

Universidad Católica San Pablo
Facultad de Ingeniería y Computación
Escuela Profesional de
Ciencia de la Computación
SILABO



CS270T. Bases de Datos I (Obligatorio)

2010-1

1. DATOS GENERALES

1.1 CARRERA PROFESIONAL	:	Ciencia de la Computación
1.2 ASIGNATURA	:	CS270T. Bases de Datos I
1.3 SEMESTRE ACADÉMICO	:	4 ^{to} Semestre.
1.4 PREREQUISITO(S)	:	CS107. Álgebra Abstracta. (3 ^{er} Sem)
1.5 CARÁCTER	:	Obligatorio
1.6 HORAS	:	2 HT; 2 HP; 4 HL;
1.7 CRÉDITOS	:	5

2. DOCENTE

3. FUNDAMENTACIÓN DEL CURSO

La gestión de la información (IM) juega un rol principal en casi todas las áreas donde los computadores son usados. Esta área incluye la captura, digitalización, representación, organización, transformación y presentación de información; algoritmos para mejorar la eficiencia y efectividad del acceso y actualización de información almacenada, modelamiento de datos y abstracción, y técnicas de almacenamiento de archivos físicos.

Este también abarca la seguridad de la información, privacidad, integridad y protección en un ambiente compartido. Los estudiantes necesitan ser capaces de desarrollar modelos de datos conceptuales y físicos, determinar que métodos de (IM) y técnicas son apropiados para un problema dado, y ser capaces de seleccionar e implementar una apropiada solución de IM que refleje todas las restricciones aplicables, incluyendo escalabilidad y usabilidad.

4. SUMILLA

1. IM/Modelos de Información. 2. IM/Sistemas de Base de Datos. 3. IM/Modelamiento de Datos. 4. IM/Indexación. 5. IM/Base de Datos Relacionales. 6. IM/Lenguajes de Consultas de Base de Datos. 7. IM/Diseño de Bases de Datos Relacionales.

5. OBJETIVO GENERAL

- Que el alumno aprenda a representar información en una base de datos priorizando la eficiencia en la recuperación de la misma
- Que el alumno aprenda los conceptos fundamentales de gestión de bases de datos. Esto incluye aspectos de diseño de bases de datos, lenguajes de bases de datos y realización de bases de datos
- Discutir el modelo de bases de datos con base en el álgebra relacional, cálculo relacional y en el estudio de sentencias SQL.

6. CONTRIBUCIÓN A LA FORMACIÓN PROFESIONAL Y FORMACIÓN GENERAL

Esta disciplina contribuye al logro de los siguientes resultados de la carrera:

- b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. [Nivel Bloom: 4]
- d) Trabajar efectivamente en equipos para cumplir con un objetivo común. [Nivel Bloom: 3]
- e) Entender correctamente las implicancias profesionales, éticas, legales, de seguridad y sociales de la profesión. [Nivel Bloom: 3]
- i) Utilizar técnicas y herramientas actuales necesarias para la práctica de la computación. [Nivel Bloom: 3]
- j) Aplicar la base matemática, principios de algoritmos y la teoría de la Ciencia de la Computación en el modelamiento y diseño de sistemas computacionales de tal manera que demuestre comprensión de los puntos de equilibrio involucrados en la opción escogida. [Nivel Bloom: 3]

7. CONTENIDOS

UNIDAD 1: IM/Modelos de Información.(14 horas)

Nivel Bloom: 3

OBJETIVO GENERAL

CONTENIDO

- Comparar y contrastar la información con datos y conocimiento.
- Criticar y defender las aplicaciones de información de tamaño pequeño y mediano con respecto a la satisfacción de las necesidades reales del usuario.
- Mostrar explícitamente la relación entre metadata/schema almacenados y los datos.
- Explicar el uso de consultas declarativas.
- Dar una versión declarativa de una consulta de navegación.
- Describir varias soluciones técnicas para problemas relacionados a la privacidad, integridad, seguridad y preservación de la información.
- Explicar las medidas de eficiencia (estimación, tiempo de respuesta) y efectividad (*precision - recall*).
- Describir métodos para asegurar que los sistemas de información pueden escalar de lo individual a lo global.
- Identificar asuntos relacionados a la persistencia de datos en una organización.
- Describir vulnerabilidades de la integridad de datos en escenarios específicos.

