Top 5 Use Cases of WhoisXML API's New Website Categorization API

Posted on July 13, 2021

WhoisXML API's website categorization products have been helping organizations determine the authenticity and reliability of sites by scanning the meta tags and content of more than 152 million websites. The machine learning (ML)-driven process allows organizations to detect suspicious domains, align their site categories with their marketing messages, and target the right leads, to name a few.

Today, the tools have been made more massive by adopting the classifications used by the Internet Advertising Bureau (IAB). IAB's content classification taxonomy has become a standard in the industry, along with other solutions that aim to promote brand safety, ad fraud prevention, and consumer privacy.

Besides the number of categories, Website Categorization Lookup and Website Categorization API also have improved functionalities that provide users with much-needed accuracy and confidence.

What's New with the Website Categorization Engine?

WhoisXML API understands that users of any website categorization tool want the results to be data-driven, precise, and inclusive. With this in mind, we ramped up our ML engine and armed it with natural language processing (NLP) to provide users with tiered web categorization results supported by confidence scores.

The new website categorization tools feature three significant improvements:

- **Increased number of web categories:** Started at 25, the new and improved website categorization tools now support more than 500 standard categories, aligned with the IAB's list. This ensures the tool can be used in conjunction with other standardized advertising tools and software.

- **Tiered results:** The tools now identify websites into multiple categories, with two tiers under each classification. The top category is labeled "tier 1" and the second "tier 2." The corresponding IAB ID number is also returned.

- **Confidence score:** Another new feature is the confidence score, which reflects how relevant

a category is to the queried website. The higher the score, the more accurate the categorization is.

Top Website Categorization Use Cases

A robust and comprehensive website categorization tool has more effective use cases. We show the website categorization engines in action below.

1. Content Filtering to Improve Productivity

Website Categorization Lookup and API can help organizations implement content filtering by detecting nonwork-related sites.

Content filtering is the process of screening network user access to specific websites. One of the goals is to improve employee productivity by ensuring that they can't visit nonwork-related websites. A survey involving 1,000 employees in the U.S. revealed that 66% of the respondents shopped online during work hours while 95% use social media. The distraction can result in wasted workweeks per year.

Hence, most organizations prohibited access to shopping, fashion, and sports websites during work hours. Using our web categorization engine, we found 528 of such sites on the Alexa top 10,000 sites. These sites are classified under Shopping, Style & Fashion, and Sports. The including of websites categorized under Movies and Video Gaming would add dozens more to the list.

2. Malware Protection through Content Filtering

Malware infection remains a favorite threat actor tactic. Verizon's 2021 Data Breach Investigation Report (DBIR) revealed that malware is used for 20% of infiltration, 61% of malicious payload deployment, and 28% of data exfiltration. Thus, malware plays a significant role in data breaches and other cyberattacks.

Malware can use several entry points, from phishing emails to malicious websites. Magecart

malware attacks, for one, mainly target e-commerce sites. In this context, blocking adult and other sites that belong to IAB's Sensitive Topics category could help, as these websites may carry malware.

Out of the Alexa Top 10,000 sites, the website categorization engine found 210 websites classified under Sensitive Topics. Several of these are adult websites while others are probably just not safe to access—including those domains marked either "malicious" or "suspicious" on VirusTotal:

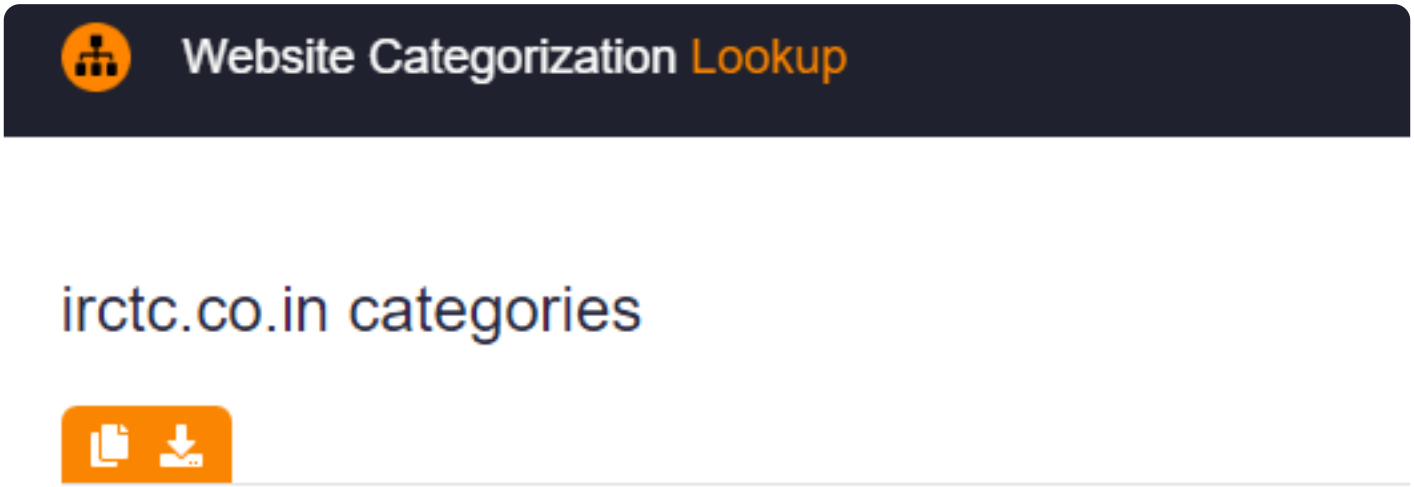
- livejasmin[.]com
- pornhubpremium[.]com
- truecaller[.]com
- redgifs[.]com
- subscene[.]com
- daftsex[.]com

3. Brand Protection through Third-Party Risk Assessment

A crucial part of brand protection is third-party assessment. After all, a company's reputation is affected by the failures committed by third parties. A Deloitte study revealed that 30% of listed companies believe their share prices could fall by 10% or more after a third-party-related security incident.

