

**Universidad Autónoma del Estado de México**  
**Facultad de Ciencias Políticas y Sociales**  
**Licenciatura en Gestión de la Información en Redes Sociales,**  
**2017**



Programa de estudio de la unidad de aprendizaje:

**Web Crawler**

**Elaboró:** Ing. Francisco Ignacio Chávez Castañeda

<b>Fecha de aprobación:</b>	H. Consejo Académico	H. Consejo de Gobierno
	Acta 729	Acta 811
	28 de enero 2021	29 de enero 2021

---

**Facultad de Ciencias Políticas y Sociales**





## Índice

	Pág.
I. Datos de identificación.	3
II. Presentación del programa de estudios.	4
III. Ubicación de la unidad de aprendizaje en el mapa curricular.	4
IV. Objetivos de la formación profesional.	5
V. Objetivos de la unidad de aprendizaje.	7
VI. Contenidos de la unidad de aprendizaje y su organización.	7
VII. Acervo bibliográfico.	8
VIII. Mapa curricular.	9





### I. Datos de identificación

Espacio educativo donde se imparte

Licenciatura

Unidad de aprendizaje  Clave

Carga académica	<input type="text" value="1"/>	<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="5"/>
	Horas teóricas	Horas prácticas	Total de horas	Créditos

Período escolar en que se ubica

<input type="text" value="1"/>	<input type="text" value="2"/>	<input type="text" value="3"/>	<input type="text" value="4"/>	<input type="text" value="5"/>	<input type="text" value="6"/>	<input type="text" value="7"/>	<input checked="" type="text" value="8"/>	<input type="text" value="9"/>
--------------------------------	--------------------------------	--------------------------------	--------------------------------	--------------------------------	--------------------------------	--------------------------------	---	--------------------------------

Seriación

UA Antecedente UA Consecuente

### Tipo de Unidad de Aprendizaje

Curso	<input type="checkbox"/>	Curso taller	<input type="checkbox"/>
Seminario	<input type="checkbox"/>	Taller	<input checked="" type="checkbox"/>
Laboratorio	<input type="checkbox"/>	Práctica profesional	<input type="checkbox"/>
Otro tipo (especificar)	<input type="text"/>		

### Modalidad educativa.

Escolarizada. Sistema rígido	<input type="checkbox"/>	No escolarizada. Sistema virtual	<input type="checkbox"/>
Escolarizada. Sistema flexible	<input type="checkbox"/>	No escolarizada. Sistema a distancia	<input type="checkbox"/>
No escolarizada. Sistema abierto	<input type="checkbox"/>	Mixta	<input checked="" type="checkbox"/>

### Formación común



No presenta



## II. Presentación del programa de estudios

Los datos son representaciones simbólicas que utilizamos para describir entidades. Estas representaciones simbólicas en sí mismas no necesariamente cambian el nivel de conocimiento de quien las consulta, no son información.

Necesitamos procesarlas y relacionarlas para poder construir información a partir de ellas. Esta necesidad de procesar datos se intensificó con la aparición de la *world wide web* (www) de Internet. Esta red de alcance mundial, proporciona una cantidad inmensa de datos, que deben ser procesados para generar información.

Los datos contenidos en los documentos de los servidores de la www pueden ser procesados por programas conocidos como *crawlers* (rastreadores) que son capaces de navegar entre servidores, leer los documentos en los mismos, analizar los documentos encontrados y generar bases de datos con la información recolectada.

Los *crawlers* son herramientas utilizadas, principalmente, por los motores de búsqueda para generar índices para saber en qué documento de qué servidor se encuentra un determinado conjunto de datos. Estos índices pueden utilizarse para generar un catálogo de importancia de los servidores al que conocemos como posicionamiento web. Otros usos de la información obtenida por los *crawlers* son: comparar precios en sitios de venta de productos, encontrar datos específicos de empresas, recopilar datos relativos a las visitas de sitios web, etc.