- Almacenamiento y recuperación de información (IS&R).
- Aplicaciones de administración de la información.
- Representación y captura de la información.
- Asociación de Metadata/schema con los datos
- Indexación y análisis.
- Búsqueda, recuperación, enlace, navegación.
- Privacidad, integridad, seguridad y preservación de la información.
- Escalabilidad, eficiencia y efectividad.
- Escalabilidad, eficiencia y efectividad.
- Conceptos relacionados con asegurar información (persistencia de datos).

Lecturas: [Veryard, 1994], [Elmasri and Navathe, 2004], [Oppel, 2004], [Korth and Silberschatz, 2002]

UNIDAD 2: IM/Sistemas de Base de Datos.(14 horas)	
Nivel Bloom: 3	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Explicar las características que distinguen a las bases de datos de los métodos tradicionales de programación con archivos de datos. ▪ Citar el objetivo, funciones, modelos, componentes, aplicaciones y el impacto social de los sistemas de bases de datos. ▪ Describir los componentes de un sistema de base de datos y dar ejemplos de su uso. ▪ Identificar las funciones superiores DBMS y describir su rol en un sistema de base de datos. ▪ Explicar los conceptos de independencia de datos y su importancia en un sistema de base de datos. ▪ Usar un lenguaje de consulta para elicitar la información de una base de datos. 	<ul style="list-style-type: none"> ▪ Historia y motivación de los sistemas de base datos. ▪ Componentes de los sistemas de base de datos. ▪ Funciones DBMS. ▪ Arquitectura de base de datos e independencia de datos. ▪ Uso de un lenguaje de consultas declarativo.
Lecturas: [Rob and Coronel, 2004], [Elmasri and Navathe, 2004], [Date, 2005], [Korth and Silberschatz, 2002]	

UNIDAD 3: IM/Modelamiento de Datos.(14 horas)	
Nivel Bloom: 4	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Categorizar los modelos de datos basados en los tipos de conceptos que ellos proveen para describir la estructura de las bases de datos, esto es, el modelo de datos conceptual, el modelo de datos físico y el modelo de datos representacional. ▪ Describir los conceptos de modelado y la notación del modelo entidad-relación y UML, incluyendo su uso en modelamiento de datos. ▪ Describir los principales conceptos del modelo OO tal como la identidad del objeto, constructores de tipos, encapsulación, herencia, polimorfismo y creación de versiones. ▪ Definir la terminología fundamental usada en el modelo de datos relacional. ▪ Describir los principios básicos del modelo de datos relacional. ▪ Ilustrar los conceptos de modelamiento y notación del modelo de datos relacional. ▪ Describir las diferencias en los modelos de datos relacional y semiestructurado. ▪ Generar un modelo semiestructurado (DTD o XMLSchema) equivalente a un esquema relacional dado. 	<ul style="list-style-type: none"> ▪ Modelamiento de datos. ▪ Modelos conceptuales (incluyendo entidad-relación y UML). ▪ Modelo orientado a objetos. ▪ Modelo de datos relacional. ▪ Modelos de datos semiestructurados (expresados utilizando DTD o XMLSchema).
Lecturas: [Simsion and Witt, 2004], [Elmasri and Navathe, 2004], [Korth and Silberschatz, 2002]	

