

Web Categorization Data Enriches Web Tracking Analysis

Posted on October 10, 2022

WhoisXML API continues to work with security researchers, scholars, and other experts as part of our goal to raise cybersecurity awareness and our overarching mission of making the Internet a safer place.

Among those who recently heeded our call is Fergus Smith, a software development MST student at the University of Glasgow. He worked on a distributed multiplayer game aimed at educating Internet users about the prevalence of web tracking.

The project called “TrackerHunt” was developed as a Chrome extension that enables users to learn what specific websites are tracking them. WhoisXML API’s [Website Categorization API](#) helped deepen the web tracking insights by allowing users and researchers to see the categories the web pages that employ web trackers fall under.

The Challenge: Detecting the Categories of Web Pages Tracking User Data

Web tracking is a common practice where websites and third parties collect, analyze, and share the data of online visitors. Tracking is usually done through cookies. While web tracking can make Internet browsing faster and more convenient, it’s essential for users to understand who is tracking their data and what type of information is collected.

TrackerHunt addressed this challenge in a fun and educational way. Users earn points whenever they identify web trackers while browsing the Internet. Their learning experience is based on their browsing activities rather than simulations.

However, while it's great to learn what websites were tracking the game participants, Smith also wanted to see what types of web pages these were.

The Solution: Game Algorithm Enriched with Web Categorization API

To deepen user education about web tracking, the developer integrated WhoisXML API's Web Categorization API with the game algorithm. Whenever a tracker is detected, the game sends a request to WhoisXML API's Web Categorization API. The list of categories connected to the website would then be returned and displayed.

The developer and game users had a seamless experience from API integration to game deployment. To quote Smith, "With other APIs, there's a tendency for requests to fail. But I didn't notice that with Website Categorization API. I noticed that whenever a request was made, the website was always categorized even if it was under an unknown category."

Another feature that was helpful to the developer was the availability of confidence ratings. Website Categorization API not only returns the web categories but also provides confidence ratings that show how relevant the category is to the website. The feature allowed the developer to identify and remedy those with low confidence ratings to ensure better accuracy.

Internet Transparency: WhoisXML API's Advocacy and TrackerHunt's Goal

Learning what categories websites that track user data fall under helps participants understand web tracking more deeply. For example, when most trackers belong to the retail category, users become aware that their data can be used to suggest products they may find interesting based on their browsing activities.

On the other hand, web tracking may be more concerning when done mainly by websites under the social media and news categories, as user data can be used in misinformation and disinformation campaigns.

TrackerHunt aimed to widen user education about web tracking, which resounds with WhoisXML API's vision of a transparent and more secure Internet.

In the words of CEO Jonathan Zhang, “Projects like TrackerHunt motivate us to continue and enhance our collaboration with members of the cyber community. Our WHOIS, DNS, and IP intelligence allow security researchers, teams, and solutions to widen their visibility of the Internet and help them make sense of online activities, whether adversarial or not.”