

PROGRAMA DE CURSO

Código	Nombre			
IN 5526	WEB INTELLIGENCE			
Nombre en Inglés				
Web Intelligence				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3.0	1.5	5.5
Requisitos			Carácter del Curso	
IN3501 Tecnologías de Información y Comunicaciones para la Gestión IN3702 Investigación de Operaciones			Electivo de la carrera de Ingeniería Civil Industrial	
Resultados de Aprendizaje				
Al finalizar el curso: El alumno aprende a desarrollar aplicaciones de Web Intelligence que permitan extraer información y conocimiento desde los datos web, para luego ser usados en el desarrollo de nuevas tecnologías web para el futuro, como sitios web adaptativos. A través del material visto en clases y el proyecto de implementación se espera que el alumno adquiera los siguientes conocimientos y habilidades:				
• Conceptos clave de Web Intelligence y su importancia en los negocios • Metodología KDD para extracción de conocimiento desde datos • Algoritmos y modelos para extracción de patrones • Metodologías y algoritmos de Web Mining: (Algoritmos Page Rank de Google, Algoritmos de reconstrucción de sesiones , Algoritmos de minado de texto) • Comprensión y modelamiento del comportamiento de usuarios web, apoyado en técnicas de eye-tracking y neurociencia • Principales temas legales en relación a la privacidad de datos en la web				

Metodología Docente	Evaluación General
Este curso tiene una connotación teórico-práctica, es decir, durante las clases de cátedra se estudiará la teoría que sustenta el desarrollo de un proyecto de levantamiento e implementación de un proyecto de Web Mining desde el punto de vista del negocio como así también del punto de vista de la implementación del mismo. Luego se aplicará la teoría a la resolución de casos simples, que permitirán al alumno prepararse para desarrollar con éxito las tareas del curso. Tanto la teoría como el desarrollo de las tareas serán soportadas en las clases auxiliares.	El curso cuenta con dos actividades que deben ser aprobadas por separado: Controles y Tareas. Se realizarán dos controles (C_i) durante el curso, más un examen (NE). La nota de controles (NC) se calcula según: $NC = \frac{\sum_{i=1}^n C_i}{n} * 0.6 + NE * 0.4$ Se realizarán tres tareas aplicadas de implementación de aplicaciones alineadas con la teoría vista en cátedra, siendo NT el promedio de las notas de tareas. La nota final del curso NF=NC*0.5+NT*0.5.

UNIDADES TEMÁTICAS

Número	Nombre de la Unidad	Duración en Semanas
1	INTRODUCCIÓN A WEB INTELLIGENCE	2.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Conceptos Iniciales 2. KDD 3. Datos Web 4. Minería Web	El estudiante: Conoce los conceptos fundamentales en Web Intelligence y la principal metodología para la extracción de conocimiento a partir de datos: KDD (Knowledge Discovery in Data Bases) y conocer los principales problemas en Data Mining (clasificación, asociación, segmentación). Introducirse a la estructura y complejidad de los datos contenidos y generados en la Web y como se puede aplicar KDD sobre ellos.	5, 6, 7

Número	Nombre de la Unidad	Duración en Semanas
2	MODELOS DE DATA MINING: MACHINE LEARNING	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción al aprendizaje de máquinas 2. Redes Neuronales a. Perceptrón b. Perceptrón multicapa c. Redes de Kohonen (SOFMs) 3. K-means 4. Árboles de Decisión 5. Redes Bayesianas 6. K-Nearest Neighbor (KNN) 7. Máquinas de soporte vectorial (SVM)	El estudiante: Conoce el funcionamiento de las principales herramientas provistas por el aprendizaje de máquinas (algoritmos para clasificación, asociación, regresión, etc.) para solucionar los diversos problemas en Data Mining.	5, 6, 7

Número	Nombre de la Unidad		Duración en Semanas
3	MINERÍA DE USO WEB (WEB USAGE MINING)		2.0
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción 2. El proceso de reconstrucción de sesiones 3. Clusterización de sesiones 4. Clasificación del comportamiento del usuario en un sitio Web 5. Aplicación de reglas de asociación para descubrir patrones de navegación 6. Aplicación de patrones de secuencia para descubrir caminos comunes de acceso. 7. Una metodología para recomendaciones de navegación 8. Aplicaciones de WUM		El estudiante: Aprende a extraer patrones sobre los registros generados por el uso de los usuarios de un sitio Web y algunas aplicaciones que se construyen usando estos datos.	1, 2, 3

