



**MAESTRIA EN CIENCIAS DE LA COMPUTACION**

**Área: Base de Datos y Recuperación de Información**

**Programa de Asignatura: Recuperación de Información**

**Código: MCOM 21200**

**Tipo: Obligatoria**

**Créditos: 9**

**Fecha: Noviembre 2012**



## 1. DATOS GENERALES

Nombre del Programa Educativo:	Maestría en Ciencias de la Computación
Modalidad Académica:	Escolarizada
Nombre de la Asignatura:	Recuperación de Información (Obligatoria)
Ubicación:	Segundo semestre

## 2. REVISIONES Y ACTUALIZACIONES

Autores:	Dr. David Eduardo Pinto Avendaño
Fecha de diseño:	Noviembre 2012
Fecha de la última actualización:	Noviembre 2012
Revisores:	No aplica, Materia nueva
Sinopsis de la revisión y/o actualización:	No aplica, Materia nueva



### **3. OBJETIVOS GENERALES:**

#### ***Específicos:***

El estudiante conocerá y aplicará los métodos actuales para recuperar información a partir de colecciones de documentos semi-estructurados y no estructurados.

### **4.- ESPECIFICOS**

- 1.- El estudiante comprenderá los conceptos genéricos de los sistemas de recuperación de información.
- 2.- El estudiante entenderá el concepto de indexación, las estructuras de datos involucradas así como los retos en el tratamiento de grandes volúmenes de información
- 3.- El estudiante revisará modelos de indexación y de consulta a colecciones de documentos semi-estructurados. Analizará el rendimiento de sistemas basados en consultas sobre este tipo de información, así como la precisión de los resultados.
- 4.- El estudiante conocerá y explorará las técnicas actuales para la recuperación de información en bases de datos no estructuradas.



#### 4. CONTENIDO

Unidad	Contenido Temático
I. INTRODUCCIÓN	1.1 El concepto de recuperación de información 1.2 Información estructurada 1.3 Información semi-estructurada 1.4 Información no estructurada 1.5 Componentes de un sistema de recuperación de información
2.RECUPERACIÓN BOOLEANA	2.1 Ejemplos de recuperación de información 2.2 El modelo booleano de recuperación de información 2.3 Índices invertidos 2.4 El modelo booleano extendido de recuperación de información 2.5 Ranking
3.CONSULTAS AVANZADAS	3.1 Índices posicionales 3.2 Pre-procesamiento 3.2.1 Tablas de códigos de caracteres 3.2.2 Identificación de idiomas 3.2.3 Tokenización 3.2.4 Aplicación de módulos de procesamiento del lenguaje natural 3.2.5 Extracción de términos 3.3 Diferentes niveles de consulta (letra, palabra, sentencia, etc.) 3.4 Corrección ortográfica 3.5 Corrección fonética
4.OTROS MODELOS DE RECUPERACIÓN DE INFORMACIÓN	4.1 El modelo de espacio vectorial 4.2 Modelos probabilísticos de recuperación de información 4.3 Recuperación de información semi-estructurada 4.4 Indexación distribuida 4.5 Indexación dinámica 1.6 Compresión de índices
5.INDEXACIÓN DE GRANDES VOLÚMENES DE INFORMACIÓN	5.1 Conceptos básicos de la Web 5.2 Crawling 5.3 Índices distribuidos 5.4 Conectividad de Servidores 5.5 Pageranking



<b>Bibliografía</b>	
<b>Básica</b>	<b>Complementaria</b>
<p>1. Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2008.</p> <p>2.- Witten, I.H., Moffat, A., and Bell, T.C. Managing Gigabytes: Compressing and indexing documents and images. Second Edition, Morgan Kaufmann, Spring 1999.</p> <p>3.- Information Retrieval: Data Structures and Algorithms. William B. Frakes, Ricardo Baeza-Yates. Prentice Hall PTR. 1992.</p> <p>4.- Information Retrieval: Algorithms and Heuristics(The Information Retrieval Series). David A.Grossman and Ophir Frieder. Springer; 2nd edition, 2004.</p>	

## 5. CRITERIOS DE EVALUACIÓN

Criterios	Porcentaje
• Exámenes	40%
• Participación en clase	
• Tareas	20%
• Exposiciones	
• Simulaciones	
• Trabajo de investigación y/o de intervención	
• Prácticas de laboratorio	20%
• Visitas guiadas	
• Reporte de actividades académicas y culturales	
• Mapas conceptuales	
• Proyecto final	20%
Total	100%