



## MAESTRÍA EN TIAE – MODALIDAD A DISTANCIA

<b>Representación, Almacenamiento y Recuperación del Conocimiento</b>  <b>Año 2019</b>	<b>Docente Responsable: Lanzarini Laura</b>
	<b>Docentes tutores:</b> <b>Bertone Rodolfo</b> (académico) <b>Thomas Pablo</b> (académico/tecnológico) <b>Lanzarini Laura</b> (académico/tecnológico)
	<b>Tutores:</b> <b>Hasperué Waldo</b> (académico/tecnológico) <b>Quiroga Facundo</b> (académico/tecnológico) <b>Mesía Salazar Natalí</b> (tecnológico) <b>Meza Maitén</b> (administrativo)
	La cantidad de tutores se incrementa acorde a la cantidad de alumnos inscriptos
	<b>Cantidad de horas presenciales/VC: 40 hs.</b> <b>Cantidad de horas de actividades en línea y de trabajo final: 130 hs.</b>
	Se dicta de agosto a diciembre.

**ACLARACION INICIAL**

*Esta asignatura está dividida en dos Módulos. El Módulo 1 está vinculado con los fundamentos y conceptos que se involucran con la representación y almacenamiento de la información. El Módulo 2 se focaliza en los aspectos más vinculados a la recuperación del conocimiento. Estos dos módulos se integran para dar una visión más amplia de la temática al alumno y una vinculación directa al área educativa.*



*Para facilitar la presentación, del programa, se presentarán por separado los ítems de objetivos, contenidos y bibliografía. La metodología y evaluación son estrategias compartidas y se presentan integradas.*

### **OBJETIVOS GENERALES:**

Que los alumnos sean capaces de:

- Conocer los conceptos básicos sobre Bases de Datos (BD)
- Reconocer estrategias de diseño de BD
- Realizar una representación de la información en una BD a partir de modelos de entidad – relación para un escenario educativo concreto.
- Realizar consultas en una BD a través de instrucciones SQL básicas
- Conocer y aplicar técnicas vinculadas al área de Minería de Datos y Sistemas Expertos, que les permitan extraer información de interés a partir de Bases de Datos para sus prácticas docentes.

#### **Objetivos Específicos del Módulo I**

En este módulo se orienta a que los alumnos identifiquen y ensayen la representación de datos de forma estructurada a partir de modelos entidad relación y practiquen realizar consultas a partir de instrucciones SQL simples.

#### **Objetivos Específicos del Módulo II**

Este módulo se propone que los alumnos reconozcan y ensayen distintas estrategias basadas en Inteligencia Artificial y Minería de Datos para la recuperación de conocimiento. Se focalizará en la recuperación de conocimiento con el objetivo de brindar una mejora en los procesos de enseñar y aprender.

En ese sentido, se analiza el uso de Sistemas Expertos y Sistemas Adaptativos en entornos educativos. También se describen y ejemplifican distintas técnicas de Minería de Datos con el objetivo de caracterizar y modelizar la información disponible sobre procesos educativos para luego extraer conocimiento de interés para los docentes.

### **MODALIDAD DE EVALUACIÓN Y ACREDITACIÓN**



Para aprobar esta asignatura el alumno deberá participar de los encuentros obligatorios ya sea por VC y/o presencialmente. Deberá aprobar el 100% de las actividades planificadas para el curso ya sea en los encuentros o mediadas por el entorno virtual de enseñanza y aprendizaje utilizado en el marco del SIED. Entre las actividades se incluye la lectura de trabajos de investigación actualizados en el área y ejercicios de aplicación que se van resolviendo a lo largo del curso para poner en juego los temas trabajados. Los ejercicios se resuelven con herramientas de software específicas, que se ponen a disposición de los estudiantes a través del entorno virtual de enseñanza y aprendizaje.

Además se debe aprobar una evaluación sumativa que consiste en realizar/analizar un modelo de representación BD para un contexto educativo particular, y la aplicación de consultas del tipo SQL para extraer información. El trabajo se continúa durante el segundo módulo, donde como cierre del módulo se deberán proponer técnicas de minería de datos, que posibiliten la obtención de información de interés para un contexto educativo particular.

Los trabajos son compartidos entre los estudiantes para socializar las estrategias empleadas, luego de su entrega.