UNIDAD 4: IM/Indexación.(4 horas)	
Nivel Bloom: 5	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Generar un archivo índice para una selección de recursos. ▪ Explicar el rol de un índice invertido en la localización de un documento en una colección. ▪ Explicar como el proceso de encontrar la raíz de una palabra (<i>stemming</i>) y las palabras no relevantes (<i>stop words</i>) afectan la indexación. ▪ Identificar los índices apropiados para un determinado esquema relacional un una consulta dada. ▪ Estimar el tiempo de recuperación de la información con y sin índices. 	<ul style="list-style-type: none"> ▪ El impacto masivo de los índices en el desempeño de consultas. ▪ La estructura básica de un índice. ▪ Manteniendo un <i>buffer</i> de datos en memoria. ▪ Creación de índices con SQL. ▪ Indexación de texto. ▪ Indexación de la web y como trabajan los motores de búsqueda.
Lecturas: [Whitehorn and Marklyn, 2001], [Date, 2005], [Korth and Silberschatz, 2002]	

UNIDAD 5: IM/Base de Datos Relacionales.(14 horas)	
Nivel Bloom: 5	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Preparar un esquema relacional de un modelo conceptual usando el modelo entidad-relación. ▪ Explicar y demostrar los conceptos de restricciones de la integridad de la entidad y restricciones de la integridad referencial (incluyendo la definición del concepto de llave foránea). ▪ Demostrar el uso de las operaciones del álgebra relacional desde la teoría de conjuntos matemáticos (unión, intersección, diferencia y producto cartesiano) y las operaciones de álgebra relacional desarrolladas específicamente para bases de datos relacionales (<i>select (restrict)</i> , <i>product</i>, <i>join</i> y <i>division</i>). ▪ Demostrar consultas en el álgebra relacional. ▪ Demostrar consultas en el cálculo relacional de tuplas. 	<ul style="list-style-type: none"> ▪ Mapeo del esquema conceptual al esquema relacional. ▪ Entidad e integridad referencial. ▪ Álgebra relacional y cálculo relacional.
Lecturas: [Whitehorn and Marklyn, 2001], [Date, 2005], [Korth and Silberschatz, 2002]	

UNIDAD 6: IM/Lenguajes de Consultas de Base de Datos.(12 horas)	
Nivel Bloom: 5	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Crear un esquema de base de datos relacional en SQL que incorpora restricciones de integridad referencial, integridad-entidad, llaves. ▪ Demostrar la definición de datos en SQL y recuperar información de una base de datos usando la sentencia SQL SELECT. ▪ Evaluar un conjunto de estrategias de procesamiento de consultas y seleccionar la óptima. ▪ Crear una consulta no procedimental por medio de llenado de plantillas de relaciones para construir un ejemplo del resultado de la consulta deseada. ▪ Incrustar consultas orientadas a objetos en un lenguaje tal como C++ o Java (ejemplo, <i>SELECT Col.Method() FROM Object</i>). ▪ Escribir un procedimiento almacenado que reciba parámetros y que tenga algún flujo de control para proveer alguna funcionalidad. 	<ul style="list-style-type: none"> ▪ Síntesis de los lenguajes de bases de datos. ▪ SQL (definición de datos, formulación de consultas, actualización del sublenguaje, restricciones e integridad). ▪ <i>Query by Example</i> (QBE) y entornos de 4ta generación. ▪ Consultas no procedurales incrustadas en un lenguaje procedimental. ▪ Introducción al lenguaje de consultas orientado a objetos. ▪ Procedimientos almacenados.
Lecturas: [Dietrich, 2001], [Elmasri and Navathe, 2004], [Celko, 2005], [Korth and Silberschatz, 2002]	

