

## Minería de Textos / Inducción de Estructuras Sintácticas desde una Perspectiva Computacional

Programa:

En este breve curso presentaremos una perspectiva computacional a la inducción automática de estructuras sintácticas. Existen diversos paradigmas para la inducción de estructuras y cada uno intenta optimizar una métrica bien definida. En este curso mostraremos dos paradigmas distintos. En el primero, basado en un esquema supervisado, se intenta aprender estructuras a partir de estructuras ejemplos. En el segundo, se intenta descubrir estructuras a partir de oraciones de lenguaje natural que no tienen visible ninguna estructura. Para el primero mostraremos los algoritmos clásicos de inducción de gramáticas probabilísticas y para el segundo mostraremos como se pueden inducir gramáticas utilizando modelos de Machine Learning y modelos basados en compresión de datos.

El objetivo de la minería de texto consiste en encontrar, verificar o descubrir información que está presente en textos mediante técnicas de extracción automática. Típicamente se extrae información sobre entidades, propiedades de entidades, acontecimientos o relaciones entre entidades, de modo que la información extraída pueda ser tratada con métodos más convencionales, por ejemplo volcada en una base de datos, u ofrecida al usuario que la requiere en un formato más amigable.

Se verá el uso de clasificadores como herramienta fundamental para el procesamiento de textos tanto desde la perspectiva de Procesamiento del Lenguaje Natural, como de la Recuperación de Información. Se verán las técnicas para el minado de opinión o sentimiento y las técnicas para extracción de entidades, atributos y relaciones. Asimismo se hará énfasis en la evaluación de los resultados.

El objetivo del curso es que los alumnos adquieran las bases fundamentales y una primera aproximación a los problemas más complejos.

12.- BIBLIOGRAFÍA (indicar título del libro, autor, Editorial y año de publicación)(adjuntada)

- Introduction to Information Retrieval

Autores: Christopher D. Manning, Prabhakar Raghavan y Hinrich Schütze,  
Cambridge University Press. 2008

- Web Data Mining

Exploring Hyperlinks, Contents and Usage Data

Autor: Bing Liu,  
Springer, 2006

- The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data

Ronen Feldman y James Sanger  
Cambridge University Press, 2007.

- Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999

- Daniel Jurafsky, James H. Martin. Speech and Language Processing.  
Prentice Hall

- Carrascosa, R., Coste, F., Gallé, M., Infante-Lopez, G., "Searching  
for Smallest Grammars on Large Sequences and Application to DNA.",  
Journal of Discrete Algorithms, 2011



A handwritten signature in black ink, consisting of several loops and a long horizontal stroke at the end.

Dr. Alejandro N. Rios  
Departamento de Computación  
FCEyN UBA



Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 500.323/2011

Buenos Aires,

14 NOV 2011

**VISTO:**

la nota presentada por el Dr. Sebastián Uchitel, Director del Departamento de Computación, mediante la cual eleva la información y el programa del curso de posgrado **Minería de textos / Inducción de Estructuras sintácticas desde una perspectiva computacional**, dictado en el segundo cuatrimestre de 2010 dictado por el Dr. José Castaño,

el CV del Dr. José Castaño

**CONSIDERANDO:**

que las presentes actuaciones no fueron presentados en tiempo y forma por un error administrativo,

lo actuado por la Comisión de Doctorado de esta Facultad el 25/10/2011,

lo actuado por la Comisión de Enseñanza, Programas, Planes de estudio y Posgrado

lo actuado por este Cuerpo en la sesión realizada en el día de la fecha,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD DE  
CIENCIAS EXACTAS Y NATURALES  
RESUELVE:**

**Artículo 1°:** Dar validez al dictado del curso de posgrado **Minería de textos / Inducción de Estructuras sintácticas desde una perspectiva computacional** de 15 hs. de duración, en el segundo cuatrimestre de 2010.

**Artículo 2°:** Aprobar el programa del curso de posgrado **Minería de textos / Inducción de Estructuras sintácticas desde una perspectiva computacional**, obrante a fs 3 y 4 del expediente de referencia.

**Artículo 3°:** Aprobar un puntaje máximo para la Carrera de Doctorado de 0,5 (cero coma cinco) puntos

**Artículo 4°:** Aprobar un arancel de 20 Módulos. Disponer que los montos recaudados serán utilizados conforme a lo dispuesto por Resolución CD N° 072/03.

**Artículo 5°:** Comuníquese al Director del Departamento de Computación, a la Biblioteca de la FCEyN, a la Biblioteca de la FCEyN y a la Subsecretaría de Postgrado (con fotocopia del Programa incluido)

**Artículo 6°:** Comuníquese al Departamento de Alumnos (sin fotocopia del programa incluido). Cumplido Archívese.

Resolución CD N°  
SP/psa/25/10/2011

2870 ==

  
Dr. JAVIER LÓPEZ DE CASENAVE  
SECRETARIC ACADEMICO

  
Dr. JORGE ALIAGA  
DECANO