

Nombre de la asignatura: **BASES DE DATOS Y BASES DE CONOCIMIENTO**

Línea de Trabajo: **Ingeniería de Software/ Inteligencia Artificial**

Tiempo de dedicación del estudiante a las actividades de:

| | | | | |
|-----|-----|-----|---------------|----------|
| DOC | TIS | TPS | Horas Totales | Créditos |
| 48 | 20 | 100 | 168 | 6 |

DOC: Docencia; **TIS:** Trabajo independiente significativo; **TPS:** Trabajo profesional supervisado

1. Historial de la asignatura

| Fecha revisión/actualización | Participantes | Observaciones, cambios o justificación |
|------------------------------|--|--|
| 21/Noviembre/2011 | Dr. Claudio Castellanos Sánchez MSI. María Gabriela Salazar de León | Versión preliminar |

2. Pre-requisitos y Co-requisitos

Pre-requisito asignatura Base de Datos.

3. Objetivo de la asignatura

Que el estudiante modele e implemente aplicaciones relacionadas con bases de conocimiento, que permitan la recolección, organización y recuperación computarizada de conocimiento.

4. Aportación al perfil del graduado

El alumno será capaz de:

- Modelar e implantar un sistema de almacenamiento de datos.
- Implantar un sistema de almacenamiento de datos para la extracción de conocimiento aplicando técnicas básicas de minería de datos.
- Recolectar y recuperar información en la web para tratarla y extraer información útil a un sistema de monitoreo estratégico.

5. Contenido temático

| Unidad | Temas | Subtemas |
|--------|-------------------------|---|
| 1 | Almacenamiento de datos | 1.1 Conceptos e historia del almacenamiento de datos. 1.2 Almacenamiento de datos versión 2.0 1.3 Ciclo de vida Kimball 1.4 Construyendo un almacén de datos 1.4.1 Correspondencia dimensional en el ciclo de vida del negocio 1.4.2 Definición de requerimiento del negocio 1.4.3 Ciclo de vida tecnológico 1.4.4 Diseño y arquitectura técnica 1.4.5 Modelado dimensional 1.4.6 Diseño y desarrollo de la preparación de datos 1.4.7 Expansión 1.4.8 Mantenimiento y crecimiento 1.5 Metodología para el registro seguro a distancia de datos |
| 2 | Minería de datos | 2.1 Introducción general a la minería de datos 2.2 Entrada: Conceptos, instancias y atributos 2.3 Salida: Representación del conocimiento |

| | | |
|---|-----------------------------------|---|
| | | 2.3.1 Tablas 2.3.2 Modelos lineales 2.3.3 Árboles 2.3.4 Reglas 2.3.5 Categorías 2.4 Algoritmos: los métodos básicos 2.4.1 Modelado estadístico 2.4.2 Construcción de árboles de decisión 2.4.3 Construcción de reglas 2.4.4 Modelos lineales 2.4.5 Categorización 2.4.6 Aprendizaje multi-instancia 2.5 Credibilidad: Evaluando lo que se ha aprendido 2.5.1 Entrenamiento y prueba 2.5.2 Desempeño predictivo 2.5.3 Validación cruzada 2.5.4 Contabilizando el costo |
| 3 | Minería en la Web | 3.1 Recuperación de información y búsqueda web 3.2 Texto y pre-procesamiento de páginas web 3.3 Índice invertido y su compresión 3.4 Indexado semántico latente 3.5 Búsquedas web 3.6 Meta-búsquedas 3.7 Basura en la web 3.8 Análisis de redes sociales 3.9 Arañas en la web 3.10 Extracción de datos estructurados 3.11 Integración de información 3.12 Minería de la opinión y análisis de sentimientos |
| 4 | Sistemas de monitoreo estratégico | 4.1 Principios 4.2 Algoritmos básicos 4.3 Aplicación |

6. Metodología de desarrollo del curso

Clases impartidas por el profesor, trabajo colaborativo y de integración.

