



CENTRO UNIVERSITARIO
ENRIQUE DÍAZ DE LEÓN A.C.

PRESENTACIÓN

Carrera

Ingeniería en software

Materia

Programación de Sistemas

Alumno

Leonardo Samuel Rodríguez Navarrete

Profesor

Ing. Mónica Hernández

Fecha

15 de enero del 2020

Investigacion



CONTENIDO

Formas de prevenir el scrapping:

- **Usar cookies o Javascript** para verificar que el visitante es un navegador web

Como la mayoría de los web scrapers no procesan código javascript complejo, para comprobar que el usuario es un navegador web real se puede insertar en la página un cálculo javascript complicado, y verificar que ha sido correctamente calculado

- **Introducir Captchas** para asegurarse de que el usuario es un humano

Sigue siendo una buena medida para eliminar visitantes robots; aunque últimamente éstos se han sofisticado y consiguen saltárselos.

- **Establecer límites de peticiones y conexiones**

Se pueden mitigar las visitas de los scrapers ajustando el número de peticiones a la página, y conexiones; puesto que un usuario humano es más lento que uno automático.

- **Ofuscar ó esconder los datos**

Los web scrapers rastrean datos en formato texto; por tanto es una buena medida publicar los datos en formato imagen o flash.

- **Detectar y bloquear fuentes maliciosas conocidas**

- Localizar y frenar el acceso de site scrapers conocidos; entre los que se pueden encontrar nuestros competidores; y a los que podríamos bloquear su dirección IP

- **Detectar y bloquear las herramientas de site scraping**

La mayoría de las herramientas usan una firma identificable que sirve para su detección y consiguiente bloqueo

- **Actualizar constantemente los HTML tags** de la página

Los scrapers se programan para buscar determinados contenidos en los tags de la página web. Si cambiamos frecuentemente los tags mediante la introducción, por ejemplo, de espacios, comentarios, nuevas etiquetas, etc ... se puede evitar que una mismo scraper repita el ataque

- **Utilizar contenido web falso para atrapar a los atacantes**

Si sospechas que tu información está siendo plagiada, se puede publicar contenido ficticio y monitorizar sus accesos para descubrir al scraper.



CENTRO UNIVERSITARIO
ENRIQUE DÍAZ DE LEÓN A.C.

REFERENCIAS

<https://www.baquia.com/emprendedores/2014-01-26-imperva-seguridad-innovae-como-evitar-que-alguien-se-apropie-de-los-contenidos-de-nuestra-pagina-web>