
Formulario de Aprobación Curso de Posgrado

Asignatura: Introducción al Procesamiento de Lenguaje Natural

Profesor de la asignatura ¹:

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

Profesor Responsable Local ¹:

Otros docentes de la Facultad:

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

Msc. Ing. Mathias Etcheverry, gr. 2, Instituto de Computación

Docentes fuera de Facultad:

Programa(s):

Maestría en Informática del Pedeciba

Doctorado en Informática del Pedeciba

Instituto ó Unidad: Instituto de Computación

Departamento ó Area: Departamento de Programación

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)

Horas Presenciales: 60

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

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 la enseñanza)

Público objetivo y Cupos: Ingenieros en Computación o con formación similar. No hay 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 en el Depto. de Posgrado, hasta completar el cupo asignado)

Objetivos:

Presentar los principales problemas del área de PLN: traducción automática, resumen automático, reconocimiento del habla, recuperación de información, etc. Dar un panorama del estado de las aplicaciones existentes hoy día.

Desarrollar los enfoques y técnicas comúnmente utilizados en el área: análisis morfológico, análisis sintáctico de superficie, análisis semántico, análisis pragmático y modelos del discurso, métodos de estado finito, aprendizaje automático, etc. Estudiar las limitaciones teóricas y materiales de los diferentes enfoques propuestos.

Presentar algunas de las herramientas y recursos más utilizados: taggers, diccionarios, ontologías, gramáticas, corpus paralelos, etc. Utilizar estas herramientas en la resolución de un problema práctico.

Conocimientos previos exigidos:

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

Conocimientos previos recomendados:

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)

La metodología es la siguiente:

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

2. Presentación y cuestionamiento de artículos

Los estudiantes – en forma individual - deberán realizar presentaciones de artículos seleccionados. Los docentes del curso ofrecerán un conjunto inicial de artículos y los estudiantes (en grupos de dos personas) deberán elegir un artículo a presentar y un artículo (que será presentado por otro grupo) a cuestionar. La dinámica será la siguiente: el grupo que presenta el artículo debe, además, entregarle al docente, el día de la presentación, un resumen de una página conteniendo los conceptos clave del artículo. El grupo que cuestiona el artículo, deberá entregarle al docente, el día de la presentación, una lista de al menos tres preguntas a hacer al grupo que presenta el artículo. Las preguntas las formularán los estudiantes de este grupo y el docente será el moderador de la discusión (en la que puede participar el resto de la clase), y evaluará la pertinencia e interés de las preguntas.

2. Laboratorio

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

3. Prueba final

Se prevé una prueba también individual escrita final, con el objetivo de evaluar los conocimientos del curso.

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

- Horas clase (teórico/práctico): 54 / Dos clases teóricas semanales de 2 horas
- Horas clase (laboratorio): 2 / Dos presentaciones de 1 hora
- Horas evaluación individual: 3
- Horas de presentación y evaluación de artículos: 1
 - o Subtotal horas presenciales: 60
- Horas estudio y preparación de artículo: 35
- Horas resolución ejercicios/prácticos: 15
- Horas proyecto final/monografía: 70
 - o Total de horas de dedicación del estudiante: 180

Forma de evaluación:

El curso se evaluará a partir de:

- la presentación de artículo: 10%
 - el cuestionamiento de artículo: 10%
 - el laboratorio: 20%
 - la prueba individual escrita: 60%
-

Temario:

- Introducción al área. Motivación. Aplicaciones más comunes. Enfoques simbólicos, estadísticos, híbridos.
- Nociones básicas de gramática del español.
- Análisis léxico. Tokenización. Segmentación en oraciones. Entidades con nombre. Morfología.
- Realización de POS tagging usando métodos estadísticos. Hidden Markov Models
- Análisis sintáctico. Gramáticas. Métodos de parsing.
- Análisis semántico. Ambigüedad. Representación de conocimiento del mundo.
- Recuperación de Información, Clasificación
- Traducción Automática
- Aplicaciones

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.
- Manning, Christopher; Schütze, Hinrich (1999) *Foundations of Statistical Natural Language Processing*, MIT Press Cambridge. ISBN: 9780262133609.
- Bird, Steven; Klein, Ewan; Loper, Edward (2009) *Natural Language Processing with Python*. O'Reilly Media. ISBN:0596516495
- Alarcos Llorach, Emilio. (1994) *Gramática de la lengua española*. Real Academia Española. Madrid. Espasa Calpe.

Complementaria:

- Dale, Robert; Moisl, Hermann; Somers, Harold (2000) *Handbook of natural Language Processing*, ed., Marcel Dekker, New York. ISBN: 0824790006.
 - Allen, James. (1995) *Natural Language Understanding* (Benjamin/Cummings, Menlo Park, CA, 1987) Publishing Company. ISBN: 0805303340
 - Di Tullio, Ángela, Malcouri Marisa (2013) *Gramática del español Para maestros y profesores del Uruguay* Montevideo Codicen.
 - Hopcroft, John E.; Ullman, Motwani, Rajeev; Ullman, Jeffrey D. (2001). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley ISBN 13- 9780201441246
- Durante el curso se indicarán artículos científicos a ser presentados.

Datos del curso

Fecha de inicio y finalización: Agosto a Noviembre de 2022

Horario y Salón: Miércoles y Viernes de 9:00 a 11:00 (salones a confirmar).
