Check a Website’s Reputation with Website Categorization API and Other Tools

Posted on October 15, 2019

In an era riddled with highly skilled threat actors and sophisticated attack methods, determining whether you can safely access a website or not is critical. After all, the only certain way of preventing a breach is to keep well away from every potential threat source online, including anywhere inappropriate you may land on the Web.

With that purpose in mind, this long-form article presents four case studies on how Website Categorization API, in conjunction with other domain and IP data feeds and APIs, can help organizations avoid the pitfalls that come with visiting harmful web pages.

Table of Contents

Check Website Trust or Land in a Sea of Trouble

Many system infections result from the simple act of visiting a compromised or outright malicious website. While some sites have been specially crafted to host malware or exploit that automatically

get dropped onto vulnerable systems, others are legitimate but have been under the control of hackers and/or used in attacks. Whatever the case may be, these malware- or exploit-laden sites often employ the same tactic to rein in victims — a drive-by download.

So how does this work? What happens once a piece of malware makes your computer its new home? Cyber attacks that rely on drive-by downloads often use these elements:

- **Entry point:** Attackers create credible sites or hijack popular ones to act as malware hosts. These pages are designed to silently drop a piece of malware onto unsuspecting visitors' systems.

- **Distribution and exploration:** The malware that initially ends up on a user's computer is designed to pinpoint exploitable vulnerabilities on it, its apps, and the devices connected to it. You can think of it as an added reconnaissance tool to reach the attacker's end goal.

- **Exploitation and infection:** After a comprehensive diagnosis of the infected system, the initial malware identifies the exploit that would work best on it. Attackers typically have commercially available (from underground marketplaces or the Deep Web) exploit kits in their attack arsenals. From there, cybercriminals choose what would run on a victim's computer and drop this onto it to continue the attack. In turn, the attackers gain control of one or more devices and so initiate the loss of data and related breach.

- **Infrastructure hacking:** Every attack designed to siphon off confidential data from an infected system makes use of a command-and-control (C&C) server owned and controlled by the attackers. In a ransomware attack, for instance, the C&C server issues commands to the actual payload such as "look for files with the .doc extension, copy then, and send the copies back."

What can companies do to prevent such malicious incidents from happening? While it may be impossible to tell if a website is a potential threat carrier at first glance, avoiding those that have been identified as unsafe to visit is highly recommended. Blacklisting sites so even the most reckless employees won't end up visiting them is good practice too.

Additionally, the aid of a Website Categorization API and other tools configured to adhere to

security policies can help organizations safeguard their digital assets against e-commerce formjacking, phishing, and ransomware attacks. Let’s see how with the following use cases.

