
Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: Análisis de Textos

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado

Educación permanente

Profesor de la asignatura ¹: Dra. Ing. Aiala Rosá, grado 3, InCo

(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹:

(título, nombre, grado, instituto)

Otros docentes de la Facultad: MSc. Ing. Juan José Prada, Grado 3, InCo

(título, nombre, grado, instituto)

Docentes fuera de Facultad:

(título, nombre, cargo, institución, país)

¹ 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)

[Si es curso de posgrado]

Programa(s) de posgrado: Especialización en Ciencia de Datos, Maestría en Ciencia de Datos Aplicada

Instituto o unidad: Instituto de Computación

Departamento o área: Grupo Procesamiento de Lenguaje Natural

Horas Presenciales: 30 hs

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 4

[Exclusivamente para curso de posgrado]

(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: Estudiantes de Posgrado de Informática y áreas afines.

Cupos: sin cupo

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

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. Se presentará un panorama completo de los desafíos que involucra el área y se experimentará con las herramientas y los recursos habituales. Se incluirá alguna práctica orientada a problemas actuales del área como Análisis de Sentimientos y Extracción de Información en Redes Sociales, aplicando principalmente técnicas de Aprendizaje Automático.

Conocimientos previos exigidos: Programación

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

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:

El curso consistirá en sesiones teóricas y sesiones en modalidad de laboratorio práctico donde los estudiantes aplicarán los conceptos teóricos vistos y probarán las diferentes herramientas.

Detalle de horas:

- Horas clase (teórico): 10
- Horas clase (práctico): 2
- Horas clase (laboratorio): 12
- Horas consulta: 4
- Horas evaluación: 2
 - Subtotal horas presenciales: 30
- Horas estudio: 10
- Horas resolución ejercicios/prácticos: 6
- Horas proyecto final/monografía: 14
 - Total de horas de dedicación del estudiante: 60

Forma de evaluación:

[Indique la forma de evaluación para estudiantes de posgrado, si corresponde]

[Indique la forma de evaluación para estudiantes de educación permanente, si corresponde]

Para todos los estudiantes:

- Informe del proyecto final (grupal).
- Prueba escrita individual.

Temario:

- Introducción al Procesamiento de Lenguaje Natural: conceptos generales, niveles clásicos de análisis, aplicaciones más comunes del área, herramientas y recursos usuales.
- Conceptos generales de gramática del español: oración y sintagma, categorías gramaticales, funciones sintácticas y semánticas, análisis gramatical básico de oraciones.
- Tratamiento previo y procesamiento usual de los textos: obtención y pre-procesamiento de textos (limpieza, tokenización, lematización, stemming), análisis morfológico, análisis sintáctico.

- Métodos clásicos y actuales para el PLN: métodos simbólicos (reglas), métodos de aprendizaje automático basados en atributos, representaciones vectoriales de palabras, redes neuronales.
- Problemas actuales del área: por ejemplo, Análisis de Sentimiento, Extracción de Información, Búsqueda de Respuestas.

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/>. 2020.

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.
