

PROGRAMA DE CURSO

Código	Nombre			
IN 4522	WEB MINING			
Nombre en Inglés				
WEB MINING				
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	
IN55A/(IN3501,IN3702)			Electivo de la carrera Ingeniería Civil Industrial	
Resultados de Aprendizaje				
<p>Que el alumno aprenda a desarrollar aplicaciones de web mining que permitan extraer información y conocimiento desde la web data, para luego ser usado en la mejora continua de un sitio web institucional. A través del material visto en clases y el proyecto de implementación se espera que el alumno adquiera los siguientes conocimientos y habilidades:</p> <ul style="list-style-type: none"> • Aplicaciones de negocio de web mining • Metodología Vector Space Model • Algoritmos Page Rank de Google • Algoritmos de data mining aplicados a la web • Algoritmos de minado de texto 				

Metodología Docente	Evaluación General
<p>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</p>	<p>El curso cuenta con dos actividades que deben ser aprobadas por separado: Controles y Tareas.</p> <p>Se realizarán tres controles durante más un examen (NE). La nota de controles se calcula según:</p> $NC = \frac{\sum_{i=1}^n C_i}{n} * 0.6 + NE * 0.4$ <p>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</p> <p>La nota final del curso $NF = NC * 0.5 + NT * 0.5$.</p>

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	INTRODUCCIÓN	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. La Wolrd Wide Web 2. Negocios electrónicos (E-business) 3. Portales Web adaptativos	Entender el rol la Web en la economía digital. Comprender el rol de un portal Web dentro de un negocio y como los portales Web adaptativos pueden mejorar la experiencia del usuario.	

Número	Nombre de la Unidad	Duración en Semanas
2	DATOS DE LA WEB	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Operaciones en la Web 2. La información tras los clicks 2.1. Reconstrucción de sesiones 2.2. Identificación de sesiones reales 3. Información contenida en una página Web 3.1. Contenido de una página Web 3.2. Hipervínculos de una página Web	Que el alumno se introduzca a la estructura y complejidad de los datos contenidos y generados en la Web.	

Número	Nombre de la Unidad	Duración en Semanas
3	DESCUBRIMIENTO DE CONOCIMIENTO EN LOS DATOS DE LA WEB	6.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. El proceso KDD 2. Fuentes de Datos y Limpieza 3. Consolidación de datos y repositorios de información 4. Minería de datos y aprendizaje de máquinas 4.1. Reglas de asociación 4.2. Clasificación 4.3. Segmentación (clustering) 5. Algoritmos de minería de datos 5.1. Redes neuronales artificiales 5.2. Self-Organization Feature Maps 5.3. K-medias 5.4. Árboles de decisión	Que el alumno comprenda la metodología KDD (Knowledge Discovery in Data Bases) y su aplicabilidad de uso sobre los datos de la Web para la extracción de conocimiento. Además que sepa diferenciar las distintas herramientas provistas por la minería de datos y el aprendizaje de máquinas según el problema que se quiere solucionar (clasificación, asociación, segmentación).	

<p>5.5. Redes Bayesianas 5.6. K-Nearest Neighbor (KNN) 5.7. Máquinas de soporte vectorial 6. Extracción de conocimiento para la extracción de conocimiento 7. Validación del Conocimiento extraído 8. Minando la Web</p>		
--	--	--

Número	Nombre de la Unidad	Duración en Semanas
4	MINERÍA DE ESTRUCTURA WEB	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<p>1. Introducción 2. El algoritmo HITS 3. El algoritmo Page Rank 4. Identificación de comunidades en la Web</p>	<p>Se espera que alumno comprenda 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.</p>	2

Número	Nombre de la Unidad	Duración en Semanas
5	MINERÍA DE CONTENIDO WEB	2.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<p>1. Introducción 2. Clasificación del texto contenido en una página Web. 3. Segmentación de grupos de páginas con contenido similar 4. Aplicaciones 4.1. WEBSOM 4.2. Sumarización automática del texto de una página Web 4.3. Extracción de componentes clave de texto de una página Web.</p>	<p>Que el alumno aprenda a procesar el contenido de un sitio Web, principalmente el texto. Se espera que aprenda a manejar el texto contenido en un sitio de manera vectorial y que pueda aplicar técnicas de minería de datos sobre éste.</p>	

Número	Nombre de la Unidad	Duración en Semanas
6	MINERÍA DE USO WEB	2.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Métodos estadísticos 2. Segmentación de sesiones 3. Clasificación del comportamiento del usuario en un sitio Web 4. Aplicación de reglas de asociación para descubrir patrones de navegación 5. Aplicación de patrones de secuencia para descubrir caminos comunes de acceso.	Que el alumno aprenda a extraer patrones sobre los registros generados por el uso de los usuarios de un sitio Web.	5

Número	Nombre de la Unidad	Duración en Semanas
7	APLICACIONES	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. La propuesta de sitios Web adaptativos 2. Selección de datos de la Web 3. Extracción de información sobre datos de la Web 3.1. El modelo estrella para la construcción de repositorios de información Web 3.2. Proceso de reconstrucción de sesiones 3.3. Preprocesamiento del contenido de una página Web 4. Aplicación de técnicas de minería de datos 4.1. Análisis del comportamiento navegacional del usuario 4.2. Análisis de preferencias de usuario 5. Generación de recomendaciones con el contenido extraído 5.1. Recomendaciones en offline 5.2. Recomendaciones online 5.3. Validación de la efectividad de las recomendaciones 5.4. Almacenamiento del conocimiento extraído	Se espera que el alumno consolide el aprendizaje adquirido en aplicaciones reales de minería de la Web. Que aprenda además a construir y mantener sitios Web adaptativos.	1,5

Bibliografía General

1. J.D. Velásquez and L.C. Jain "Advanced Techniques in Web Intelligence part 1", Springer, 2010.
2. J.D. Velásquez and V. Palade "Adaptative Web Site". IOS Press, Netherland, 2007.
3. 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>
4. G. Myatt, "Making Sense of Data: A practical Guide to exploratory data análisis and data mining". Wiley Interscience 2007.
5. S. Chakrabarti, "Mining The Web, Discovering Knowlege From HyperText Data". Morgan Kaufmann Publisher 2003.
6. A. Scime, "Web Mining: Application and techniques.". IDEA Group Publishing 2005.

Vigencia desde:	Primavera 2010
Elaborado por:	Juan D. Velásquez