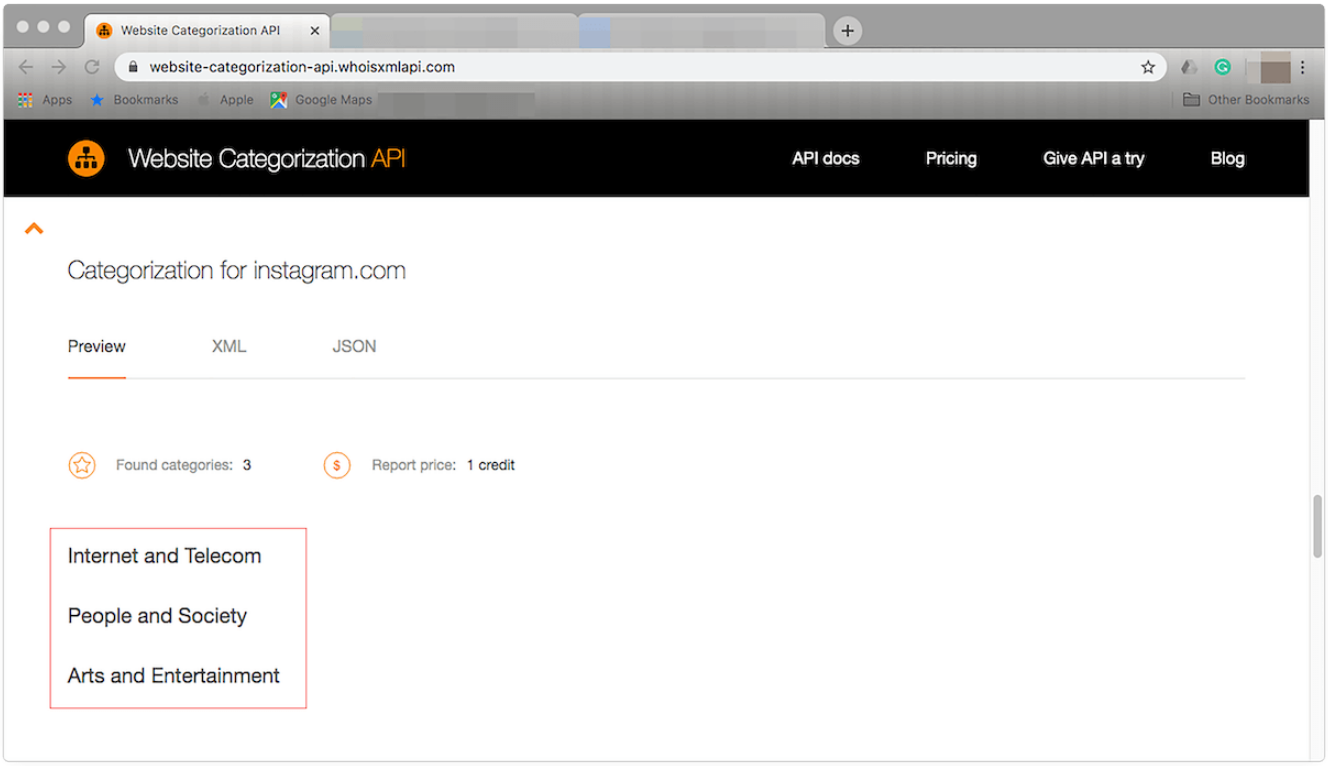
Case Studies

Case Study 1: E-Commerce (Magecart)

Before digging into this first case study, let’s take a look at what website categorization entails. In short, website classification is an easy way for businesses to get to know their customers as well as to flag inconsistencies and potential cases of fraud.

In fact, in three simple steps, users would already know more about a specific customer or potential threat source. For instance:

- Log in to https://website-categorization-api.whoisxmlapi.com.
- Click Give API a Try. You should land on the desired section.
- Type the domain name into the Get website categorization input field and hit the Enter key. Up to three categories to which the user’s site belongs should appear.



These categories are just a few examples among many. The API currently has 25 categories that include Autos and Vehicles, Beauty and Fitness, and more. If the categories you're interested in aren't on the current list, you can send WhoisXML API a request.

So what do you do this with data? Pooling the customers' domains into categories can help sales and marketing teams identify which industries to prioritize. They can then come up with informed strategies that would yield a more significant profit margin for their companies.

Not all site visitors are prospective customers, however. And neither should everyone be welcome with open arms. If you're a cybersecurity specialist, you already know that businesses must be wary of threat actors that want to gain access to their networks and prey on their customers. This is illustrated by one threat in particular — e-commerce formjacking.

E-commerce formjacking, which we explored by using Magecart attacks, involves implanting malicious code into the forms that online buyers fill in when placing their orders on online shops. This code allows attackers to steal users' credit card information as they input it into the checkout page.

Importantly, Magecart refers to an attack category, that is, e-commerce formjacking, and not a specific organization or entity. Several cybercriminal groups have used Magecart in high-profile attacks. like those that targeted British Airways, Ticketmaster, and Newegg.

Magecart attacks use a malicious JavaScript code that listens for and collects personal information. Some monitor all of a user's keypresses within a page while others only intercept inputs into specific parts of a form like credit card numbers and card verification values (CVVs). In general, however, all attackers hide the malicious code inside benign-looking code to evade detection.

The latest report on Magecart incidents revealed that more than 17,000 domains fell prey to the threat. Even worse, experts say they see no end to the attacks any time soon. The only course of action left for businesses then is to beef up the security of their e-commerce sites.

Organizations that want to safeguard their infrastructure and customers can use their website categorization findings with reports that give out indicators of compromise (IoCs) to identify unwanted site visitors. URL blacklisting, in the event of a potential e-commerce formjacking attempt, would help users safeguard their digital properties and customer data from malicious individuals.

Case Study #2: Brand Protection (PayPal)

Web categorization is a worthwhile endeavor when it comes to brand protection. For any business to succeed, its brand always has to be reputable. Succumbing to a cyber attack can leave a lasting negative impression on a company's existing and potential customers.

