



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD CIENCIAS DE LA COMPUTACION

PROGRAMA DE LA MATERIA CORRESPONDIENTE A LA LICENCIATURA EN CIENCIAS DE LA COMPUTACIÓN.

Coordinación:

Área de

NOMBRE DE LA MATERIA: Tópicos Selectos en Análisis de Datos.

Clave: LCC 597

Créditos: 10

Modalidad: Escolarizada

Nivel de Ubicación: Formativo

Tipo de Materia: Optativa

PRE-REQUISITOS: Nivel Básico

MATERIA CONSECUENTE: Ninguna

TIEMPO TOTAL ASIGNADO: 80 Hrs.

PRIMAVERA – OTOÑO

HRS. TEÓRICAS/SEM: 5 **HRS. PRÁCTICAS/SEM:** 0

VERANO

HRS. TEÓRICAS/SEM: 10 **HRS. PRÁCTICAS/SEM:** 0

AUTOR(ES) DEL PROGRAMA:

Lic. María del Rocío Boone Rojas

Dra. María Josefa Somodevilla García

Dra. Darnes Vilariño Ayala

REVISADO POR: Área de Bases de Datos e Ingeniería de Software

APROBADO POR: FCC

AUTORIZADO POR: Vicerrectoría de Docencia

FECHA DE ELABORACIÓN/REVISIÓN: Septiembre 2007

VIGENCIA: A partir de Enero 2008

JUSTIFICACIÓN:

Las necesidades actuales del procesamiento de información hacen que las empresas de gestión tanto públicas como privadas estén interesadas en una explotación inteligente de sus bases de datos. En este curso se imparten los conocimientos necesarios para el preprocesamiento, análisis y modelado efectivo de los datos operativos e informacionales de la organización, mediante el uso de técnicas y herramientas de minería de datos (aprendizaje automático). A lo largo del curso, se describen las metodologías que permiten paso a paso analizar y modelar datos en casos reales.

OBJETIVO GENERAL DE LA MATERIA:

Que el alumno conozca la necesidad y conceptos generales de la tecnología de almacenes de datos y OLAP y saber aplicar las técnicas de minería de datos apropiadas para problemas concretos de extracción de conocimiento, útil para el análisis o la toma de decisiones.

CONTRIBUCIÓN DE LA ASIGNATURA AL PERFIL DE EGRESO:

En el perfil del egresado se establece que éste tendrá una visión general de las Ciencias de la Computación y poseerá conocimientos sólidos para la construcción de soluciones basadas en Sistemas de Software. Un área importante de la computación son los Sistemas Inteligentes, los cuales soportan el análisis de datos. Esta disciplina computacional debe ser conocida en detalle en virtud de que la información se ha convertido en el activo más importante de la mayoría de las organizaciones. Por otra parte, gracias al advenimiento de Internet, la información se ha hecho compleja y extremadamente voluminosa, entonces son imperativas las técnicas para el almacenamiento y manipulación de este tipo de información, las cuales son: los almacenes de datos y las técnicas de minería de datos. Por ello, la asignatura de Análisis de Datos tiene una contribución indispensable en la formación del estudiante.

CONTENIDO TEMÁTICO

MATERIA:

UNIDAD: 1			TÍTULO: Introducción			
<p>OBJETIVO ESPECÍFICO:</p> <ul style="list-style-type: none"> Que el estudiante comprenda y sepa definir el problema del análisis de sistemas de información para la toma de decisiones y las herramientas genéricas y las tecnologías más avanzadas para llevarlo a cabo. <p>Bibliografía: [1]</p>						
CONTENIDO DE LA UNIDAD		Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
		HT	HP			
1.1	Evolución de los Sistemas de Información	2		Ubicar en el contexto actual el desarrollo de los sistemas de información.	Exposición del profesor, discusión grupal.	Salón, pizarrón, plumones, proyector de video o acetatos.
1.2	Sistemas Transaccionales	2		Entender el concepto de transacción como una actividad algorítmica	Exposición del profesor, discusión grupal.	Salón, pizarrón, plumones, proyector de video o acetatos.
1.3	Sistema para la toma de Decisiones	4		Diferenciar el proceso de toma de decisiones, como un proceso que posee un alto nivel de abstracción	Exposición del profesor, discusión grupal.	Salón, pizarrón, plumones, proyector de video o acetatos.
HORAS TOTALES:		8				