UNIDAD 7: IM/Diseño de Bases de Datos Relacionales.(12 horas)	
Nivel Bloom: 4	
OBJETIVO GENERAL	CONTENIDO
<ul style="list-style-type: none"> ▪ Determinar la dependencia funcional entre dos o más atributos que son un subconjunto de una relación. ▪ Conectar las restricciones expresadas como llave primaria y llaves foráneas con dependencias funcionales. ▪ Calcular cerradura de un conjunto de atributos bajo una dependencia funcional dada. ▪ Determinar si un conjunto de atributos forma o no una super llave y/o es candidato a ser llave para una dependencia funcional dada. ▪ Evaluar una descomposición propuesta para decir si tiene o no tiene <i>lossless-join</i> y preservación de dependencia. ▪ Describir que significa 1NF, 2NF, 3NF y BCNF. ▪ Identificar si una relación es una 1NF, 2NF, 3NF o BCNF. ▪ Normalizar una 1NF en un conjunto de relaciones en 3NF (o BCNF) y desnormalizar un esquema relacional. ▪ Explicar el impacto de la normalización sobre la eficiencia de las operaciones de base de datos, especialmente la utilización de consultas. ▪ Describir que es una dependencia multivaluada y que tipo de restricciones ésta específica. ▪ Explicar por qué 4NF es útil en el diseño del esquema. 	<ul style="list-style-type: none"> ▪ Diseño base de datos. ▪ Dependencia funcional. ▪ Descomposición de un esquema: <i>lossless-join</i> y propiedades de preservación de independencia de una descomposición. ▪ Llaves candidatas, super llaves, cerradura de un conjunto de atributos. ▪ Formas normales (1NF, 2NF, 3NF, BCNF). ▪ Dependencia multivaluada (4NF). ▪ <i>Join dependency</i> (PJNF, 5NF). ▪ Representación teórica.
Lecturas: [Harrington, 2002], [Elmasri and Navathe, 2004], [Date, 2005], [Korth and Silberschatz, 2002]	

8. METODOLOGÍA

El profesor del curso presentará clases teóricas de los temas señalados en el programa propiciando la intervención de los alumnos.

El profesor del curso presentará demostraciones para fundamentar clases teóricas.

El profesor y los alumnos realizarán prácticas

Los alumnos deberán asistir a clase habiendo leído lo que el profesor va a presentar. De esta manera se facilitará la comprensión y los estudiantes estarán en mejores condiciones de hacer consultas en clase.

9. EVALUACIONES

Evaluación Permanente 1 : 20 %

Examen Parcial : 30 %

Evaluación Permanente 2 : 20 %

Examen Final : 30 %

Referencias

- [Celko, 2005] Celko, J. (2005). *Joe Celko's SQL Programming Style*. Elsevier.
- [Date, 2005] Date, C. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Elsevier.
- [Dietrich, 2001] Dietrich, S. W. (2001). *Understanding Relational Database Query Languages, First Edition*. Prentice Hall.
- [Elmasri and Navathe, 2004] Elmasri, R. and Navathe, S. B. (2004). *Fundamentals of Database Systems, Fourth Edition*. Addison Wesley.
- [Harrington, 2002] Harrington, J. L. (2002). *Relational Database Design Clearly Explained, Second Edition*. Morgan Kaufmann.
- [Korth and Silberschatz, 2002] Korth, H. F. and Silberschatz, A. (2002). *Fundamentos de Base de Datos*. McGraw-Hill.
- [Oppel, 2004] Oppel, A. (2004). *Databases Demystified*. Mc Graw Hill Osborne.
- [Rob and Coronel, 2004] Rob, P. and Coronel, C. (2004). *Database Systems: Design, Implementation and Management, Sixth Edition*. Morgan Kaufmann.
- [Simsion and Witt, 2004] Simsion, G. and Witt, G. (2004). *Data Modeling Essentials, Third Edition*. Morgan Kaufmann.
- [Veryard, 1994] Veryard, R. (1994). *Information Coordination: The Management of Information Models, Systems and Organizations*. Prentice Hall.
- [Whitehorn and Marklyn, 2001] Whitehorn, M. and Marklyn, B. (2001). *Inside Relational Databases, Second Edition*. Springer.