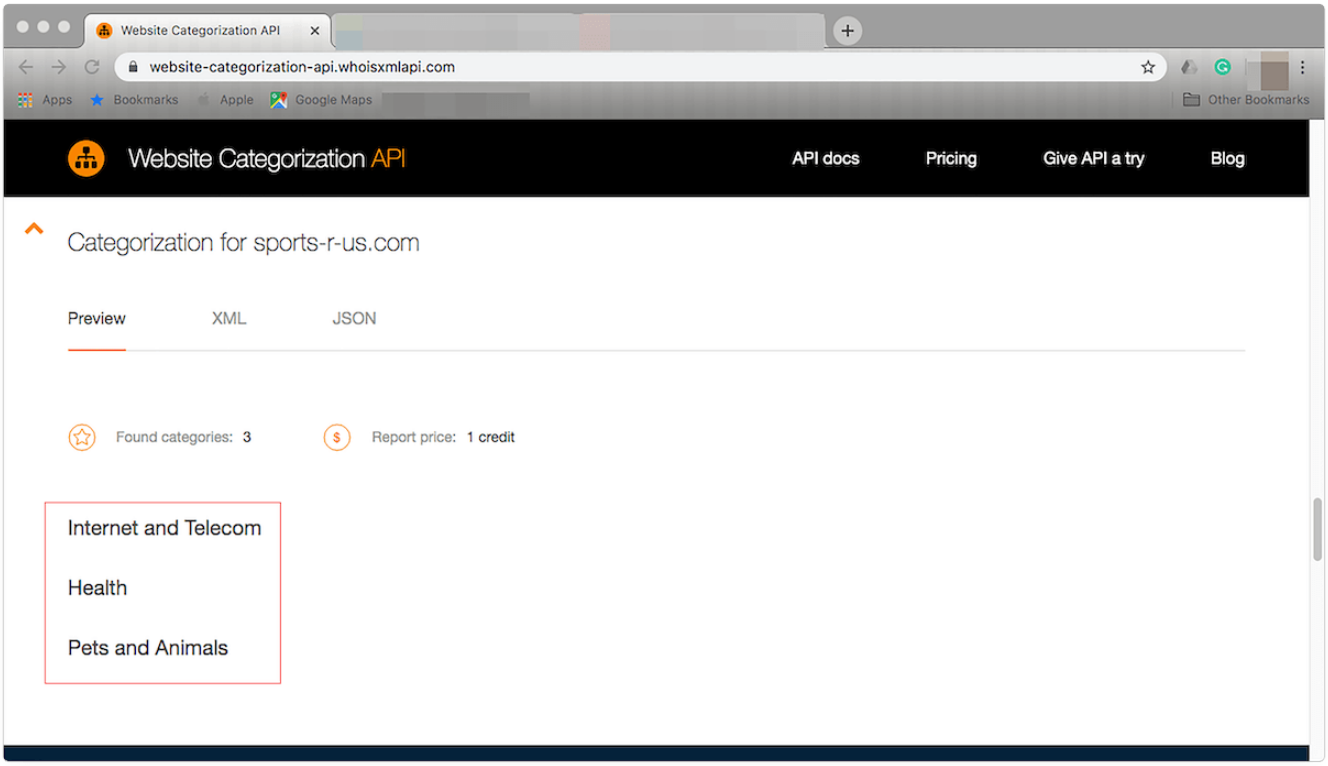
Phishing, for example, is an age-old but ever-reliable threat that remains one of the most significant challenges even for today's biggest brands. While lesser-known targets also struggle with it, it isn't surprising that the most popular vendors comprise phishers' list of go-to targets. There's a straightforward reason for that — the bigger the brand, the wider the potential victim pool is and the profit margin for attackers.

If you're wondering about the prevalence of phishing in actual numbers, the latest Anti-Phishing Working Group (APWG) Phishing Activity Trends Report showed a constant increase in the number of unique phishing websites month-over-month — from 48,663 in January 2019 to 50,983 in February to 81,122 in March. Verizon's 2019 Data Breach Investigations Report (DBIR) identified phishing as the leading data breach attack vector. Another report said one in every 99 messages is a phishing email. These are alarming trends, but the danger can usually be avoided by employing tools that add an extra layer of defense against the threat.