El presente programa proporcionará a los estudiantes de la Licenciatura en Gestión de Redes sociales los principales elementos de la creación de programas rastreadores para obtener información que les permita generar ventajas competitivas para empresas con sistemas de ventas en línea o para identificar de forma automática necesidades de la población que permitan diseñar políticas públicas significativas.

## III. Ubicación de la unidad de aprendizaje en el mapa curricular

Núcleo de formación:	Integral
Área Curricular:	Ingeniería y tecnología
Carácter de la UA:	Optativa





#### IV. Objetivos de la formación profesional

##### Objetivos del programa educativo:

La Licenciatura en Gestión de la Información en Redes Sociales forma profesionales que contribuyen al progreso social, económico y cultural del país, a través de los siguientes objetivos:

##### Generales

- Ejercer el diálogo y el respeto como principios de la convivencia con sus semejantes, y apertura al mundo.
- Reconocer la diversidad cultural y disfrutar de sus bienes y valores.
- Convivir con las reglas de comportamiento socialmente aceptables, y contribuir en su evolución.
- Adquirir los valores de cooperación y solidaridad.
- Cuidar su salud y desarrollar armoniosamente su cuerpo; ejercer responsablemente y de manera creativa el tiempo libre.
- Desarrollar la sensibilidad y el arte como base a la creatividad.
- Ampliar su universo cultural para mejorar la comprensión del mundo y del entorno en que vive, para cuidar de la naturaleza y potenciar sus expectativas.
- Participar activamente en su desarrollo académico para acrecentar su capacidad de aprendizaje y evolucionar con autonomía como Licenciado(a) en Gestión de la Información en Redes Sociales.
- Asumir los principios y valores universitarios, y actuar en consecuencia.
- Emplear habilidades lingüístico-comunicativas de inglés como segunda lengua.
- Desarrollar su forma de expresarse, su creatividad, iniciativa y espíritu emprendedor.
- Desarrollar un juicio profesional basado en la responsabilidad, objetividad, credibilidad y justicia.

##### Particulares

- Crear sistemas de bases de datos mediante algoritmos, modelos matemáticos y modelos de software; para almacenar, recuperar y procesar datos de la sociedad en red.
- Seleccionar métodos y técnicas cualitativas y cuantitativas en el análisis de datos, comprender su significado, procesarlo y convertirlo en información útil para las organizaciones públicas y privadas.
- Analizar el marco normativo, la seguridad de las TIC y los delitos cibernéticos, para evaluar decisiones y formular soluciones racionales y éticas sobre el uso, acceso y protección de datos e información de la sociedad en red.



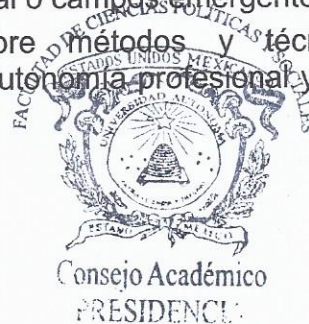


- Analizar datos mediante la selección de principios, métodos y técnicas estadísticas, modelos de inteligencia artificial, minería de datos y teoría de juegos; para localizar patrones, identificar tendencias, necesidades y problemáticas de la sociedad en red.
- Investigar mercados a través de métodos cuantitativos y cualitativos para mejorar o innovar productos tangibles e intangibles e identificar oportunidades de negocio en las organizaciones públicas y privadas.
- Desarrollar habilidades tecnológicas mediante el uso de las tecnologías de la información y comunicación, para operar plataformas digitales y comunidades virtuales de la sociedad en red.
- Crear estrategias discursivas mediante la composición editorial y diseño gráfico, para comunicar información, contenido e imagen de personas, productos tangibles e intangibles, organizaciones públicas y privadas en plataformas y redes sociales digitales.
- Gestionar información en las organizaciones públicas y privadas a través del proceso administrativo, para tomar decisiones estratégicas en los ámbitos de intervención profesional de los sectores primario, secundario y terciario.
- Seleccionar los canales y medios de comunicación para difundir información, promover la identidad, imagen y reputación de personas y organizaciones; así como comercializar productos tangibles e intangibles vía *internet*.
- Colaborar en la formulación de políticas, legislación y lineamientos en torno al acceso, uso y protección de datos e información, para personas y organizaciones públicas y privadas.
- Administrar sistemas de información en plataformas tecnológicas y redes sociales digitales, respetando las políticas, legislación y lineamientos sobre el uso, acceso y protección de datos e información.
- Proponer información para la toma de decisiones en el desarrollo de políticas públicas, para atender necesidades y resolver problemas en materia de planeación y desarrollo urbano, demográfico, educativo, salud, trabajo, seguridad social, vivienda, entre otros.

