



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC2513 - Sección 1 — Tecnologías y Aplicaciones Web

# Tarea 1

## Entrega

- **Fecha y hora:** Viernes 1 de septiembre del 2023, a las 22:00
- **Lugar:** Repositorio individual en la organización del curso en GitHub

## Objetivos

- **Construir** programas usando herramientas para interactuar con aplicaciones web
- **Utilizar** la herramienta *DevTools* para explorar páginas web
- **Producir** documentación efectiva y clara que permita el entendimiento del proceso realizado

## Descripción

En esta tarea, tendrán la oportunidad de adentrarse en el mundo del *web scraping*, una habilidad esencial en el campo de la ciencia de datos que consiste en extraer datos de páginas web mediante la simulación de la navegación, similar a lo que haría un usuario. Para esto, en esta tarea utilizarán dos herramientas comúnmente utilizadas: [Selenium](#), una herramienta de automatización de navegadores, y [BeautifulSoup](#), una biblioteca para analizar y extraer datos de páginas web, especializada en extraer datos de archivos HTML y XML.

Tanto Selenium como BeautifulSoup, son librerías de Python que deberás instalar para realizar esta tarea. A continuación puedes encontrar los links de instalación:

- [Selenium](#) (ver la sección *Python*)
- [BeautifulSoup](#)

Los objetivos de esta tarea son: que aprendan a extraer datos de manera efectiva, se familiaricen con la estructura HTML de páginas web funcionales, y que utilicen *DevTools* para inspeccionar páginas web.

## Simulación de acciones en BuscaCursos

Para esta tarea, deberán “scrapear” la página <https://buscacursos.uc.cl> con Python. Para ello, la navegación deberá comenzar desde esa URL. Luego, su programa deberá automatizar algunas acciones utilizando Selenium para llegar a la página web desde la que se desee extraer información.

Mediante el formulario de búsqueda, se deberán encontrar aquellos cursos que fueron dictados el primer semestre de 2023 y que comiencen con la sigla 'IIC3'. Se deberá rellenar los campos necesarios y seleccionar ítems desde las listas desplegables para que la búsqueda sea efectiva (no puedes acceder a esta información directamente mediante un *link* u algún otro, sino que es necesario simular la navegación, lo que será revisado por el cuerpo de ayudantes).

**IMPORTANTE:** Recordar que uno de los objetivos de esta tarea es familiarizarse con las *DevTools*, por lo que es indispensable que antes de utilizar Selenium y BeautifulSoup se inspeccione la página y analice su composición para saber dónde está la información deseada y pensar posibles maneras de acceder a esta. De ese modo, se podrían encontrar múltiples formas de obtener los datos solicitados.

### Extracción de información útil

Ya teniendo los resultados de la búsqueda, se deberán extraer los datos de los ramos que cumplan las condiciones pedidas, y entregar un listado de los ramos en un `.csv` con el siguiente formato:

SIGLA DEL CURSO;NOMBRE DEL CURSO;PROFESORES DEL CURSO

En el caso de que un curso sea dictado por dos o más profesores, estos deben ir separados por una coma (','):

SIGLA DEL CURSO;NOMBRE DEL CURSO;PROFESOR1,PROFESOR2,PROFESOR3

### Documentación del proceso

Este ítem tiene como propósito comprender el proceso detallado de cómo se llevó a cabo la obtención de datos. En primer lugar, se espera que expliquen cómo se familiarizaron con la composición de la página y sus respectivas herramientas de desarrollo del navegador (*Dev Tools*). Deben describir los pasos realizados para navegar hasta secciones específicas de la página, así como la ubicación de información relevante dentro de los archivos y componentes.

Posteriormente, es necesario describir por qué esta exploración previa es crucial para poder automatizar la navegación y llevar a cabo la búsqueda de los datos requeridos de manera eficiente.

Finalmente, deberán explicar el funcionamiento del código desarrollado e indicar qué elementos de la página (por ejemplo, clases, identificadores, entre otros) habrían facilitado o simplificado el proceso de extracción de información. Esto permitirá comprender por qué es beneficioso escribir código organizado y bien estructurado para el scraping y la navegación automatizada.

Para asegurar una correcta entrega de la documentación, se les hará entrega de una plantilla con la estructura que debe seguir su descripción y las preguntas que deben responder.

### Entregables

Cada entrega deberá incluir los siguientes archivos:

- Un archivo de extensión `.py` con el código utilizado para realizar el proceso de *web scraping* (desde la simulación de acciones hasta la preparación del documento de salida que contiene la información).
- `README.md` con el paso a paso de las acciones realizadas durante el proceso de *web scraping*, así como también de una descripción de que fue lo que extrajeron y cómo lo hicieron.

## Rúbrica

A continuación, se describe el criterio de cada uno de los ítems de la rúbrica con la que se evaluará esta entrega. La nota se calcula al 50 % considerando un total de 10 puntos.

- **Utilización de Selenium [2.5 puntos]:** Que al ejecutar el código comiencen las acciones automáticas en el buscacursos. Se deberá rellenar el formulario de búsqueda y dar inicio a esta, para finalmente obtener todos los cursos que la UC ofrece que comiencen con la sigla 'IIC3'.
- **Utilización de BeautifulSoup [2.5 puntos]:** Analizar el documento HTML del buscacursos mediante esta herramienta y extraer la información solicitada de cada uno de los cursos que comiencen con la sigla pedida, para entregarlo con el formato solicitado.
- **Comprensión y documentación del proceso [5 puntos]:** Que se explique de manera clara y ordenada el paso a paso de la realización de la tarea. Comenzar con el comando que se debe ejecutar para dar inicio al *web scraping* y con la explicación del funcionamiento de tu código. Luego comentar el análisis y familiarización de la página web, seguir con como este se relaciona con el uso de Selenium y BeautifulSoup y terminar con las sugerencias que te hubieran ayudado a realizar de manera más fácil el scraping.

## Dudas

Pueden hacer sus consultas acerca del enunciado de la tarea en las [issues](#) del repositorio del curso. No se van a contestar dudas por correo.