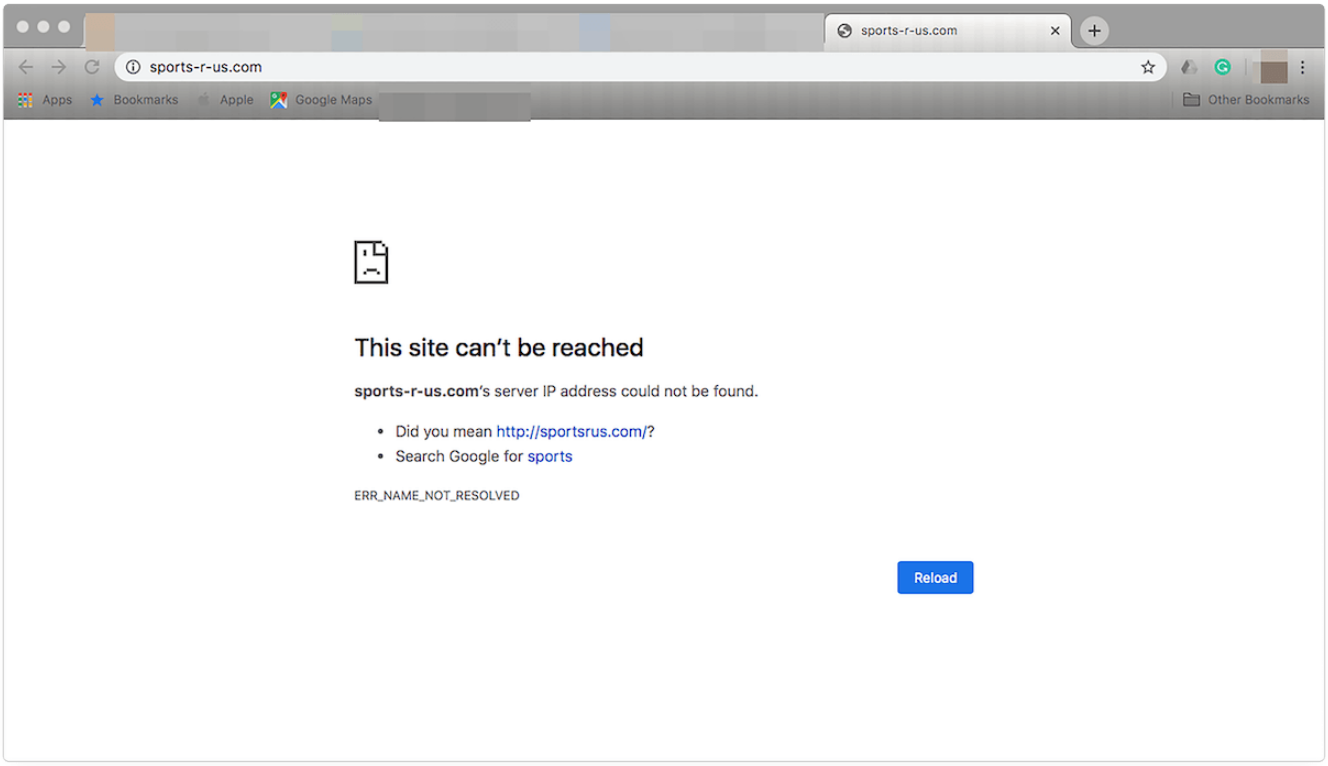
Like in the e-commerce case study, users can rely on Website Categorization API to determine if a potential client, for instance, is worth trusting or should potentially be flagged as a "phisher." A payment processor like PayPal can follow the same three steps to verify if a user's claims are valid. It can search for the client's domain to check if it corresponds to the email sender's supposed company category. Calling the organization for confirmation, of course, enhances the verification process.

Let's take a look at a hypothetical scenario. Say that John Smith wishes to sign up for a PayPal account to start his new business. He claims to have recently put up an online shopping site that sells sports apparel called Sports R Us. (**IMPORTANT NOTE:** We used a randomly chosen domain name for this scenario. That domain name is not malicious.)

The PayPal representative in charge of John Smith's account registration can look up sports-r-us.com (his domain) on Website Categorization API to verify its existence. To do that, type the domain name into the "Get website categorization input" field and hit the Enter key. A list of the categories the customer's site belongs to should appear.



