los estudiantes deberán realizar un laboratorio a lo largo del curso. Los docentes del curso pondrán a disposición de los estudiantes las herramientas y recursos.

4. Prueba final

Se realizará una prueba escrita final individual.

**Detalle de horas:**

Se estima un total de aproximadamente 180 horas de trabajo del estudiante, desglosadas de la siguiente forma:

- Horas clase (teórico/práctico): 56
- Horas evaluación individual: 3
- Horas de presentación y evaluación de artículos: 1
  - Subtotal horas presenciales: 60
- Horas estudio y preparación de artículo: 20
- Horas resolución ejercicios/prácticos: 30
- Horas proyecto final/monografía: 70
  - Total de horas de dedicación del estudiante: 180

---

#### Forma de evaluación:

El curso se evaluará a partir de:

- la presentación y cuestionamiento de artículo: 25%
- el laboratorio: 35%
- la prueba individual escrita: 40%

Todas las instancias son obligatorias y se requerirá como mínimo el 60% del puntaje de la prueba individual escrita. El curso se aprueba obteniendo más del 60% de los puntos totales.

---

#### Temario:

- Introducción al área. Motivación. Aplicaciones. Enfoques simbólicos, estadísticos, híbridos.
- Nociones básicas de gramática del español.
- Análisis léxico. Tokenización. Segmentación en oraciones. Morfología. POS tagging. Entidades con nombre.
- Análisis sintáctico. Formalismos y métodos.
- Análisis semántico. Representación de conocimiento del mundo. Representaciones distribuidas de palabras.
- Métodos y aplicaciones: Modelos de lenguaje, Recuperación y Extracción de Información, Traducción Automática, Análisis de Sentimiento, entre otros.

---

#### Bibliografía:

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

- Jurafsky, Daniel; Martin. James H. (2009) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd Edition)*, Prentice Hall, Upper Saddle River, N.J. ISBN: 0131873210.
- Jurafsky, Daniel; Martin. James H. (2023) *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, 3rd edition draft*. <https://web.stanford.edu/~jurafsky/slp3/>.
- Bird, Steven; Klein, Ewan; Loper, Edward (2009) *Natural Language Processing with Python*. O'Reilly Media. ISBN:0596516495.
- Di Tullio, Ángela, Malcouri Marisa (2013) *Gramática del español para maestros y profesores del Uruguay*. Montevideo. Codicen.

#### Complementaria:

- Hopcroft, John E.; Ullman, Motwani, Rajeev; Ullman, Jeffrey D. (2001). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley ISBN 13- 9780201441246.
- Manning, Christopher; Schütze, Hinrich (1999) *Foundations of Statistical Natural Language Processing*, Mit Press Cambridge. ISBN: 9780262133609.

Durante el curso se indicarán artículos científicos a ser presentados.



## Facultad de Ingeniería Comisión Académica de Posgrado

---

### Datos del curso

---

**Fecha de inicio y finalización:**

Marzo a Junio de 2023

**Horario y Salón:**

A confirmar.

**Arancel:**

[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

**Arancel para estudiantes inscriptos en la modalidad posgrado:** No corresponde

**Arancel para estudiantes inscriptos en la modalidad educación permanente:** No corresponde

---