## PROGRAMA

### I. MOTIVACIÓN

La temática de bases de datos (BD) constituye un eje central dentro de la disciplina Informática. Desde el punto de vista de las metas de las carreras, en esta materia se abordan temas conceptuales de BD, que permiten a los alumnos conocer cómo se representan los datos de un problema, y sus relaciones. El acceso posterior a estas bases de datos es una tarea asidua de muchos docentes, pero sin conocer la naturaleza de los datos y sus formas de modelización. En este sentido, se busca que los alumnos alcancen una mayor comprensión de estos temas para poder luego poder extraer información de interés de bases de datos para el contexto educativo en el que se desempeñan. Por poner un caso, actualmente la mayoría de las universidades nacionales, trabajan con sistemas de gestión como, por ejemplo, Siu Guaraní. Los docentes incorporan información que se almacena en bases de datos, asociadas a estos sistemas. En varias instituciones hoy se está realizando estudios sobre variables que pueden afectar en el desgranamiento y la deserción de sus estudiantes, a partir de la extracción de información, a partir de estas bases de datos. Esta materia se propone que los alumnos puedan conocer y aplicar técnicas de modelado y extracción para poder participar en proyectos institucionales vinculados a estas temáticas, en su investigación, y/o poder realizar tareas sencillas de consultas a bases de datos.



## II. CONTENIDOS

### Módulo I

#### 1- Conceptos generales:

Definición de bases de datos. Diferencias entre Bases de datos y almacenamiento en archivos

Etapas en la implementación de una base de datos. Distintos tipos de DBMS (Data Base Management System).

#### 2- Modelos de datos

Diseño de una BD.

Esquema Conceptual

Esquema Lógico

Esquema Físico

#### 3- Modelo de Entidad Relación

Introducción al Modelo de E/R

Conceptos del modelo de E/R:

Entidad/Tipo de Entidad. Atributo, Dominio, Clave

Relación/Tipo de relación. Cardinalidad

Roles

Generalización/Especialización

Agregación

Pasaje a tablas

#### 4- Recuperación de la información

Lenguaje de consultas SQL: definición

Operaciones básicas: selección, proyección, producto cartesiano, unión, diferencia.

Operadores más complejos: unión natural, cociente, intersección

SQL como lenguaje de consulta. Transformación de AR a SQL. Consultas básicas.

Clausulas SELECT, FROM, WHERE, GROUP BY, ORDER BY

#### 5- Modelado de ejercicios usando el modelo de entidades y relaciones.

#### 6- Ejemplos y ejercicios de consultas usando SQL básico

### Módulo II

#### 1- Introducción



- Obtención de conocimiento a partir de los datos. El concepto de patrón. El proceso KDD. Fases del proceso de extracción del conocimiento. La Minería de Datos como fase del proceso KDD. Relación con otras disciplinas.
- 2- Recuperación de información vs recuperación de datos. Proceso de recuperación de información.
  - 3- Preparación de Datos. Metadatos. Análisis de la información de entrada. Construcción y análisis de representaciones gráficas. Limpieza y transformación. Transformación y creación de atributos. Discretización y Numerización, Normalización de rango, escalado y centrado. Exploración mediante visualización y selección de datos.
  - 4- Técnicas de Minería de Datos. Extracción de Patrones. Introducción. Tareas y Métodos. Tareas predictivas y descriptivas. Aprendizaje supervisado y aprendizaje no supervisado. La Minería de Datos y el aprendizaje inductivo. Comparación de las técnicas de Minería de Datos.
  - 5- Árboles de decisión. Métricas de selección de atributos. Entropía. Ganancia de Información. Tasa de Ganancia. Índice Gini. Poda y Sobreajuste. Algoritmos Id3, C4.5 y Random Forest. Construcción de árboles para grandes volúmenes de datos.
  - 6- Reglas de clasificación. Partición vs cobertura. Métodos ZeroR, OneR, PRISM y PART. Métricas de una regla: soporte, cobertura, confianza, interés y convicción.
  - 7- Reglas de asociación. Calidad de las reglas. Algoritmo A priori. Concepto de ítem frecuente. Mejoras del algoritmo a priori: FP-Growth y FP-Tree.
  - 8- Técnicas de Agrupamiento. Métricas de calidad del agrupamiento. Tipos de agrupamiento: Jerárquico, particivo y probabilista. Medidas de distancia y de conectividad. Proceso de agrupamiento. Clustering particivo.
  - 9- Redes Neuronales Feedforward. Descripción de la arquitectura. Resolución de problemas de clasificación y predicción.

### III. MODALIDAD Y RECURSOS/MATERIALES DE ESTUDIO

En este curso se combinan los encuentros sincrónicos con actividades mediadas a través del entorno virtual de enseñanza y aprendizaje que propone el SIED. Los encuentros son teórico – prácticos, donde se presentan los ejes temáticos, y luego se abordan ejemplos prácticos, para que los alumnos reflexionen y analicen sobre los casos presentados.

Los encuentros son presenciales y/o por videoconferencia, y para ello se utiliza la herramienta de VC del SIED. El encuentro final del curso es presencial para la presentación para la evaluación sumativa e integración de los temas abordados.