For our made-up scenario, the results don't necessarily confirm John Smith's claims. You may need to employ further Internet research by visiting the said site and seeing if it is indeed an online shopping site for sports apparel. Our search for the domain shows this:



A result like that makes John Smith's claims about putting up a sports goods shopping site less credible. If you do happen to land on an active site, you can check for visible signs of credibility. For instance:

- The URL should start with HTTPS instead of the usual HTTP. The additional S at the end means the site is encrypted and is thus harder to compromise than one that isn't. A lock icon preceding the URL also indicates the website security.

- The presence of a website privacy policy is also a good indication of a site's reliability. That means its owner adheres to the stipulations of data privacy laws such as the General Data Protection Regulation (GDPR).

- Every reputable company provides accurate contact information on its website. You can

check these out by emailing or calling them.

- Some vendors even go the extra mile and have their sites verified by certified authorities to guarantee that they look out for their customers' welfare.

- No reputable vendor site has third-party ads that offer things for free. Even an accidental hover or click on a malicious ad can lead to a drive-by download.

In light of these approaches of checking a website reputation, financial service providers like PayPal can support or reject an applicant's account registration. These are just some of the ways in which they can avoid being abused by a potential phisher.

Cybersecurity professionals can also opt for specially designed tools to check the validity of a site's Secure Sockets Layer (SSL) certificate. All reputable vendors' sites have valid SSL certificates which digitally bind a cryptographic key to their organizations.

Finally, when complemented by brand protection and monitoring tools, IT security teams and employees can avoid dealings with malicious individuals and engage in unintended interactions that may have severe repercussions for their reputation.

Case Study #3: Cybersecurity Against Ransomware (Businesses in Major U.S. Cities)

We've seen many organizations worldwide lose massive amounts of data and incur huge financial losses after suffering a ransomware attack. A Florida city is likely to be holding the record for shelling out the biggest ransomware payout amounting to US$600,000 to date. It gave in to the hackers' demand when it lost access to all its records and when its email system was disabled. What's more, it had to resort to paying employees and vendors by check, and its 911 dispatchers were left unable to pass on calls to the responders.

Another example of a ransomware attack involved Maersk, a global container shipping giant, in one of the most prominent casualties of the NotPetya outbreak in June 2017. The company

reportedly lost an estimated US$300 million due to the attack that resulted in a severe business interruption across 600 of its sites located in 130 countries.

Ransomware isn't just a problem for large enterprises, though. Small and medium-sized businesses (SMBs) are also prone to attack. Take the case of a local medical service provider in Michigan, for instance, whose owners preferred to close shop rather than deal with an attack's aftereffects. Not paying the ransom is justifiable as those who opt to sometimes end up with nothing but a gaping hole in their bank accounts.