UNIDAD: 2		TÍTULO: Almacenes de datos			
<p>OBJETIVO ESPECÍFICO:</p> <ul style="list-style-type: none"> • Que el estudiante conozca las técnicas de Almacenes de Datos y OLAP para facilitar el procesamiento analítico de apoyo a la toma de decisiones estratégicas. • Que el estudiante conozca las arquitecturas actuales que soportan los almacenes de datos. • Que el estudiante conozca y aprenda a identificar los distintos aspectos a considerar en el diseño y la implementación de un almacén de datos. <p>Bibliografía: [1]</p>					
CONTENIDO DE LA UNIDAD	Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
	HT	HP			
2.1 Estructura de un Almacén de Datos (Data Warehouse) 2.1.1. Cubos De Información (DataMarts)	2		Introducción y Motivación; Comprensión y Elicitación de Ideas.	Exposición del Profesor.	Salón, pizarrón, plumones, proyector de video o acetatos.
2.2 Elementos que integran un Data Warehouse 2.2.1. Metadata 2.2.2. Middleware 2.2.3. API 2.2.4. Mecanismos de Extracción (OLAP) 2.2.4.1. Mecanismos de Carga	2		Comprensión y Elicitación de Ideas. Conocer y comprender definiciones	Exposición del Profesor.	Salón, pizarrón, plumones, proyector de video o acetatos.
2.3 Diseño de un Data Warehouse	4		Comprensión y Elicitación de Ideas. Conocer las condiciones y elementos para el diseño de una data warehouse	Exposición del Profesor; Solución de preguntas y/o problemas (ejercicio resuelto en equipo donde se determinen diferencias entre bases de datos relacionales y data warehouses)	Salón, pizarrón, plumones, proyector de video o acetatos.

2.4	Modelado conceptual del almacén de datos 2.4.1. Características del Modelo Multidimensional 2.4.1.1. Esquema Estrella 2.4.1.2. Esquema copo de nieve (snowflake) 2.4.1.3. Tabla Fact o de Hechos 2.4.1.4. Tablas Lock-up o Dimensionales	4	Comprensión y Elicitación de Ideas. Conocer las condiciones para la implementación de un modelo multidimensional ROLAP	Exposición del Profesor; Solución de preguntas y/o problemas (ejercicio resuelto en equipo donde se identifiquen y apliquen los elementos de una arquitectura multidimensional estrella o copo de nieve)	Salón, pizarrón, plumones, proyector de video o acetatos.
2.5	Metodología para el Modelado Multidimensional	2			
HORAS TOTALES:		14			

UNIDAD: 3			TÍTULO: Introducción a la Minería de Datos			
OBJETIVO ESPECÍFICO:						
<ul style="list-style-type: none"> Que el estudiante comprenda la necesidad del análisis inteligente y automático de la información para el descubrimiento de conocimiento útil. Bibliografía: [1,2]						
CONTENIDO DE LA UNIDAD		Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
		HT	HP			
3.1	Motivación	2		Introducción y Motivación; Comprensión y Elicitación de Ideas. Conocer las condiciones y elementos para el diseño de una data warehouse	Exposición del Profesor; Solución de preguntas y/o problemas (ejercicio resuelto en equipo donde se identifiquen problemas a ser resueltos aplicando minería de datos)	Salón, pizarrón, plumones, proyector de video o acetatos.
3.2	Problemas tipo y aplicaciones	4				
3.3	Relación de MD con otras disciplinas	4				

	HORAS TOTALES:	10		
--	----------------	----	--	--

UNIDAD: 4	TÍTULO: El proceso de KDD
------------------	----------------------------------