### Objetivos del núcleo de formación:

Proveerá al alumno de escenarios educativos para la integración, aplicación y desarrollo de los conocimientos, habilidades y actitudes que le permitan el desempeño de las funciones, tareas y resultados ligados a las dimensiones y ámbitos de intervención profesional o campos emergentes de la misma.

Comprenderá aprendizajes sobre métodos y técnicas especializadas, y capacidades para desarrollar la autonomía profesional y el desempeño aceptable en el campo laboral.





### Objetivos del área curricular o disciplinaria:

Crear herramientas de inteligencia artificial ad hoc, a través de modelos matemáticos, algoritmos y sistemas de software, para el procesamiento, modelado y simulación de grandes volúmenes de datos, y manejo de plataformas tecnológicas de negocio electrónico y comunidades virtuales.

### V. Objetivos de la unidad de aprendizaje

Crear sistemas informáticos capaces de realizar mapeo y compilación de datos de *internet* de forma automática, para la toma de decisiones,

### VI. Contenidos de la unidad de aprendizaje y su organización.

#### Unidad 1. Fundamentos básicos de internet.

**Objetivo:** Analizar los fundamentos básicos de la estructura de internet a través de su lenguaje y protocolo para asociarlos a la construcción de un *crawler*.

#### Temas:

- 1.1. Estructura de la WWW.
- 1.2. Estructura del lenguaje html y del protocolo http.
- 1.3. Diferencia entre *crawling* y *scraping*.
- 1.4. Bibliotecas para *crawling*.
- 1.5. Bibliotecas para *scraping*.

#### Unidad 2. Uso de bibliotecas para el rastreo en internet.

**Objetivo:** Establecer requerimientos técnicos, a través de la selección y utilización de bibliotecas para el proceso web de *crawling* y *scraping*

#### Temas:

- 2.1. Requerimientos técnicos para el uso de las bibliotecas.
- 2.2. La biblioteca *urllib* de *python*.
- 2.3. Bibliotecas *python* para el acceso a la web.
- 2.4. Bibliotecas de *parseo* de *html* (*LXML*, *Xpath* y selectores *CSS*).
- 2.5. *Beautiful Soup* y *Pyquery*.
- 2.6. *Apis* basadas en web (*requests*, *json* y *collections*).





### Unidad 3. Tratamiento de textos

**Objetivo:** Establecer herramientas de expresiones regulares a través del software *Python* para lograr procesar textos.

**Temas:**

- 3.1. Expresiones regulares.
- 3.2. Requerimientos técnicos para utilizar expresiones regulares en *Python*.
- 3.3. Uso de expresiones regulares.

## VII. Acervo bibliográfico.

### Básico:

Chapagain, Anish. (2019). *Hands-On Web Scraping with Python*. Estados Unidos: Packt Publishing.

Mitchell, Ryan. (2015). *Hands-On Web Scraping with Python*. Estados Unidos: O'Reilly Media, Inc.

Peicevic, Antun. (2016). *Create a web crawler in Python*. Estados Unidos: Geek University Press.

A. Del Coso Santos. (2009). *Desarrollo de infraestructuras para el modelado de usuarios*. España: Universidad Carlos III de Madrid.

