

UNIVERSIDAD NACIONAL DEL SUR  
BAHIA BLANCA

1  
4

DEPARTAMENTO DE CIENCIAS E INGENIERÍA DE LA COMPUTACION

MINERÍA DE LA WEB

CODIGO: 7815

AREA N°: VI

**PROFESOR RESPONSABLE:**

Dra. Ana G. Maguitman – Profesor Adjunto con Dedicación Exclusiva

<b>CARGA HORARIA</b>	<b>Teoría 64</b>	<b>Práctica 64</b>	<b>Laboratorio</b>	<b>CANTIDAD DE SEMANAS</b>	<b>16</b>
----------------------	----------------------	------------------------	--------------------	----------------------------	-----------

**CORRELATIVAS**

<b>PARA CURSAR LA MATERIA</b>		<b>PARA APROBAR LA MATERIA</b>	
<b>APROBADAS</b>	<b>CURSADAS</b>	<b>APROBADAS</b>	<b>CURSADAS</b>
- Matemática Discreta - Programación Orientada a Objetos	- Teoría y Diseño de Bases de Datos	- Matemática Discreta - Programación Orientada a Objetos - Teoría y Diseño de Bases de Datos	

**DESCRIPCIÓN**

El objetivo de este curso es proveer al alumno con una introducción a los conceptos básicos de la minería de la Web. En la primera parte del curso se estudiarán los conceptos fundamentales de la recuperación de la información clásica para pasar luego al análisis de técnicas de crawling, indexación y búsqueda en la Web.

La segunda parte del curso estará orientada al estudio de aplicaciones diversas y temas especiales. Se discutirán aplicaciones de algoritmos de aprendizaje en la Web, sistemas de personalización y recomendación, técnicas de minería de datos en casos prácticos y sistemas para el soporte de toma de decisiones basados en información extraída de la Web.

Este curso dará la oportunidad a los estudiantes de explorar tecnologías actuales de minería de la Web. Asimismo el curso permitirá que el alumno gane experiencia en la presentación y discusión crítica de material técnico, el desarrollo de un proyecto y la escritura de un reporte técnico.

**METODOLOGÍA DE ENSEÑANZA**

El dictado de clases teóricas semanales se complementa con clases prácticas y discusiones grupales e individuales sobre proyectos.

Las clases teóricas se dictan utilizando diapositivas. La formación teórica está orientada a que el alumno se familiarice con los conceptos más relevantes del área de recuperación de la información tales como modelo vectorial, generación de índices, técnicas de crawling, algoritmos de clustering y algoritmos de clasificación.

La formación práctica está orientada a la aplicación de los conceptos teóricos en la

resolución de problemas concretos. Dichos problemas se postulan en ejercicios que conforman las guías de trabajos prácticos. Por otra parte los alumnos desarrollan un proyecto final aplicando diversas herramientas estudiadas durante el cursado de la materia.

### **MECANISMO DE EVALUACIÓN**

Para cursar la materia el alumno debe desarrollar y aprobar dos trabajos prácticos. Debe además presentar tres informes de avance del proyecto desarrollado durante el cursado y realizar una presentación oral del mismo al final del cuatrimestre.

Existe la posibilidad de promocionar la materia. Para la promoción se requiere concluir el proyecto, presentar un informe final escrito del mismo y aprobar un examen oral sobre los temas teóricos vistos durante la materia. Aquellos alumnos que cursen la materia pero no accedan a la promoción, una vez presentado el informe final del proyecto, deberán rendir un examen final regular consistente en una entrevista oral.

### **PROGRAMA SINTÉTICO**

1. Introducción a los conceptos básicos de la minería de la Web.
2. Recuperación de la Información.
3. Web Crawlers.
4. Buscadores.
5. Clustering.
6. Clasificación.
7. Análisis de Redes Sociales.
8. Búsqueda Basada en Contexto.
9. La Web Semántica.
10. La Web Invisible.

