

**UNIVERSIDAD DE BUENOS AIRES**  
**FACULTAD DE CIENCIAS EXACTAS Y NATURALES**

1. DEPARTAMENTO: Computación
2. CUATRIMESTRE: Segundo de 2004.
3. ASIGNATURA: **Aprendizaje Automático y Recuperación de Información**
4. CARRERA: Licenciatura en Ciencias de la Computación
5. CARACTER DE LA MATERIA: Optativa
6. NUMERO DE CODIGO DE CARRERA: 18
7. NUMERO DE CODIGO DE MATERIA:
8. PUNTAJE: 1 punto
9. PLAN DE ESTUDIOS AÑO: 1993
10. DURACION: 1 semana
11. HORAS DE CLASE SEMANAL:  
a) TEORICAS/PRACTICAS: 15 horas    b) LABORATORIO:                    c) PRACTICAS:
12. CARGA HORARIA TOTAL: 15 horas
13. ASIGNATURAS CORRELATIVAS: **Conocimientos en estadística o probabilidad**  
**Conocimientos básicos en informática**
14. FORMA DE EVALUACION: final
15. PROGRAMA Y BIBLIOGRAFIA:

Profesor  
**Dr. Hugo Zaragoza**

  
Dr. Enrique Carlos Segura  
Director  
Depto. de Computación  
F. C. E. y N - UBA

## 15) PROGRAMA:

**Introducción:** La Recuperación de Información intenta resolver el problema de "búsqueda en textos". Las mejores soluciones del problema de Recuperación de Información surgieron de la estadística hace más de veinte años; se ocupaban de situaciones con pocos textos y, estos, completamente inestructurados.

Este viejo problema está ganando una atención creciente en los últimos años con el tremendo aumento de la información textual en línea producida por la WWW. Mientras que los métodos tradicionales de IR continúan formando el núcleo de la mayor parte de la tecnología de búsqueda actual, están surgiendo muchos nuevos desafíos. En los últimos 10 años el Aprendizaje Automático, y en particular los campos de la inferencia probabilística y el aprendizaje estadístico se han aplicado con éxito a la recuperación de información.

**Objetivos:** En este curso se dará una introducción al Aprendizaje Automático, y en particular a algunas de las técnicas más exitosas de inferencia probabilística y aprendizaje estadístico. A lo largo de la introducción se utilizarán ejemplos de la investigación en Recuperación de Información para ilustrar las técnicas utilizadas actualmente y los desafíos que permanecen abiertos. Finalmente, se introducirán algunos de los problemas más desafiantes para la Recuperación de Información futura y se analizará cómo pueden resolverse utilizando técnicas de Aprendizaje Automático.

**Programa:** Inferencia probabilística y aplicaciones a la búsqueda textual

- Modelos probabilísticos & inferencia
- El principio de ranking probabilístico y el Modelo de Relevancia
- Modelos de Recuperación de Información: el modelo probabilístico, el modelo lingüístico
- Extensiones: modelos semánticos latentes, recuperación estructurada.

**Métodos Kernel y aplicaciones a la Categorización de Textos.**

- Teoría de aprendizaje estadístico
- Métodos Kernel, Algoritmo Perceptrón y Máquinas de Soporte Vectorial (VSM)
- Kernels para datos textuales: kernel polinómico & de secuencias, aprendizaje de kernels
- Problemas y algunas soluciones: optimización, dimensionalidad, jerarquía de clases, desbalanceo de clases.

**Direcciones futuras:**

- Metadatos, XML y Recuperación Estructurada
- Aprendizaje en línea a partir de realimentación de la búsqueda
- Búsqueda personalizada

## 16) BIBLIOGRAFIA:

Apunte de cátedra. No fue adjuntada otra bibliografía por parte del docente a cargo.

  
Dr. Enrique Carlos Segura  
Director  
Depto. de Computación  
F. C. E. y N - UBA