



## Trabajo Práctico

### Identificación

---

Curso: Estructura de los Lenguajes

### Objetivos

---

- Fijar los conceptos y usos de los diferentes lenguajes de programación.
- Integrar los conocimientos sobre la descripción de los lenguajes con la implementación.
- Facilitar la comprensión de los temas relacionados con la implementación de los lenguajes de programación.
- Aplicar criterios y características para evaluar lenguajes de programación.
- Enunciar y describir características y propiedades de los lenguajes de programación.
- Comparar estructuras de diferentes lenguajes de programación.
- Adquirir experiencia en el uso de al menos un lenguaje de programación.

### Actividades

---

El trabajo práctico consiste en programar en 4 lenguajes, que pueden ser: **Python, Go, R, Ruby, C, Java, C++, C# u otro de su preferencia**, utilizando las herramientas de scraping de cada lenguaje para realizar los ejercicios. Deberá justificar la selección de los lenguajes en función a la tarea a realizar.

El trabajo se divide en dos partes de 70% y 30% y se indican a continuación.

**Tema 1 (70%): Ratings de lenguajes más populares en GitHub** considerando el listado de los 20 lenguajes del índice TIOBE disponible en <https://www.tiobe.com/tiobe-index/>

- 1.1 Realice un programa del tipo web scraping sobre la web de GitHub <https://github.com/topics> para extraer automáticamente el número total de repositorios públicos de cada uno de los 20 lenguajes populares (NRO\_APARICIONES).

Here are **163,010** public repositories matching this topic...

Language: All ▼

Sort: Best match ▼

- 1.2 Almacene los datos extraídos en un archivo de texto (.txt o .csv) y nombrarlo como "Resultados". Guardar los datos en la forma: NOMBRE LENGUAJE, NRO\_APARICIONES
- 1.3 Calcule el Rating de GitHub de cada uno de los 20 lenguajes. Para ello, deberá obtener el valor máximo y el valor mínimo de NRO\_APARICIONES, y calcular el Rating según la siguiente ecuación:



$$RATING\_GITHUB = \frac{(x - MIN)}{(MAX - MIN)} \times 100$$

- 1.4 Ordene descendientemente por RATING\_GITHUB e imprimir en pantalla: NOMBRE LENGUAJE, RATING\_GITHUB, NRO\_APARICIONES. Se deben visualizar los 20 lenguajes.
- 1.5 Realice un gráfico de barras para visualizar los 10 lenguajes con mayor número de apariciones entre los resultados. Visualizar en el eje de ordenadas el NRO\_APARICIONES y en el eje de abscisas el NOMBRE LENGUAJE. Hacer el gráfico en orden descendente.

## Tema 2 (30%): Topics de interés en la web de GitHub.

- 2.1 Seleccione un Topic de su interés disponible en <https://github.com/topics>. A continuación, realice una extracción automática de todos los Topics asociados a su interés y que fueron actualizados en los últimos 30 días. Guardarlos en un archivo.

- Por ejemplo: el topic seleccionado de interés es **go** y los **Topics asociados son los resaltados** en la siguiente imagen:

The screenshot shows two GitHub repository pages. The top page is for the 'go' topic, showing a list of associated topics: 'static-site-generator', 'cms', 'hugo', 'documentation-tool', 'blog-engine', 'hacktoberfest', and 'content-management-system'. The bottom page is for the 'gin-gonic / gin' repository, showing a list of associated topics: 'middleware', 'performance', 'framework', 'router', 'server', and 'gin'. Both pages have a red box highlighting the associated topics.

- 2.2 Contar el número de veces (NRO\_APARICIONES) en que cada una de las palabras (TOPIC) se encuentra en el conjunto de datos recuperados y ordenarlos de forma descendente. Imprimir los resultados en pantalla.



- 2.3 Realice un gráfico de barras visualizando las 20 palabras con mayor número de apariciones contadas. Visualizar en el eje de ordenadas el NRO\_APARICIONES y en el eje de abscisas el TOPIC. Hacer el gráfico en orden descendente.

**Elaborar la documentación del trabajo realizado.** Por cada lenguaje, describe su experiencia en cuanto a las implementaciones realizadas, según el contenido solicitado a continuación, y presentar el documento en formato PDF.