A number of brand protection strategies may help, such as including website categorization in third-party assessment and monitoring. To illustrate, consider a business-to-business (B2B) company catering to the automotive industry. Third-party vendors and suppliers can automatically pass the initial stage of third-party assessment when Website Categorization API's ML and NLP engines detect their websites under the Automotive category. The B2B organization can further set a certain confidence level to ensure that only those highly relevant to the industry can pass the assessment and proceed to the succeeding phases.

For example, irctc[.]co[.]it is also classified as an Automotive site, with Cars as its tier 2 category. However, the confidence level is low so the third-party assessment tool may flag it.



4. Lead Generation

For B2B companies, gathering leads involves obtaining a list of businesses in the target industry. If an organization develops programs and software for the education sector, for example, it would need a list of schools, universities, and other educational institutions before it can market its products.

Website Categorization API can help by instantly determining whether or not a website belongs to the target market. In the Alexa Top 10,000 sites classified by the tool, about 144 websites fell within the Education category. These include the following:

- udemy[.]com
- duolingo[.]com
- mit[.]edu
- khanacademy[.]org
- harvard[.]edu
- codecademy[.]com
- quizizz[.]com
- brainly[.]in
- stanford[.]edu
- illinois[.]edu

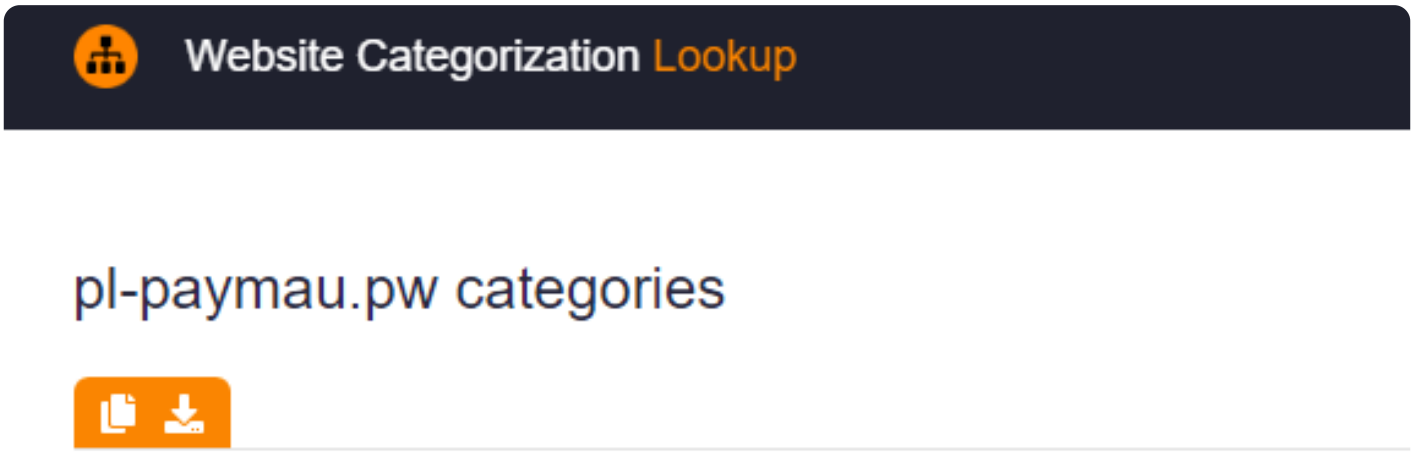
The B2B company can further narrow down the list by including tier 2 classifications in the process. For instance, it may target those that also belong in the Online Education category if its product is specifically for institutions that offer online courses. Confidence scores can then be used to rank which sites to prioritize in their marketing efforts.

5. Fraud Detection and Prevention

Fraud detection, such as detecting anomalies in particular events, can be simplified with the aid of ML algorithms that detect even the smallest traces of inconsistencies. ML-powered website categorization engines can, therefore, be used in fraud detection and prevention.

A money transfer request from invoice@pl-paymau[.]pw, for instance, could seem believable, especially if the email imitates a company employee or an executive. However, fraud detection systems with Website Categorization API would detect that the domain falls under Sensitive Topics and Spam or Harmful Content.

These categories would raise a red flag since they are inconsistent with a string found in the domain, which is “pay.” The alert would then enable security teams to avoid possibly fraudulent transactions.



Conclusion

In an effort to make the Internet a safer place, WhoisXML API improved the functionality of its website categorization engines while ensuring that categories conform to widely accepted industry

standards. Adopting a tiered format not only supports IAB standards but also helps users obtain precise results. On the other hand, the confidence scoring mechanism establishes the relevance of a particular category, as dictated by NLP signals.

The enhanced website categorization products have several uses, and five of them were discussed in this article. These are:

- Content filtering to improve workplace productivity
- Malware protection through content filtering
- Brand protection through third-party risk assessment
- Lead generation
- Fraud detection and prevention

An ongoing and repetitive theme in these applications is that website categorization tools can enhance crucial business security and marketing processes, making them more inclusive and comprehensive. In particular, Website Categorization API could be integrated into fraud detection and prevention platforms, content filtering systems, and malware detection solutions.