7. Sugerencias de evaluación

Se recomienda la siguiente ponderación:

| | |
|----------------------------------|-----|
| Examen de un tercio de término | 15% |
| Examen de dos tercios de término | 15% |
| Examen final | 20% |
| Tareas | 20% |
| Proyecto Final | 30% |

8. Bibliografía y software de apoyo

Bibliografía

- "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", Second Edition, Ralph Kimball, Margy Ross, 464 pages, Wiley; 2 edition (April 26, 2002), ISBN-13: 978-0471200246

- “Kimball's Data Warehouse Toolkit Classics: The Data Warehouse Toolkit”, 2nd Edition; “The Data Warehouse Lifecycle”, 2nd Edition; “The Data Warehouse ETL Toolk”, Ralph Kimball, Margy Ross, Bob Becker, Joy Mundy, Warren Thornthwaite, 1628 pages, Wiley; 1 edition (April 6, 2009), ISBN-13: 978-0470479575
- “The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence”, Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, 744 pages, Wiley; 1 edition (February 8, 2010), ISBN-13: 978-0470563106
- “The Business of Data Vault Modeling”, Daniel Linstedt, Kent Graziano, Hans Hultgren, 92 pages, lulu.com (November 19, 2010), ISBN-13: 978-1435719149
- “Data Warehousing: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors”, Kevin Roebuck, 94 pages, Tebbo (June 23, 2011), ISBN-13: 978-1743047040
- “DW 2.0: The Architecture for the Next Generation of Data Warehousing”, William H. Inmon, Derek Strauss, Genia Neushloss, 400 pages, Morgan Kaufmann; 1 edition (July 9, 2008), ISBN-13: 978-0123743190
- “Data Warehouse Implementations: Critical Implementation Factors Study”, Joe Ganczarski, 136 pages, VDM Verlag (July 29, 2009), ISBN-13: 978-3639185898
- “Data Mining: Practical Machine Learning Tools and Techniques”, Ian H. Witten, Eibe Frank, Mark A. Hall, Third Edition, 664 pages, Morgan Kaufmann; 3 edition (January 20, 2011), ISBN-13: 978-0123748560
- “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data”, Bing Liu, 642 pages, Springer; 2nd Edition. edition (July 1, 2011), ISBN-13: 978-3642194597
- “Mining the Web: Discovering Knowledge from Hypertext Data”, Soumen Chakrabarti, 344 pages, Morgan Kaufmann; 1 edition (October 23, 2002), ISBN-13: 978-1558607545
- “Handbook of Statistical Analysis and Data Mining Applications”, Robert Nisbet, John Elder IV, Gary Miner, 864 pages, Academic Press; 1 edition (June 5, 2009), ISBN-13: 978-0123747655
- “Mathematical Methods for Knowledge Discovery and Data Mining”, Giovanni Felici, Carlo Vercellis, 371 pages, Idea Group Reference (October 4, 2007), ISBN-13: 978-1599045283
- “Applications of Data Mining in E-Business and Finance”, Z.-H. Zhou, 120 pages, IOS Press (August 15, 2008), ISBN-13: 978-1586038908
- “Intelligent Data Warehousing: From Data Preparation to Data Mining”, Zhengxin Chen, 256 pages, CRC Press; 1 edition (December 13, 2001), ISBN-13: 978-0849312045
- “Survey of Text Mining I: Clustering, Classification, and Retrieval”, Michael W. Berry, 261 pages, Springer; 1 edition (September 9, 2003), ISBN-13: 978-0387955636
- “Data Mining and Decision Support: Integration and Collaboration”, Dunja Mladenic, Nada Lavrac, Marko Bohanec, Steve Moyle, 304 pages, Springer; 1 edition (September 30, 2003), ISBN-13: 978-1402073885

Software de apoyo:

- Weka, <http://www.weka.net.nz/> o en <http://www.cs.waikato.ac.nz/ml/weka>

9. Actividades propuestas

| Unidad | Actividad |
|--------|---|
| 1 | Realizar un Mapa Conceptual sobre los conceptos e historia del almacenamiento de datos (Data Warehouse) |
| 1 | Resolver el caso práctico propuesto siguiendo la Metodología del Ciclo de Vida Kimball para construir un DATA WAREHOUSE |
| 2 | Investigar ejemplos en los que se puede utilizar la Minería de Datos, exponerlos y discutirlos en el aula |
| 2 | Establecer la estrategia de Minería de Datos para extraer y presentar la información requerida del Data Warehouse construido en la unidad 1. |
| 3 | Establecer la estrategia de extracción y tratamiento de datos de una página web que tenga relación con el caso práctico que se utilizó en las dos unidades anteriores |
| 4 | Proponer un sistema de monitoreo estratégico para el caso práctico resuelto anteriormente |

10. Catedrático (s) responsable (s)



Dr. Claudio Castellanos Sánchez