
Formulario de Aprobación Curso de Posgrado 2015

Asignatura: Recuperación de Información y Recomendaciones en la Web (webir)

(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura ¹: Dra. Libertad Tansini, gr. 3, InCo
(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local ¹:
(título, nombre, grado, Instituto)

Otros docentes de la Facultad:
(título, nombre, grado, Instituto)

Docentes fuera de Facultad:
(título, nombre, cargo, Institución, país)

Instituto o Unidad: InCo
Departamento o Area: Investigación Operativa

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)

Fecha de inicio y finalización: segundo semestre 2015
Horario y Salón: A determinar

Horas Presenciales: 76
(se deberán discriminar las mismas en el ítem Metodología de enseñanza)

Nº de Créditos: 9
(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: Estudiantes avanzados de la carrera de Ingeniería en Computación y estudiantes de Maestría y Doctorado de PEDECIBA.
Cupo de 30 estudiantes.

(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:

El objetivo principal de este curso es enseñar, en una primer lugar, los fundamentos básicos de la recuperación de la información, en particular para Internet y luego profundizar en temas avanzados.

La primera parte del curso presenta los temas principales relacionados con la recuperación de la información, en particular relacionados con el funcionamiento de los motores de búsqueda de Internet, hasta el tema 7(ver Temario).

En la segunda parte del curso se discuten temas avanzados, como por ejemplo la optimización de la recuperación de información mediante "relevance feedback" o expansión de consultas.

Finalmente se tratan los temas de filtrado colaborativo y el análisis de enlaces, para estudiar como el comportamiento colectivo ayuda en la recuperación de la información en Internet.

Conocimientos previos exigidos:

Bases de Datos

Análisis de algoritmos

Algoritmos y estructura de datos

Conocimientos previos recomendados:

Investigación Operativa

Conocimientos básicos de Probabilidad y Estadística

Metodología de enseñanza:

Son 2 clases teórico-prácticas semanales y los estudiantes deben realizar un proyecto en la temática del curso. El seguimiento de los proyectos de hará mediante monitoreos semanales de 1-2 horas. Para la evaluación del curso, los estudiantes deberán escribir un informe del trabajo realizado y hacer una presentación.

- Horas clase (teórico): 52
 - Horas clase (práctico):
 - Horas clase (laboratorio): 20
 - Horas consulta:
 - Horas evaluación: 4
 - Subtotal horas presenciales: 76
 - Horas estudio: 20
 - Horas resolución ejercicios/prácticos:
 - Horas proyecto final/monografía: 40
 - Total de horas de dedicación del estudiante: 136
-

Forma de evaluación:

- Ejecución de un proyecto en la temática del curso.
 - Informe escrito del proyecto.
 - Presentación oral del proyecto.
-

Temario:

1. Recuperación booleana.
 2. Vocabularios y listas de ocurrencias de palabras.
 3. Diccionarios y recuperación tolerante.
 4. Construcción y compresión de índices.
 5. Puntaje y asignación de pesos en el modelo de espacio vectorial.
 6. Calculo de puntajes en un sistema completo de búsqueda.
 7. Evaluación de la recuperación de la información.
 8. Relevance feedback y expansión de consultas.
 9. Recuperación de la información probabilística.
 10. Temas avanzados: Modelos de lenguajes para la recuperación de la información, Clasificación de texto con la regla de Bayes y mediante el modelo de espacio vectorial, SVM y aprendizaje automático para documentos, Clustering y clustering jerárquico.
 11. Búsquedas web, web crawling e índices.
 12. Filtrado colaborativo.
 13. Análisis de enlaces.
-

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <http://nlp.stanford.edu/IR-book/>