### **PROGRAMA ANALÍTICO**

1. Introducción a los conceptos básicos de la minería de la Web
  - Historia de la Web
  - Recuperación de la información
  - Caracterización formal de modelos de recuperación de la información
  - Métodos clásicos para la evaluación de sistemas de recuperación de la información
  - Técnicas de crawling, indexación y búsqueda en la Web
  - Directorios de paginas Web
  - Análisis de contenido
  - Análisis de enlaces
  - Análisis de uso

DEPARTAMENTO DE CIENCIAS E INGENIERÍA DE LA COMPUTACION

MINERÍA DE LA WEB

CODIGO: 7815

AREA N°: VI

- Datos estructurados y no estructurados
- 2. Técnicas de Crawling
  - Estándares y protocolos
  - Herramientas de crawling
  - Crawlers generales y por tema
  - Ingeniería de crawlers de gran escala
- 3. Técnicas de Indexación y Búsqueda
  - Construcción de índices
  - Creación de consultas
  - Optimización de consultas
  - Expansión de consultas
  - Técnicas de ranking
  - Búsquedas incrementales y basadas en contexto
- 4. Aplicaciones y Temas Especiales
  - Aprendizaje mecánico basado en la Web
  - La Web invisible
  - Toma de decisiones basadas en información extraída de la Web
  - Agentes inteligentes en la Web
  - Personalización y sistemas de recomendación
  - Análisis de redes sociales y comunidades en la Web
  - La Web semántica
  - Seguridad y privacidad
  - Medidas y análisis del grafo de la Web
  - Interacción y visualización

Bibliografía Básica

1. Mining of Massive Datasets. 2nd Edition. Jure Leskovec, Anand Rajaraman, Jeffery David Ullman. Cambridge University Press, 2014.
2. Information Retrieval: Implementing and Evaluating Search Engines. Stefan Büttcher, Charles L. A. Clarke, and Gordon V. Cormack. MIT Press, 2010.
3. Modern Information Retrieval. The concepts and technology behind search. Second Edition. R. Baeza-yates and B. Ribeiro-Neto, Addison-Wesley, New York, 2011.

Bibliografía Adicional

1. Collective Intelligence in Action. Satnam Alag. Manning Publications Co. 2008.
- Algorithms of the Intelligent Web Haralambos Marmanis and Dmitry Babenko. Manning Publications Co. 2009.
2. Search Engines: Information Retrieval in Practice. Bruce Croft, Donald Metzler, Trevor Strohman. Addison Wesley. 2010.
3. Introduction to information retrieval. By Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze Cambridge Press. 2009.

**UNIVERSIDAD NACIONAL DEL SUR  
BAHIA BLANCA**

4  
4

**DEPARTAMENTO DE CIENCIAS E INGENIERÍA DE LA COMPUTACION**

**MINERÍA DE LA WEB**

**CODIGO: 7815**

**AREA N°: VI**

4. Programming Collective Intelligence. Building Smart Web 2.0 Applications. Toby Segaran. O'Reilly. 2007.
5. TREC: Experiment and Evaluation in Information Retrieval. Ellen M. Voorhees and Donna K. Harman, editors. Cambridge, Massachusetts: MIT Press, 2005.
6. Mining the Web: Discovering Knowledge from Hypertext Data, Soumen Charkabarti, Morgan Kaufmann, 2002.
7. Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten, Eibe FrankMorgan Kaufmann, 2005.
8. Artículos disponibles en la Web, en particular aquellos publicados en las siguientes conferencias: Internacional World Wide Web Conference, ACM WSDM, ACM HT, ACM SIGIR, ACM KDD y ACM CIKM.

**AÑO**

**2017**

**FIRMA PROFESOR RESPONSABLE**

**VISADO**

**COORDINADOR AREA**

**SECRETARIO ACADÉMICO**

**DIRECTOR  
DEPARTAMENTO**