

Tarea 1 - SOAP

El SOAP (Seguro Obligatorio de Accidentes Personales) es el seguro que debe ser contratado por todo propietario de un vehículo motorizado en Chile, debido a esto, existen muchas compañías de seguro que lo proporcionan, entregando datos privilegiados a cualquier usuario que intentase cotizar con el simple hecho de ingresar la patente. Además, ocurre lo mismo con el permiso de circulación que también lo ofrecen todas las municipalidades de Chile.

Su objetivo principal en esta tarea es demostrar que se pueden obtener datos de personas y/o vehículos a partir de sitios utilizados con otros fin y guardarlos en una base de datos.

Objetivos específicos:

Usted debe elegir una página la cual entregue dichos datos y escribirla en el documento compartido por el profesor.

Observaciones:

- La URL que seleccione, no debe ser igual a alguna previamente escrita, por lo que se recomienda que haga una temprana búsqueda de la que usará. Recuerde que la página debe ser a la cual usted realizará las consultas y no sirve colocar páginas que sólo son intermediarias con la base de datos.
- La modificación de la URL de cualquier otra persona que no sea usted, será evaluado con nota 1.0, sin ninguna excepción.
- La URL que elija DEBE retornar algún dato personal (nombre, rut, número telefónico, dirección, vehículo, tipo de vehículo, motor, etc.).
- En caso de no existir más sitios de SOAP o Permiso de Circulación que pidan sólo la patente, podrá utilizar una bd que se le entregará con las duplas de ruts y números de patente para buscar si están registradas en alguna municipalidad.

Debe crear un script (patentegen.sh) que creará la lista con todas las patentes chilenas, las cuales deben ser válidas con el formato antiguo utilizado entre los años 1985 y 2007 (BB*10*00). Ref [Wikipedia](https://es.wikipedia.org/wiki/Formato_de_patente_de conducir)

Observaciones:

- Tener en cuenta las letras que se utilizan y cuáles no en la creación de estas patentes.
- Como se menciona, DEBEN realizarse sólo sentencias en lenguaje bash.

Luego de tener las patentes listas, debe crear otro script (getdata.sh) que obtenga los datos de la URL seleccionada y los guarde en un archivo .db (utilice sqlite3). La base de datos debe contener como mínimo las siguientes columnas, en caso de existir dichos valores. Primer Nombre, Segundo Nombre, Apellido Paterno, Apellido Materno, rut, dígito verificador.

Observaciones:

- Si las consultas tardan demasiado, pregúntese si está utilizando toda la capacidad del computador.
- Las consultas de manera masiva para que no tengan ningún tipo de problema, se deben realizar desde algún proxy (proxychains), sobre todo si trabajarán desde la Universidad.
- Se recomienda que primero cree un archivo (data.csv), y luego de terminar el proceso de obtención de datos, trabaje en cómo escribir su base de datos.

Evaluación:

- Automatizar la extracción de datos sensibles a partir de una url
- Automatizar la consulta a todas las patentes generadas.
- Automatizar la creación de la base de datos con todos los registros.
- La base de datos debe ser consistente (por ejemplo, la columna rut sólo debe contener datos de tipo entero).
- En el código debe mencionar el tipo de procesador utilizado, y cantidad de urls consultadas por minuto.

Sobre la entrega:

- La tarea es totalmente individual.
- Cualquier indicio de copia será evaluado con nota 1.0
- Deberá enviar su trabajo al correo del ayudante (hector.cisternam@mail.udp.cl) con copia al profesor con solamente los 3 archivos solicitados. No se evaluarán otros archivos que sirvan como main.
- La fecha de entrega será hasta el día 8 de Abril. (Recuerde que el servicio SOAP está disponible sólo por un tiempo limitado y no necesariamente hasta el 8 de Abril).