1. Identificar el documento con una carátula. Título: "Extracción de datos web en diferentes lenguajes de programación", seguido de el/los nombre(s) y apellido(s).
2. Presente el documento en diferentes secciones como: carátula, introducción, contenido, conclusión y bibliografía. Mencionar todas las referencias utilizadas, tanto para el código y la documentación. Se espera una redacción correcta del texto, evite el uso excesivo de lenguaje coloquial.
3. **Contenido solicitado:**
  - 3.1 Justificar la selección de los 4 lenguajes utilizados.
  - 3.2 Por cada lenguaje, describir las herramientas (paquetes, librerías, módulos) de scraping utilizados y realizar comentarios generales sobre los códigos realizados.
  - 3.3 Por cada lenguaje, describir el entorno de desarrollo utilizado.
  - 3.4 Por cada lenguaje, escribir las ventajas y desventajas encontradas.
  - 3.5 Indicar la palabra de su interés, utilizada para realizar el Tema 2, y comentar sobre los resultados obtenidos.
  - 3.6 Escribir una comparación propia entre los 4 lenguajes utilizados, considerar los criterios propuestos por Sebesta.

Para su conocimiento: algunas páginas web utilizan medidas de seguridad utilizando técnicas para controlar de forma efectiva el acceso mediante web scrapers, por ejemplo, bloquear la dirección IP del scraper tras varios intentos de acceso que infrinjan las normas de la web.

## INDICACIONES:

1. **El trabajo podrá ser realizado de manera individual o en grupos de hasta 3 personas**, deberá ser realizado por los Alumnos (de principio a fin). Se permite hacer referencias a páginas o tutoriales utilizados, pero no un directo *copy paste* porque se espera una propia codificación y redacción.
2. No se permite el uso ninguna API (existente) o Programa como alternativa a su codificación web scraping, a menos que sirvan para resolver temas puntuales donde no exista una alternativa evidente.
3. **Copiar será castigado con la pérdida total de los trabajos/actividades. Además, se pasarán los antecedentes a la instancia correspondiente.**
4. La documentación del trabajo deberá estar en formato .PDF
5. Enviar los códigos fuentes de sus ejercicios en archivos legibles .txt .ipynb (cuaderno de Jupyter) .py . java .c u otra. Nombre correctamente sus archivos. Además, debe acompañar un archivo leame.txt con las indicaciones necesarias para ejecutar los programas.
6. Preparar y enviar todos los archivos del trabajo en una carpeta comprimida de la siguiente manera:
  - Si el trabajo es individual, nombrarlo **NOMBRE\_APELLIDO**.
  - Si el trabajo es grupal, solamente deberá ser enviado por **1 de los integrantes** y nombrarlo **GRUPO\_APELLIDOS**



7. **Entrega intermedia:** está planificada para dar un seguimiento del trabajo a ser presentado a fin de tomar medidas correctivas y recibir sugerencias, así como compartir entre los compañeros durante la sesión *Meet* en una de las dos fechas asignadas (ver plazo).
8. Las horas de clases podrán ser utilizadas para hacer el trabajo práctico y realizar consultas.

**Para la Evaluación se tendrán en cuenta:**

**Presentación final:** se verificarán los temas desarrollados, mediante una sesión *Meet* de presentación de trabajos. Además, se deberán enviar los archivos (código fuente y documentación) al classroom.

**Documentación:** la documentación contiene todo lo solicitado de manera correcta. La estructura y el formato del documento están correctos. Ortografía y claridad en la redacción.

**Responsabilidad:** entrega a tiempo del trabajo finalizado.

## Plazo

La entrega del trabajo tiene como fecha de inicio el día 27 de abril de 2022 y como fecha de:

- **Entregas intermedias** de avances del trabajo, el 9 y 11 de mayo de 2022, presentación vía *Meet* en el horario de clases.
- **Entrega del trabajo finalizado**, el **18 de mayo** de 2022 (presentación vía *Meet* en horario a coordinar). Subir el trabajo completo en el aula virtual de classroom hasta las 23:00 horas.

## Evaluación

Rúbrica de evaluación:

Indicadores y niveles de logro	No	Parcialmente	Si
Presentación final, archivos y documentación	0%	Según temas correctos	100%