

Nombre Recuperación de Información y Recomendaciones en la Web (WIR)

Créditos 10

Objetivos de la Asignatura

La recuperación de la información no comenzó con Internet, pero de un tiempo a esta parte, el motor de innovación en el área ha sido justamente la Web, debido principalmente a explosión de publicaciones y a la enorme cantidad de creadores de información que hay en ella. Toda esta información no podría seguir creciendo si no se pudiera encontrar, indexar y analizar para que cada usuario pueda obtener de manera rápida la información que es pertinente a sus necesidades.

El objetivo principal de este curso es enseñarles a los estudiantes en una primera etapa 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.

Metodología de enseñanza

140 hs (60 hs. teórico, 60 hs. realización de proyectos, 20hs. estudio individual).

Curso es de carácter teórico y de tipo proyecto-presentación. El estudiante tendrá 4 horas semanales de clases (durante 15 semanas), y horario de consultas para el apoyo a la realización de un proyecto. Se estima unas 2 horas semanales de estudio adicional y 4 horas semanales para la realización del proyecto.

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 comprensión 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. Modelos de lenguajes para la recuperación de la información
11. Clasificación de texto con la regla de Bayes y mediante el modelo de espacio vectorial
12. Support vector machines y aprendizaje automático para documentos
13. Clustering y clustering jerárquico
14. Descomposición de matrices y LSI (latent semantic indexing)
15. Búsquedas web, web crawling e índices
16. Filtrado colaborativo
17. Análisis de enlaces

Bibliografía

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

Conocimientos previos recomendados

Bases de Datos
Análisis de algoritmos
Algoritmos y estructura de datos
Investigación Operativa
Conocimientos básicos de Probabilidad y Estadística

Anexo: Ingeniería en Computación

1) Cronograma tentativo

Las clases teóricas serán desde el 1/8/2010 a 16/11/2010.

En setiembre los estudiantes deberán seleccionar un proyecto, que previamente deberá ser validado por el docente. Trabajarán en el proyecto durante los meses de setiembre y octubre, las dos últimas semanas de octubre deberán escribir el informe del trabajo realizado.

Desde el 16/11/2010 hasta el 26/11/2010 se harán las entregas de los proyectos y las presentaciones de los mismos.

2) Modalidad del curso y procedimiento de evaluación

1. Teórico

Para abordar los temas propuestos se realizarán dos clases teóricas semanales de agosto a noviembre.

2. Proyecto

Los estudiantes deberán realizar un proyecto que puede constar esencialmente de tres actividades: investigación sobre un tema relacionado con el curso, diseño de una solución a un problema (usando lo aprendido) y una posible implementación de la solución. El tema específico deberá ser validado por el docente previamente. El estudiante deberá escribir un informe sobre el trabajo realizado y hacer una presentación oral del trabajo (max. 15 minutos).

3. Evaluación

El curso se evaluará a partir de:

- La asistencia a clases (30%)
- El informe sobre el proyecto realizado (40%)
- La presentación oral del proyecto (30%)

El curso se aprueba con el 60% de los puntos, teniendo un mínimo del 60% en cada actividad.

3) Materia

Bases de Datos y Sistemas de Información.

4) Previaturas

- Programación 3
- Fundamentos de base de datos

Para el Plan 87:

- Bases de datos
- Programación 3

5) Cupo

Cupo máximo de 30 estudiantes, se seleccionara 50% de los estudiantes por avance en la carrera y 50% por sorteo. Se conformará una lista de 10 estudiantes suplentes de acuerdo a los mismos criterios.

La necesidad de tener cupo se fundamenta en la cantidad de docentes (uno) y a la modalidad del curso, que incluye el seguimiento de grupos y sus tareas.

6) Esta asignatura no adhiere a resolución del consejo sobre condición de libre

APROB. RES. CONSEJO DE FAC. ING.

Fecha 29.7.10 Exp. 060120-001536-10