

1. Datos Generales de la asignatura

Nombre de la asignatura:	Análisis de Grandes Volúmenes de Datos
Clave de la asignatura:	ARW -1806
SATCA¹:	2 – 3 – 5
Carrera:	Ingeniería en Sistemas Computacionales Ingeniería en Tecnologías de la Información y Comunicaciones Ingeniería Informática

2. Presentación

Caracterización de la asignatura
<p>Esta asignatura aporta al perfil del egresado la capacidad para procesar, interpretar y graficar datos previamente obtenidos.</p> <p>El estudiante podrá aplicar sus conocimientos y habilidades para análisis de grandes cantidades de datos.</p> <p>Esta materia sirve de integración de varios conocimientos que los alumnos desarrollaron en materias anteriormente tomadas.</p>
Intención didáctica
<p>El temario está organizado en 5 unidades.</p> <p>La primera unidad retoma los conceptos de motores de búsqueda y extracción de datos que se vieron en la materia previa y que servirán para iniciar el curso.</p> <p>La segunda unidad se concentra en analizar los modelos de los datos estructurados y no estructurados que se encuentran en la red.</p> <p>La tercera unidad analiza los datos obtenidos previamente utilizando técnicas y métodos estadísticos y probabilísticos.</p> <p>La cuarta unidad analiza datos utilizando Aprendizaje Máquina y analiza el lenguaje natural.</p> <p>La quinta unidad abordará los métodos y técnicas para visualizar gran cantidad de datos en los</p>

dispositivos de los cliente finales, como son las aplicaciones de escritorio, las aplicaciones Web y los dispositivos móviles.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico de Morelia Fecha: Agosto de 2018	Miembros de la línea de investigación de “Tecnologías Web” de la Academia de Ingeniería de Sistemas Computacionales. Ing. Juan Jesús Ruiz Lagunas Dr. Heberto Ferreira Medina M.C. Rogelio Ferreira Escutia Dr. Anastacio Antolino Hernández M.I. Adrián Núñez Vieyra Ing. Kenia Aline Ayala Robles M.C. Juan Carlos Olivares Rojas	Análisis y diseño curricular de los módulos de especialidad las carreras del departamento de Sistemas y Computación. Plan de estudios de la especialidad 2019

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura

- Desarrollar aplicaciones para la búsqueda, extracción, procesamiento, interpretación y visualización de grandes cantidades de datos.

5. Competencias previas

- Conocer las arquitecturas distribuidas.
- Conocer e implementar motores de búsqueda en la Web.
- Almacenamiento de datos local y distribuida.

6. Temario

No.	Temas	Subtemas
1	Búsqueda y recolección de información	1.1 Motores de búsqueda. 1.2 Extracción de datos usando API's
2	Modelos de datos estructurados y no estructurados	2.1 Tipos de datos 2.2 Datos estructurados 2.3 Datos no estructurados
3	Métodos Estadísticos y Probabilísticos	3.1 Métodos estadísticos para análisis de datos 3.2 Métodos probabilísticos para análisis de datos
4	Aprendizaje Máquina y Lenguaje Natural	4.1 Machine Learning 4.2 Deep Learning 4.3 Análisis de lenguaje natural
5	Visualización	5.1 Tipos de gráficos 5.2 Herramientas de graficación. 5.3 Graficación de gran cantidad de información. 5.4 Impresión de gráficas en aplicaciones de escritorio, Web y móviles.

7. Actividades de aprendizaje de los temas

Nombre del tema: Búsqueda y recolección de información	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Comprender los procesos para búsqueda y recolección de información. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo 	<ul style="list-style-type: none"> • Conocer y construir software para motores de búsqueda. • Conocer e implementar la extracción de datos utilizando API's de terceros.

