

PROGRAMA DE CURSO

Código	Nombre			
CC5203	Recuperación de la Información			
Nombre en Inglés				
INFORMATION RETRIEVAL				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3	0	
Requisitos			Carácter del Curso	
Resultados de Aprendizaje				
Al finalizar el curso el alumno estará capacitado para:				
1. Conocer conceptos y modelos fundamentales de recuperación de información.				
2. Conocer estructuras de datos usadas con mayor frecuencia en la implementación de sistemas de recuperación de información.				
3. Conocer conceptos y modelos fundamentales para búsqueda e indexamiento de documentos en la web.				

Metodología Docente	Evaluación General
El Curso se desarrollará con clases expositivas de 90 minutos de duración.	El alumno deberá rendir tres controles y desarrollar un proyecto. Ponderaciones: Controles 75 %, Proyecto 25 %

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Propiedades y operaciones en texto	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Leyes Zipf - Heaps, preprocesamiento de texto, diccionarios, stemming, tesauros.	Conocer conceptos fundamentales en preprocesamiento de texto	[1] Caps. 6-7 [2] Caps. 2-3

Número	Nombre de la Unidad	Duración en Semanas
2	Modelos de recuperación de información	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Modelo booleano, Tf-Idf, LSI, probabilístico, BM25, language model.	Conocer los modelos más usados en recuperación de información	[1] Cap. 2 [2] Caps. 1,6,7,11,12,18

Número	Nombre de la Unidad	Duración en Semanas
3	Evaluación en recuperación de información	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Métricas y metodologías: precisión, recall, F-measure, colecciones de referencia.	Conocer metodologías de evaluación de desempeño de sistemas de recuperación de información	[1] Cap. 3 [2] Cap. 8

Número	Nombre de la Unidad	Duración en Semanas
4	Estructuras de datos para texto	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Índices invertidos, árboles de sufijos, arreglos de sufijos, estructuras compactas.	Conocer las estructuras de datos para texto más ampliamente usadas en recuperación de información	[1] Cap. 8 [2] Cap. 4-5

Número	Nombre de la Unidad	Duración en Semanas
5	Arquitecturas de sistemas de recuperación de información	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Procesamiento de consultas centralizado v/s distribuido, paralelismo en sistemas de recuperación de información, estrategias de caching.	Conocer las arquitecturas más ampliamente usadas en la implementación de sistemas de recuperación de información	[1] Cap. 9 [2] Cap. 20

Número	Nombre de la Unidad	Duración en Semanas
6	Recuperación de información en la Web	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Tecnologías de búsqueda en la Web, crawlers, ranking en la Web, análisis de links, HITS, PageRank, query mining.	Conocer las principales tecnologías de búsqueda que se utilizan en motores de búsqueda Web.	[1] Cap.13 [2] Caps.19-21

Bibliografía
[1] Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, 1999. [2] Christopher Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

Vigencia desde:	27 de Julio 2009
Elaborado por:	Marcelo Mendoza R.