



UNIVERSIDAD SIMÓN BOLÍVAR
DIVISIÓN DE FÍSICA Y MATEMÁTICAS
DEPARTAMENTO DE COMPUTACIÓN Y
TECNOLOGÍA DE LA INFORMACIÓN

DIVISIÓN:	Física y Matemáticas
DEPARTAMENTO:	Computación y Tecnología de la Información
ASIGNATURA:	Recuperación Inteligente de Información en Internet
CÓDIGO:	CI-6454
HORAS/SEMANA:	Teoría: 3 Práctica: 1 Laboratorio: 0
VIGENCIA:	Desde Septiembre de 1999

PROGRAMA

OBJETIVOS GENERALES:	<p>Con este curso se pretende que el estudiante:</p> <ol style="list-style-type: none">1. Se familiarice con los modelos de recuperación de información no estructurada, que se utilizan en la actualidad.2. Aprenda y aplique técnicas avanzadas de procesamiento de texto, para extraer información de documentos.3. Estudie y desarrolle sistemas de recuperación de información en Internet.
OBJETIVOS ESPECÍFICOS:	<p>A nivel específico, se espera que el estudiante:</p> <ol style="list-style-type: none">1. Aprenda y entienda el proceso de recuperación de información no estructurada, como un proceso de comunicación completo, con sintaxis, semántica y pragmática.2. Analice y compare los enfoques de recuperación de información utilizando lenguaje natural y utilizando un vocabulario controlado.3. Compare distintos enfoques para realizar la búsqueda de información no estructurada y, además, seleccione y utilice heurísticas para búsquedas particulares.4. Comprenda la organización de información en Internet.

	<ol style="list-style-type: none"> 5. Entienda y utilice sistemas de recuperación de información no estructurada en Internet. 6. Analice distintos casos de agentes inteligentes en el Web. 7. Desarrolle agentes inteligentes capaces de realizar tareas tales como: clasificar, sumarizar, capturar o filtrar mensajes o documentos en el Web.
<p>CONTENIDO DETALLADO TEORÍA:</p>	<ol style="list-style-type: none"> 1. Introducción al área de recuperación de información no estructurada. 2. Modelos de Recuperación: formulación y evaluación. 3. Lenguajes y operaciones de interrogación. 4. Representaciones de textos: metadatos, formatos, lenguajes etiquetados (HTML, SGML, XML). 5. Operaciones sobre textos: clasificación de documentos, extracción de información, entendimiento de textos. 6. Caracterización del Web. 7. Recuperación de información en Internet. 8. Nuevas tecnologías (navegadores, motores de búsqueda, metabuscadores, <i>spiders</i>, <i>crawlers</i>) y protocolos (HTTP). 9. Agentes inteligentes en el Web: casos de estudio.
<p>BIBLIOGRAFÍA:</p>	<ol style="list-style-type: none"> 1. J. Allen. <i>Natural Language Understanding.</i> Benjamin Publishing, 1995. 2. R. A. Cole et al. (eds.) <i>Survey of the State of the Art in Human Language Technology</i>, 1996. Disponible electrónicamente en http://cslu.cse.ogi.edu/HLTSurvey. 3. R. Baeza-Yates y Ribeiro-Neto. <i>Modern Information Retrieval.</i> ACM Press Series, 1999. 4. S. Russell y P. Norvig. <i>Artificial Intelligence: A Modern Approach.</i> Prentice Hall, 1995.