

**Formulario de Aprobación Curso de Posgrado 2010**

**Asignatura:**

Seminario: Métodos de aprendizaje automático en Procesamiento de Lenguaje Natural

**Profesor de la asignatura:** Dra. Dina Wonsever, G5, InCo

**Instituto ó Unidad:** InCo

**Departamento ó Area:** Departamento de Programación

**Fecha de inicio y finalización:** Lunes 25 de octubre de 2010 - Lunes 27 de diciembre de 2010

**Horario y Salón:** Lunes y Miércoles de 18 a 20 horas

**Horas Presenciales:** 50

**Nº de Créditos:** 6

**Público objetivo y Cupos:** Estudiantes de posgrado en Informática o áreas afines. Sin cupo.

**Objetivos:**

Profundizar en métodos de aprendizaje automático actualmente en uso para procesamiento de lenguaje natural. Experimentar con herramientas de software especializado disponibles. Adquirir solvencia en el análisis comparativo de diversos métodos y estrategias de evaluación de resultados.

**Conocimientos previos exigidos:**

**Conocimientos previos recomendados:** Nociones básicas de Procesamiento de Lenguaje Natural y de Aprendizaje Automático.

**Metodología de enseñanza:**

Se realizarán exposiciones a cargo de los participantes, basadas en el material establecido para el tema y en un trabajo de experimentación / demostración que el expositor realizó previamente. En cada exposición, además, los restantes participantes habrán leído el material correspondiente y formularán al menos tres preguntas al expositor. Se priorizará la discusión, la comparación de diversos métodos y el análisis de resultados y errores en la parte experimental.

Se estima un total de 90 horas de trabajo por parte del estudiante.

**Forma de evaluación:**

Se evaluará en base a la exposición y trabajo experimental previo realizado por cada participante. Se considerará también la participación en las discusiones en cada sesión.

**Temario:**

0. Una breve introducción/repaso en temas de probabilidad e inferencia.
1. Introducción al Aprendizaje Automático. Clasificador simple, árboles de decisión. Ejemplos en PLN.
2. Naive Bayes y clasificación.
3. Métodos de clustering.

4. Voted Perceptron, Support Vector machines.
5. Modelos Gráficos HMM, MEMM.
6. CRF.
7. Revisión de algunos sub-problemas en Procesamiento de Lenguaje Natural y su resolución mediante las técnicas vistas: etiquetado de categoría gramatical, desambiguación semántica, etc.
8. Parsing probabilista.
9. Traducción automática.
10. Otras aplicaciones: extracción de información, resumen automático, etc.

---

**Bibliografía básica:**

*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, Upper Saddle River, N.J., 2008. 2<sup>nd</sup> edition D. Jurafsky, James H. Martin. ISBN 0130950696.

*Foundations of Statistical Natural Language Processing*, Manning & Schütze, Mit Press Cambridge, 1999. ISBN: 9780262133609.

*Machine Learning* Thomas M. Mitchell, , McGraw-Hill Higher Education, 1997

---