- **OBJETIVO ESPECÍFICO:** Que el estudiante conozca las fases del Descubrimiento de Conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso (en especial las de limpieza y selección de datos).
Bibliografía: [3, 4]

CONTENIDO DE LA UNIDAD		Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
		HT	HP			
4.1	Las Fases del KDD	2		Introducción, Motivación, Comprensión y Elicitación de Ideas y conceptos. Diferenciar el propósito de cada una de las fases del KDD. Establecer características de cada tipo de técnica de DM e identificar tipos de problemas a ser resueltos por cada una. Comparar capacidades de los diferentes sistemas comerciales	Exposición del Profesor; Discusión grupal, lluvia de ideas, ejemplos y participación del grupo en equipos de trabajo.	Salón, pizarrón, plumones, proyector de acetatos o de video.
4.2	Tipología de Técnicas de Minería de Datos	2				
4.3	Sistemas Comerciales	2				
HORAS TOTALES:		6				

UNIDAD: 5		TÍTULO: Técnicas de Minería de Datos				
<p>OBJETIVO ESPECÍFICO:</p> <ul style="list-style-type: none"> • Que el estudiante conozca los beneficios del uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos. • Que el estudiante conozca y pueda comparar las técnicas de aprendizaje automático más habituales en minería de datos. Se estudiará la idoneidad de cada una para diferentes problemas. <p>Bibliografía: [2,3,4]</p>						
CONTENIDO DE LA UNIDAD		Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
		HT	HP			
5.1	El Problema de la Extracción Automática de Conocimiento.	2		Introducción, Motivación, Comprensión y Elicitación de Ideas y conceptos. Diferenciar las características y modo de aplicación del aprendizaje supervisado y no supervisado	Exposición del Profesor; Discusión grupal, lluvia de ideas, ejemplos y participación del grupo en equipos de trabajo.	PC, programa de software para minería de datos (Weka)
5.2	Evaluación de Hipótesis	5				
5.3	Técnicas no supervisadas y descriptivas.	10				
5.4	Técnicas supervisadas y predictivas.	10				
HORAS TOTALES:		27				

UNIDAD: 6		TÍTULO: Minería de WEB y Textos				
OBJETIVO ESPECÍFICO: <ul style="list-style-type: none"> Que el estudiante conozca la problemática especial de la minería de datos sobre la web (documentos textuales e hipertextuales) y las técnicas más usuales. Bibliografía: [1, 4]						
CONTENIDO DE LA UNIDAD		Tiempo de impartición (hrs.)		Actividades de Aprendizaje	Técnicas	Recursos Necesarios
		HT	HP			
6.1	Los Problemas de la Información No Estructurada.	2		Identificar fuentes de información no estructurada	Exposición del Profesor;	PC, programa de software para minería de datos (Weka)
6.2	Extracción de Conocimiento a partir de Documentos HTML y texto.	5		Conocer los formatos de archivos html y de texto	Exposición del Profesor;	PC, programa de software para minería de datos (Weka)
6.3	Extracción de Información semi-estructurada (XML).	8		Conocer las condiciones que diferencian la extracción de información estructurada de la no estructurada.	Exposición del Profesor; Solución de preguntas y/o problemas (ejercicio en equipo donde trabaje con documentos html y xml)	PC, programa de software para minería de datos (Weka)
HORAS TOTALES		15				

BIBLIOGRAFÍA:

1. José Hernández Orallo, M. José Ramírez Quintana, César Ferri Ramírez, Introducción a la Minería de Datos, Editorial Pearson, . ISBN: 84 205 4091 9, 2004
2. Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.
3. Ian Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann, ISBN 0120884070, 2005.
4. K. Cios, W. Pedrycz, R. Swiniarski, L. Kurgan, Data Mining: A Knowledge Discovery Approach, Springer, ISBN: 978-0-387-33333-5, 2007.

CRITERIOS DE EVALUACIÓN

	%
Asistencias y participación:	10
Exámenes parciales:	50
Tareas y Actividades:	20
Trabajos de Investigación:	20
TOTAL:	100