Número	Nombre de la Unidad		Duración en Semanas
4	MINERÍA DE CONTENIDO WEB (WEB CONTENT MINING)		2.0
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción 2. Minería de Texto (Text Mining) 3. Clasificación del texto contenido en una página Web. 4. Segmentación de grupos de páginas con contenido similar 5. WEBSOM 6. Sumarización automática del texto de una página Web 7. Extracción de componentes clave 8. Procesamiento de Lenguaje Natural (NLP) 9. Aplicaciones		El estudiante: Aprende a procesar el contenido de un sitio Web, principalmente el texto. En primer lugar, se espera que aprenda a manejar el texto contenido en un sitio de manera vectorial utilizando el enfoque de text mining clásico, así como también considerando nuevas técnicas de procesamiento de lenguaje natural para, por ejemplo, hacer minería de opiniones.	1, 2, 3, 7

Número	Nombre de la Unidad	Duración en Semanas
5	MINERÍA DE ESTRUCTURA WEB (WEB STRUCTURE MINING)	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción 2. El algoritmo HITS 3. El algoritmo Page Rank 4. Identificación de comunidades en la Web	El estudiante: Comprende la estructura de grafo de la Web y como identificar que páginas son relevantes dentro de un conjunto de páginas según su estructura de hipervínculos.	1, 3

Número	Nombre de la Unidad	Duración en Semanas
6	ENTENDIENDO EL COMPORTAMIENTO DEL USUARIO WEB	2.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Sistemas recomendadores en la web 2. Entendiendo las preferencias de los usuarios a través de Web Site Keyobjects y Eye Tracking 3. Aplicando neurociencia para modelar el comportamiento del usuario en la web 4. Web Intelligence y la Neurociencia	El estudiante: Combinando los datos y la información entregada por diversas técnicas de web mining, aprende a entender y modelar el comportamiento de los usuarios en la web, en vistas de generar aplicaciones que hagan la navegación más simple y rica. Se espera que el alumno consolide el aprendizaje adquirido y a construir y mantener sitios Web adaptativos.	2

Número	Nombre de la Unidad	Duración en Semanas
7	DATOS PERSONALES, PRIVACIDAD Y MINERÍA WEB	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Datos personales 2. Protección legal de datos personales y privacidad 3. Desafíos	El estudiante: Se familiariza con el impacto que las técnicas de minería web tienen en la conservación de la privacidad de los usuarios y que conozca los conceptos básicos de protección de estos derechos de privacidad.	9,10

Bibliografía General

1. J.D. Velásquez and L.C. Jain “Advanced Techniques in Web Intelligence part 1”, Springer, 2010.
2. J.D. Velásquez y L. Donoso “Tratamiento de Datos Personales en Internet”, Thomson Reuters, 2013
3. J.D. Velásquez, V. Palade and L.C. Jain “Advanced Techniques in Web Intelligence part 2”, Springer, 2013.
4. J.D. Velásquez and V. Palade “Adaptative Web Site”. IOS Press, Netherland, 2007.
5. C.D. Manning, P. Raghavan, H. Schutze, “Introduction to Information Retrieval”, Cambridge University Press 2008. <http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>
6. G. Myatt, “Making Sense of Data: A practical Guide to exploratory data analysis and data mining”. Wiley Interscience 2007.
7. S. Chakrabarti, “Mining The Web, Discovering Knowledge From HyperText Data”. Morgan Kaufmann Publisher 2003.
8. B. Liu, “Web Data Mining”, Springer 2011.
9. A. Scime, “Web Mining: Application and techniques”. IDEA Group Publishing 2005.
10. Tratamiento De Datos Personales En Internet, Juan Velásquez Silva, Lorena Donoso Abarca, Thomson, 2013
11. J.D. Velásquez. Web mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in web-based environments, Expert Systems with Applications, 40(13):5228–5239, 2013

Vigencia desde:	Primavera 2013
Elaborado por:	Juan D. Velásquez
Aprobado por:	Comisión de Docencia DII