



ESPECIALIZACIÓN EN

WEB SCRAPING

BROCHURE 2024

#AprendeDesdeCasa

#AprendeConLosPioneros



PRESENTACIÓN

Los profesionales del dato están en búsqueda constante de nuevas fuentes de información, y la web se ha constituido en una gran alternativa para explotar su contenido. Por ello, el dominio de las técnicas y el software diseñados para este fin, llamado Web Scraping, representa un importante diferencial frente otras organizaciones.

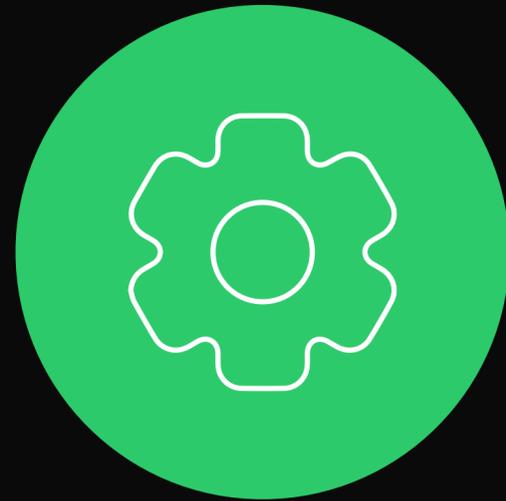
En tal sentido, DMC presenta la “Especialización en Web Scraping”, la cual tiene por objetivo capacitarte en la implementación de soluciones de scraping sencillas y aplicadas a proyectos mediante herramientas como Selenium y Playwright, complementado con técnicas para la exportación, transferencia y almacenamiento de los datos obtenidos.

 **Duración:** 36 horas cronológicas
(48 horas académicas)

 **Certificación:** Especialista en Web Scraping

 **Pre-requisito:** Conocimientos básicos en Python y en Transact-SQL (obligatorio). Deseable conocimiento de HTML.





Objetivo General

La especialización le permitirá al participante implementar soluciones de scraping sencillas y aplicadas a proyectos sobre diferentes tipos de datos mediante técnicas y herramientas para su extracción, manipulación y almacenamiento.



Objetivos Específicos

- Elaborar un diseño de plan de extracción de datos a partir de la identificación de los componentes de las fuentes (DOM).
- Realizar scraping básico de recursos web empleando herramientas como Selenium, Playwright, y Request.
- Aprender scraping avanzado aplicado a proyectos empleando las librerías Selenium y Playwright.
- Exportar, transferir y almacenar los recursos de datos de diversas fuentes y formatos; gestionar su almacenamiento SQL mediante mapeo ORM; así como manipularlos a través de dataframes y la librería Pandas del lenguaje Python.



DIRIGIDO A

- Ingenieros de datos y especialistas que trabajen en su recolección y procesamiento.
- Especialistas en inteligencia de negocios y analítica interesados en aprender técnicas de recolección de datos desde la web.



COMPETENCIAS

- Aplicar un diseño de plan de extracción de datos a partir de la identificación de los componentes de las fuentes (DOM).
- Obtener data mediante scraping básico de recursos web empleando herramientas como Selenium, Playwright, y Request.
- Realizar scraping avanzado aplicado a proyectos empleando las librerías Selenium y Playwright.
- Gestionar los recursos de datos de diversas fuentes y formatos; gestionar su almacenamiento SQL mediante mapeo ORM.

CARACTERÍSTICAS

Clases en Vivo

El 100% de las clases que se desarrollan en el programa son en vivo.

Asesoría Académica

Resuelve tus dudas con el asesor académico en línea.

Plataforma E-Learning

Accede en cualquier momento a materiales complementarios: lecturas, videos, tutoriales, clases grabadas y más.



Aprende haciendo

Desarrolla casos con datos reales, incluso puedes proponer casos de tu propio sector.

Certificación

Con una nota mínima de 14 sobre 20 y una asistencia mínima del 80%

Soporte técnico

Asistencia técnica permanente y acceso a máquinas virtuales de ser necesario.

MAPA CURRICULAR



MAPA CURRICULAR



Lógica de Extracción y Documentación

- Identificación de recursos de la fuente: Networks, Html, CSS, JS.
- Armado de flujo de extracción.
- Diseño del plan de extracción.



Scrapy Básico con Selenium

- Introducción a la técnica Scrapy con librería Selenium.
- Configuración de Driver manager: Personalización del simulador.
- Driver parámetros.
- Obtención de perfil de navegación (profile).
- Xpath: introducción, sintaxis.
- Búsqueda de elemento HTML (xpath).
- Descarga de recursos (texto, imagen, files).
- Manejo de scripts para JS.
- Estrategias para resolver Captcha Nivel 1.



Scrapy Básico con Playwright

- Introducción a la librería Playwright, comparación con Selenium.
- Configuración de Driver manager: Personalización del simulador.
- Driver parámetros.
- Obtención de perfil de navegación (profile).
- Buenas prácticas (W3C).
- Búsqueda de elemento HTML (xpath).
- Descarga de recursos (texto, imagen, files).
- Manejo de scripts para JS.



Scrapy con Request

- Introducción a la librería Request.
- Manejo de datos JSON.
- Manejo de métodos get, post.
- Formateo de Request Result (textos).
- Descarga Request recursos (imagen, file).
- Manejo de tipos de autorización (CURL).

MALLA CURRICULAR



Selenium para Proyectos

- Scraping con patrones de diseño: Builder, Prototype, Strategy.



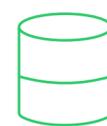
Playwright para Proyectos

- Scraping con patrones de diseño: Builder, Prototype, Strategy.



Acumulación y Gestión de Recursos

- Exportación de formatos de datos (JSON, XLSX, CSV).
- Exportación de formatos de imágenes (JPG, PNG).
- Compresión de carpetas y archivos Python.
- Transferencia de recursos mediante protocolo FTP, SFTP.
- Transferencia de recursos por API (request-post).
- Extracción y aplicación de metadatos de recursos.



Almacenamiento de Recursos

- Introducción a ORM y SQLAlchemy.
- Puerta de enlace al engine (SQL Server).
- Creación de tablas con ORM.
- Creación de constraints con ORM.
- Modelado entidad-relación.
- Acciones CRUD ORM: Insert, Select, Update, Delete.
- Sentencias SQL desde ORM: Group by, Subconsultas, Joins, With.



Manipulación de Datos

- Introducción a la librería Pandas.
- Creación de Dataframes.
- Manejo de Dataframes y series: Filtros, CRUD, Comandos DTypes, AsType, Info, Iterrows, Agregate, Merge.

Buscamos liderar la transformación de las empresas.



+14

Años de Experiencia

Desde el año 2009 capacitamos con técnicas de análisis de datos a profesionales de diferentes empresas y sectores.

+100

Expertos en Analítica

Nuestra plana docente ocupa los cargos más importantes en las áreas analíticas de todos los sectores.



+15K

Profesionales Capacitados

Nuestros alumnos inscritos pertenecen a las mejores compañías del medio y amplifican con nosotros su red de contactos especializada.

+50

Capacitaciones Especializadas

Contamos con una variedad de líneas temáticas y niveles de especialización.



+300

Empresas Asesoradas

Las empresas top del mercado buscan nuestra asesoría y les brindamos soluciones analíticas ad hoc.

+5

Big Data Analytics Summit

Organizamos el evento más grande del Perú, con los mejores ponentes internacionales.



Di:ic
Perú

Formando profesionales mediante la analítica de los datos.

Di!iC
Perú