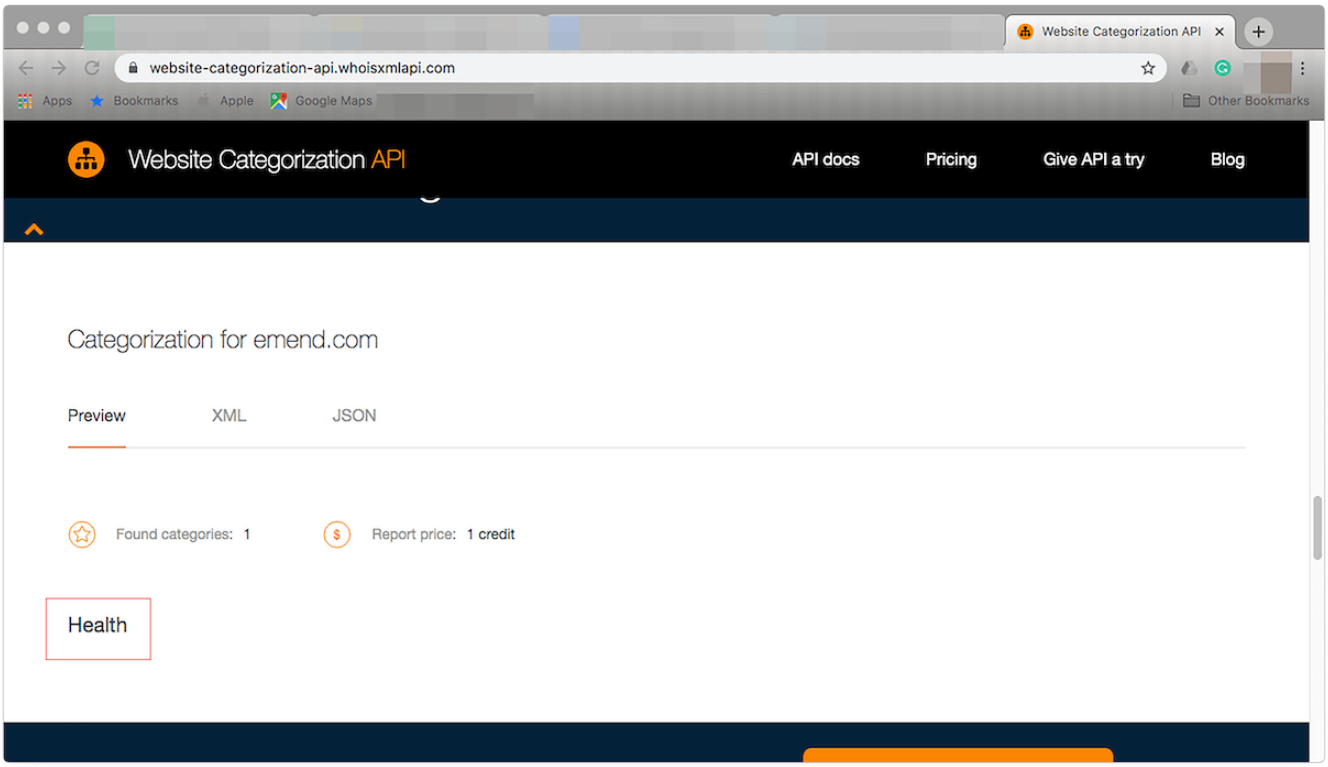
These aren't isolated cases. Security experts believe that the ransomware damage could reach as much as US$11.5 billion this year. Reasons for this include an expected rise in attack frequency and code innovations.

In light of recently reported events in major cities across the U.S. alone, we're bound to see the prediction come true. Local city halls, public libraries, and other government offices in Dallas, Baltimore, Albany, and Laredo in Texas and Lake City, Florida were just some of the recent victims.

As with any online threat, one way of countering the ill effects of a ransomware attack lies in identifying risky sites. To do that, organizations can determine where their site visitors originate from. Website categorization and IP geolocation can work hand in hand to enable that. For instance, a website categorization API can help assess if anyone under a specific domain (and its related categories) has a legitimate reason to be accessing the company website.

Let's say, for demonstration purposes, that you work as an IT security personnel for the Dallas Public Library. Because of the recent spate of attacks against similar institutions, you decided to sift through your network's traffic logs. While at it, you discover a suspicious domain such as emend.com that keeps trying to access your network. (**IMPORTANT NOTE:** The domain in this scenario isn't necessarily related to the attacks discussed in this section. It has been randomly chosen and does not have to be malicious.)

A Website Categorization API lookup using the domain as a search term should give you thisresult:

