



IIC2115 – Programación como Herramienta para la Ingeniería (I/2022)

Laboratorio 4

Aspectos generales

- **Objetivo:** evaluar individualmente el aprendizaje sobre bases de datos relacionales y Web Scraping en Python, a través de la descarga de información, la construcción de una base de datos y la generación de consultas asociadas a estos datos.
- **Lugar de entrega:** domingo 3 de julio a las 23:59 hrs. en repositorio privado.
- **Formato de entrega:** debe entregar el archivo Python Notebook (**L4.ipynb**) con la solución del laboratorio y la base de datos generada en formato **.db**. El archivo debe estar ubicado en la carpeta **L4**. Es requerimiento de formato el utilizar múltiples celdas de texto y código para la construcción de la solución. Laboratorios que no cumplan el formato de entrega tendrán un descuento de 0,5 pts.
- **Entregas atrasadas:** El descuento por atraso se realizará de acuerdo a lo definido en el programa del curso. Si su laboratorio es entregado fuera de plazo, tiene hasta el **lunes 4 de julio a las 11:59 AM** para responder el formulario de **entregas fuera de plazo** disponible en el Syllabus.
- **Issues:** Las discusiones en las *issues* del Syllabus que sean relevantes para el desarrollo del laboratorio, serán destacadas y se considerarán como parte de este enunciado. Así mismo, el uso de librerías externas que solucionen aspectos fundamental del laboratorio no podrán ser utilizadas. Solo se podrán utilizar las que han sido aprobadas en las *issues*, previa consulta de los estudiantes.
- **Laboratorios con errores de sintaxis y/o que generen excepciones en todas las ejecuciones** serán calificados con **nota 1.0**.

Introducción

En este laboratorio deberá generar una base de datos relacional académica lo más completa posible, a partir de información obtenida mediante Web Scraping. Tanto la fuente de información (sitio web de la institución académica) como el esquema de datos quedan completamente a criterio de ud. Una vez construida la base de datos, deberá responder una serie de preguntas mediante consultas en SQL.

Para el desarrollo deberá utilizar `bs4` (u otra herramienta de scraping) y `sqlite3`. Todas las misiones tienen el mismo puntaje, mientras que las submisiones dividen el puntaje de su misión de manera uniforme.

Misiones

1. **Descarga de información:** seleccione una institución académica de la cual pueda extraer información de su estructura, oferta educacional y académicos, a través de sitios web. No es necesario que la información sea completa, es decir, que contenga todas las publicaciones o todos los cursos dictados. Basta con que contenga suficiente variedad como para responder a las preguntas que se describen en la última misión. Para el proceso, utilice cuantos llamados sea necesario a la librería de Web Scraping que utilice.
2. **Construcción de la base de datos:** a partir de la información descargada en el ítem anterior, construya una base de datos relacional con `sqlite3`, que capture en tablas las distintas entidades presentes en la información, además de las relaciones entre ellas. En particular, es importante que las tablas cumplan con la estructura e información almacenada, que permita responder mediante consultas en SQL, a las preguntas que se describen la siguiente misión.
3. **Consultas:** escriba consultas SQL para responder a las preguntas que se indican a continuación. Considere que cada pregunta debe ser respondida utilizando **una única consulta**.
 - (a) Encuentre la facultad/departamento que más publicaciones genera en promedio por académico.
 - (b) Encuentre a los cinco académicos que más han colaborado (número de publicaciones) con académicos de facultades/departamentos distintos al que pertenecen, dentro de la institución académica modelada.
 - (c) Encuentre a los 10 académicos de jornada completa que más alumnos han atendido en sus cursos, después de las 13:30 hrs.
 - (d) Encuentre a la facultad/departamento que más salas distintas utiliza.
 - (e) Encuentre el bloque horario donde hay más vacantes no ocupadas en los cursos.

Corrección

Es importante que deje todas las celdas de su trabajo ejecutadas antes de subir el archivo, de lo contrario se le aplicará un descuento de 0,5 ptos. al puntaje total. Para la corrección de este laboratorio, se revisarán los procedimientos desarrollados para responder las diferentes misiones propuestas y la estructura de como utiliza los módulos para Web Scraping y *sqlite3*.

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.