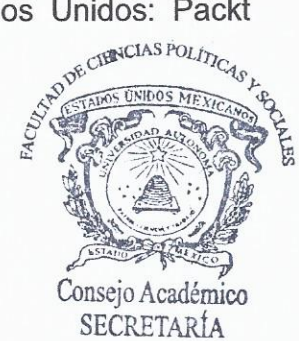
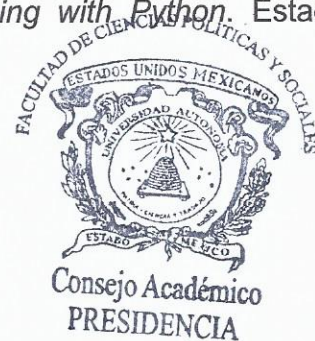
G. Martinez. (2006). *Clasificación mediante Conjuntos*. España: Tesis Doctoral, Departamento de Ingeniería Informática, Universidad Autónoma de Madrid.

### Complementario:

Kouzis-Loukas, Dimitrios. (2016). *Learning Scrapy*. Estados Unidos: Packt Publishing.

Nair, Vineeth G. (2014). *Getting Started with Beautiful Soup*. Estados Unidos: Packt Publishing.

Lawson, Richard. (2015). *Web Scraping with Python*. Estados Unidos: Packt Publishing.









Proyecto curricular de la Licenciatura en Gestión de la Información en Redes Sociales (2017)

Secretaría de Docencia • Dirección de Estudios Profesionales



Proyecto curricular de la Licenciatura en Gestión de la Información en Redes Sociales (2017)

Secretaría de Docencia • Dirección de Estudios Profesionales



DISTRIBUCIÓN DE LAS UNIDADES DE APRENDIZAJE OPTATIVAS

PERIODO 1 PERIODO 2 PERIODO 3 PERIODO 4 PERIODO 5 PERIODO 6 PERIODO 7 PERIODO 8 PERIODO 9

Psicología de la publicidad	3	3	3	3	3	3	3	3	3
Psicología del color	4	4	4	4	4	4	4	4	4
Diseño editorial	3	3	3	3	3	3	3	3	3
Ytibir sus problemas	3	3	3	3	3	3	3	3	3
Antropología visual	4	4	4	4	4	4	4	4	4
Algunas estrategias	3	3	3	3	3	3	3	3	3
Estrategias de promoción	4	4	4	4	4	4	4	4	4
Examen	3	3	3	3	3	3	3	3	3
Tercer año de sociología	3	3	3	3	3	3	3	3	3
Sociología	4	4	4	4	4	4	4	4	4

Nota: créditos

PARÁMETROS DEL PLAN DE ESTUDIOS

<b>Núcleo Básico</b>	48	<b>Núcleo Básico</b>	48
Obligatorio	20	Obligatorio	20
Optativo	28	Optativo	28
<b>Núcleo Sustantivo</b>	43	<b>Núcleo Sustantivo</b>	43
Obligatorio	22	Obligatorio	22
Optativo	21	Optativo	21
<b>Núcleo Integral</b>	17	<b>Núcleo Integral</b>	17
Obligatorio	3	Obligatorio	3
Optativo	14	Optativo	14
<b>Total del Núcleo</b>	116	<b>Total del Núcleo</b>	116
Obligatorio	50	Obligatorio	50
Optativo	66	Optativo	66
<b>Total del Núcleo Integral</b>	15	<b>Total del Núcleo Integral</b>	15
Obligatorio	3	Obligatorio	3
Optativo	12	Optativo	12

<b>TOTAL DEL PLAN DE ESTUDIOS</b>	
UA Obligatorias	50 + 2 Actividades académicas
UA Optativas	3
UA a acreditar	53 + 2 Actividades académicas
<b>Créditos</b>	381

\*Actividades académicas  
 \*\*Las horas de la actividad académica  
 Créditos a cursar por periodo escolar. Mínimo 21 y máximo 48  
 13 líneas de seriación →  
 e UA que debe impartirse, cursarse y acreditarse en la modalidad educativa escolarizada.



Consejo Académico PRESIDENCIA

