



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Programa de Introducción al Procesamiento de Lenguaje Natural

1. NOMBRE DE LA UNIDAD CURRICULAR

Introducción al Procesamiento de Lenguaje Natural

2. CRÉDITOS

12 créditos

3. OBJETIVOS DE LA UNIDAD CURRICULAR

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

4. METODOLOGÍA DE ENSEÑANZA

Los temas serán expuestos en clases teórico/prácticas a ser dictadas por los profesores. También se incluye la presentación y crítica de artículos científicos por parte de los estudiantes y un trabajo de laboratorio en grupo en el cual aplicarán los conocimientos adquiridos a lo largo del curso.

5. TEMARIO

1. Introducción al área. Motivación. Aplicaciones más comunes. Enfoques simbólicos, estadísticos, híbridos.
2. Nociones básicas de gramática del español.
3. Análisis léxico. Expresiones Regulares. Tokenización. Distancia Mínima Edición. Morfología y POS-tagging usando métodos estadísticos. Modelos de lenguaje.
4. Análisis sintáctico. Gramáticas Libres de Contexto. Otros formalismos. Métodos de parsing.
5. Análisis semántico. Semántica Formal. Semántica Léxica. Roles Semánticos.

6. Aplicaciones:

- Clasificación
- Recuperación de Información
- Extracción de Información
- Traducción Automática
- Word embeddings

6. BIBLIOGRAFÍA

Tema	Básica	Complementaria
Introducción al área	(1,2)	(4,5)
Nociones básicas de gramática del español	(4)	(7)
Análisis léxico	(1,2)	(3,5,8,9)
Análisis sintáctico	(1)	(8)
Análisis semántico	(1,2)	(5,6)

6.1 Básica

1. Jurafsky, Daniel; Martin. James H. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, Upper Saddle River, N.J. ISBN: 0130950696.
 2. Manning, Christopher; Schütze, Hinrich (1999) *Foundations of Statistical Natural Language Processing*, Mit Press Cambridge. ISBN: 9780262133609.
 3. Bird, Steven; Klein, Ewan; Loper, Edward (2009) *Natural Language Processing with Python*. O'Reilly Media. ISBN:0596516495
 4. Alarcos Llorach, Emilio. (1994) *Gramática de la lengua española*. Real Academia Española. Madrid. Espasa Calpe.
-

6.2 Complementaria

5. Dale, Robert; Moisl, Hermann; Somers, Harold (2000) *Handbook of natural Language Processing*, ed., Marcel Dekker, New York. ISBN: 0824790006.
 6. Allen, James. (1995) *Natural Language Understanding* (Benjamin/Cummings, Menlo Park, CA, 1987) Publishing Company. ISBN: 0805303340
 7. Di Tullio, Ángela, Malcouri Marisa (2013) *Gramática del español Para maestros y profesores del Uruguay* Montevideo Codicen.
 8. Hopcroft, John E.; Ullman, Motwani, Rajeev; Ullman, Jeffrey D. (2001). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley ISBN 13-9780201441246
 9. Beesley, Kenneth R.; Karttunen, Lauri; (2003) *Finite State Morphology* CSLI Publications, ISBN (Paperback): 1575864347.
-

7. CONOCIMIENTOS PREVIOS EXIGIDOS Y RECOMENDADOS

7.1 Conocimientos Previos Exigidos:

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

7.2 Conocimientos Previos Recomendados:

No tiene.

ANEXO A

A1) INSTITUTO

Instituto de Computación

A2) CRONOGRAMA TENTATIVO

Semana 1	Introducción al área. Motivación. Aplicaciones más comunes. Enfoques simbólicos, estadísticos, híbridos (2hs de clase). Nociones básicas de gramática del español (2hs de clase).
Semana 2	Nociones básicas de gramática del español (cont.) (2hs de clase). Expresiones Regulares. Tokenización. Distancia de Mínima Edición (2hs de clase).
Semana 3	Análisis léxico. Morfología y POS-tagging (4 hs de clase).
Semana 4	Aplicaciones: Clasificación (4hs de clase).
Semana 5	Modelos de lenguaje (4hs de clase)
Semana 6	Análisis sintáctico: Gramáticas Libres de Contexto y otros formalismos (4hs de clase)
Semana 7	Semana de parciales
Semana 8	Análisis sintáctico: parsing (4hs de clase)
Semana 9	Análisis semántico: Semántica Formal (4hs de clase)
Semana 10	Análisis semántico: Semántica Léxica (2hs de clase) Roles Semánticos (2hs de clase)
Semana 11	Aplicaciones: Recuperación y Extracción de Información (4hs de clase)
Semana 12	Aplicaciones: Word embeddings (2hs de clase)
Semana 13	Aplicaciones: Traducción automática (4hs de clase)
Semana 14	Presentación de artículos (estudiantes) (4hs de clase)
Semana 15	Presentación de artículos (estudiantes) (4hs de clase)

A3) MODALIDAD DEL CURSO Y PROCEDIMIENTO DE EVALUACIÓN

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

- Horas clase: 56 - Dos clases teórico/práctico semanales de 2 horas cada una
- Horas evaluación: 24 - incluye 20 hs de preparación
- Horas de presentación de artículos: 10 - incluye 8 hs preparación del artículo
- Horas del trabajo de Laboratorio (en grupo): 90

Detalle de las modalidades:

1. Presentación y cuestionamiento de artículos

Los estudiantes deberán realizar presentaciones de artículos seleccionados. Los docentes del curso ofrecerán un conjunto inicial de artículos y los estudiantes (en grupos de tres o cuatro personas dependiendo de la cantidad de estudiantes del curso) deberán elegir un artículo a presentar y se les asignará otro 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 cinco 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 (los grupos serán los mismos que los de los artículos). Los docentes del curso pondrán a disposición de los estudiantes las herramientas y recursos.

3. Prueba final

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

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.

A4) CALIDAD DE LIBRE

Esta unidad curricular no adhiere a la Calidad de Libre.

A5) CUPOS DE LA UNIDAD CURRICULAR

No tiene cupo

ANEXO B para la carrera Ingeniería en Computación (plan 97)

B1) ÁREA DE FORMACIÓN

Inteligencia Artificial y Robótica.

B2) UNIDADES CURRICULARES PREVIAS

Curso:

- Lógica (Examen)
- Programación 3 (Examen)
- Probabilidad y Estadística (Examen)
- Programación 4 (Examen)
- Teoría de Lenguajes (Examen)
- Taller de Programación

Examen:

No corresponde.

ANEXO B para la carrera Licenciatura en Computación

B1) ÁREA DE FORMACIÓN

Inteligencia Artificial y Robótica.

Esta unidad curricular pertenece al agregado Inteligencia Artificial del perfil Inteligencia Artificial.

B2) UNIDADES CURRICULARES PREVIAS

Curso:

- Lógica (Examen)
- Programación 3 (Examen)
- Probabilidad y Estadística (Examen)
- Programación 4 (Examen)
- Teoría de Lenguajes (Examen)
- Taller de Programación

Examen:

No corresponde.

ANEXO B para la carrera Ingeniería en Computación (plan 87)

B1) ÁREA DE FORMACIÓN

No corresponde.

B2) UNIDADES CURRICULARES PREVIAS

Curso:

- Previas comunes a electivas
- Lógica
- Programación III
- Teoría de la Programación I
- Taller III

Examen:

No corresponde.

Observación: esta unidad curricular se corresponde con una electiva.

APROB. RES. CONSEJO DE FAC. MIO.
RESOLUCIÓN 28/08/2018 EXP. 060120-001492-04