Las actividades que se realizan en el EVEA se orientan a la entrega de ejercicios prácticos con las herramientas utilizadas para el abordaje de los temas. Posteriormente, los alumnos realizan autoevaluaciones en donde se propone el análisis y fundamentación de casos y las técnicas que se aplicarían para su resolución.

Como materiales de estudio, se cuenta con:

- Presentaciones multimedia desarrolladas ad-hoc que introducen los diferentes ejes temáticos.
- Ejemplos que muestran ejercicios prácticos
- Píldoras formativas con la explicación de algunas temas
- Material de lectura para investigar y estudiar conceptos abordados en los encuentros
- Enlaces a artículos de actualidad de repositorios reconocidos en el área

También se presentan herramientas con las que se llevan adelante las prácticas, que se introducen durante los encuentros sincrónicos.

#### **IV. ACTIVIDADES PLANIFICADAS PARA LA APROPIACIÓN DE LOS SABERES Y LA EVALUACIÓN**

Las actividades que se proponen para este curso son:

- Encuentros vía videoconferencia para introducir temas y/o para realizar tutoría sobre ejercicios prácticos que deben realizar los alumnos
- Lectura de artículos de actualidad para conocer aplicación de los temas vistos en diferentes situaciones educativas
- Realización de ejercicios donde aplican técnicas de modelado (diagramas de entidad relación), consultas SQL, y técnicas de minería de datos usando las herramientas propuestas en el curso.
- Análisis de modelos de entidad relación para detectar posibles errores o inconsistencias y debatir sus mejoras.
- Realización de una evaluación que integre los temas de la materia. Se realiza una primera evaluación parcial que se realiza al finalizar el módulo I y una segunda que se realiza al finalizar el módulo II.

#### **V. BIBLIOGRAFÍA Y OTRO MATERIAL DE ESTUDIO**

##### **MÓDULO I**

**Diseño Conceptual de Bases de Datos. Un enfoque de entidades-interrelaciones**".Batini Ceri Navathe. Editorial Addison-Wesley/ Díaz de Santos.1994



**Conceptos de Bases de datos.** Abraham Silbrchatz, Henry Korth, S. Sudarshan.  
Editorial McGraw-Hill. 2010.

**Fundamentos de Bases de Datos.** Abraham Silberschatz, Henry F. Korth, S. Sudarshan.  
McGraw Hill. 2006.

**Fundamentos de Sistemas de Bases de Datos.** Ramez Elmasri, Shamkant Navathe.  
Editorial Pearson Addison Wesley. 2008

**Introducción a los sistemas de Bases de Datos.** C. J. Date. Pearson Educación, 2001

**Introducción a las Bases de Datos. Fundamentos y Diseño.** R. Bertone y P. Thomas.  
Pearson Education. 2011.

**Sistemas de Bases de Datos. Conceptos Fundamentales.** Ramez Elmasri, Shamkant Navathe.  
Editorial Addison-Wesley Iberoamericana. 2000.

## MÓDULO II

**Introducción a la Minería de Datos.** Hernández Orallo, Ramírez Quintana, Ferri Ramírez.  
Prentice Hall. 2004. ISBN84-205-4091-9.

**Data Mining: Practical Machine Learning Tools and Techniques.** Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal. 4ta. Edición. Editor Morgan Kaufmann, 2016.  
ISBN 0128043571, 9780128043578

**Data Mining and Learning Analytics: Applications in Educational Research.** Editor(s): Samira ElAtia, Donald Ipperciel, Osmar R. Zaiane. Series: Wiley Series on Methods and Applications in Data Mining. Publisher: Wiley, Year: 2016. ISBN 9781118998236.

**Handbook of Educational Data Mining.** Author(s): Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, Ryan S.J.d. Baker. Series: Chapman & Hall/CRC data mining and knowledge discovery series. Publisher: CRC Press, Year: 2010. ISBN: 1439804575, 978-1-4398-0457-5, 978-1-4398-0458-2

**Educational data mining: a review of the state of the art.** Authors: Cristóbal Romero, Sebastián Ventura. Journal IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews archive. Volume 40 Issue 6, November 2010. Pages 601-618



La bibliografía se complementa además material ad-hoc desarrollado por los docentes (presentaciones, ejercicios, ejemplos, píldoras formativas, etc.)

Se presentan además, de acuerdo a las tendencias de investigación, artículos de repositorios de referencia como IEEE, ACM, Springer y/o tesis de la propia maestría en el área. Por ejemplo:

- “Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios”. Formia Sonia. Dirigida por Laura Lanzarini. 2014
- “Espacios Virtuales para la formación de investigadores: elementos de análisis y diseño”. José Darío Rodríguez. Dirigida por Rodolfo Bertone. 2012
- “Contribución a la formación de docentes de las tecnologías informáticas. Recuperación y registro de estrategias de enseñanza.”. Héctor Raúl Gonzalez. Dirigido por Rodolfo Bertone. 2012.