<p>interdisciplinario.</p> <ul style="list-style-type: none"> • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	
<p>Nombre del tema: Modelos de datos estructurados y no estructurados</p>	
<p>Competencias</p>	<p>Actividades de aprendizaje</p>
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer los modelos de datos estructurados y no estructurados. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Conocer los tipos de datos estructurados. • Conocer los tipos de datos no estructurados. • Conocer la forma de extraer los datos estructurados y no estructurados de la red.
<p>Nombre del tema: Métodos Estadísticos y Probabilísticos</p>	
<p>Competencias</p>	<p>Actividades de aprendizaje</p>
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer los métodos estadísticos y probabilísticos. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los 	<ul style="list-style-type: none"> • Conocer y aplicar los métodos estadísticos para análisis de datos. • Conocer y aplicar los métodos probabilísticos para análisis de datos.

conocimientos en la práctica.	
Nombre del tema: Aprendizaje Máquina y Lenguaje Natural	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer las técnicas de Aprendizaje Máquina y el análisis del Lenguaje Natural. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Implementar soluciones de software para el Aprendizaje Máquina en la solución de problemas. • Implementar soluciones de software para el análisis del lenguaje natural.
Nombre del tema: Visualización	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer las herramientas y técnicas para visualizar información. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Conocer los tipos de gráficas que se utilizan. • Usar lenguajes y herramientas para la graficación de información. • Implementar la graficación de grandes volúmenes de información. • Implementar la graficación en dispositivos de escritorio, Web y móviles.

8. Práctica(s)

1. Implementar un motor de búsqueda.
2. Extraer datos de la red usando API's.
3. Almacenar datos estructurados en bases de datos distribuidas.
4. Almacenar datos no estructurados en bases de datos distribuidas.
5. Predecir tendencias usando métodos estadísticos.
6. Predecir tendencias usando métodos probabilísticos.
7. Aplicar Machine Learning para aprendizaje basado en información.
8. Aplicar Deep Learning para aprendizaje basado en información.
9. Usar herramientas para análisis de lenguaje natural.
10. Graficar datos usando aplicaciones de escritorio.
11. Graficar datos usando aplicaciones Web.
12. Graficar datos usando aplicaciones móviles.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

- **Fundamentación:** marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.
- **Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso de intervención empresarial, social o comunitario, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.
- **Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.
- **Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesión, social e investigativo, ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de "evaluación para la mejora continua", la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

La evaluación debe ser continua y permanente por lo que se debe considerar el desempeño en cada una de las actividades de aprendizaje, haciendo especial énfasis en:

- Participación en clase.
- Ejercicios realizados en clase.
- Información obtenida durante las búsquedas encomendadas.
- Lectura y análisis de textos.
- Autoevaluación, coevaluación y evaluación de las actividades.
- Revisión periódica del avance del proyecto (o proyectos) de la asignatura.

11. Fuentes de información

- **Cómo programar Internet & World Wide Web**
Paul Deitel, Harbey Deitel y Abbey Deitel
Quinta Edición, Pearson Educación de México, Naucalpan de Juárez, Estado de México, septiembre 2013
ISBN: 978-607-32-2290-7
Páginas: 663
- **Sistemas Distribuidos - Conceptos y Diseño**
George Coulouris, Jean Dollimore y Tim Kindberg
Tercera Edición Addison Wesley Pearson Educación, Madrid España 2001
ISBN: 84-7829-049-4
Páginas: 726
- **Principles of Distributed Database Systems**
M. Tamer Ozsú and Patrick Valduriez
Second Edition, Prentice Hall, USA 1991
ISBN: 0-13-659707-6
Páginas: 666
- **Introducción a la Minería de Datos**
José Hernández Orallo, María José Ramírez Quintana y César Ferri Ramírez
Primera Edición, Pearson Prentice Hall, Madrid España 2004
ISBN: 978-84-205-4091-7
Páginas: 656
- **MPI The Complete Reference - Volume 1, The MPI Core**
Marc Smir, Stece Otto, Steven Huss Lederman, David Walker, Jack Dongarra,
Second Edition MIT Press
Cambridge Massachusetts, London England 1998

- **Small Data**
Joan Moisés González Martínez y Montse Rodríguez Fígols
Primera Edición en México, Editorial Planeta Mexicana, Ciudad de México agosto 2017
ISBN: 978-607-07-4223-1
Páginas: 158