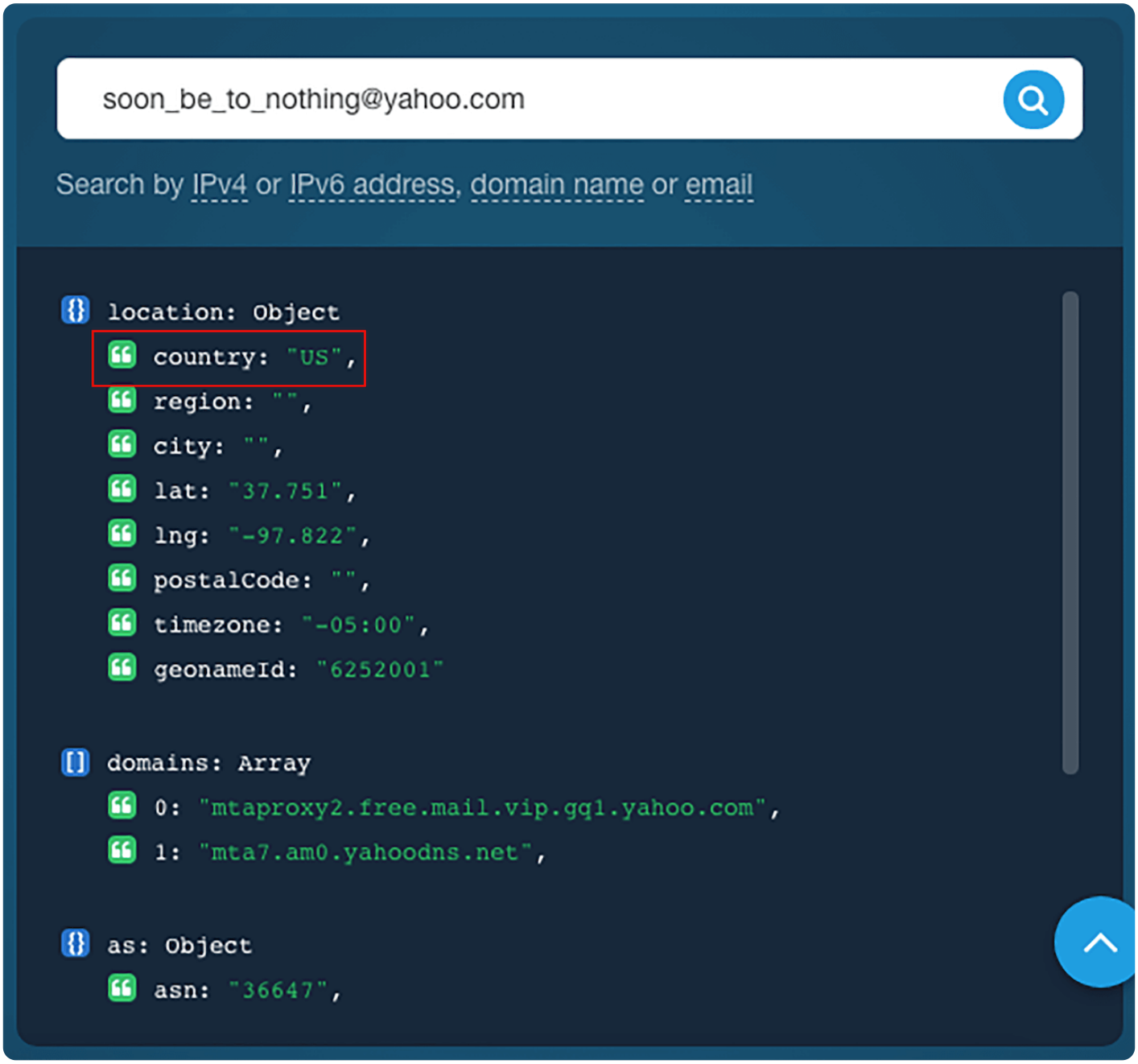


The word “emend,” according to a dictionary, means “to correct usually by text alterations.” You should probably wonder why it was classified under the Health category. It may be a good idea to dig deeper into the site. Accessing the site does take you to a healthcare provider’s site, so it’s safe to visit so long as it doesn’t appear on any blacklist such as the Abuse.ch Ransomware Tracker Blocklist. To more safely navigate the Web, including all sites in the said blocklist in your organization’s own URL blacklist is a very good idea. That way, you can avoid landing on a known ransomware-laden site.

In the case of the U.S. cities, it may be a good idea to spot inconsistencies between say what's claimed in an email and the information contained in IP addresses. For that, you can use an IP geolocation tool.

For example, let's say that your company received an email with a suspicious attachment that claims to originate from a partner in Canada. Keep in mind that a lot of ransomware variations could come in the guise of documents. You can enter the email sender's domain name into the API's search field to determine the message's real source.

For demonstration purposes, let's say the email address soon_be_to_nothing@yahoo.com was used. The IP Geolocation API should give you this result:



As you can see, the inbox isn't registered in Canada. While this alone doesn't justify distrusting the sender, you might want to dig further. You could, for example, check for website category before clicking URLs and potentially end up on unknown sites containing malware.

Case Study #4: Content Filtering

We briefly mentioned URL blacklisting as a means to prevent an e-commerce form jacking attempt. An even broader approach than URL blacklisting, though, is content filtering. Content filtering is the practice of restricting a user's Internet access to predetermined types of content only. A typical office scenario would likely allow employees to visit news and educational websites, while they would be prohibited from accessing social media, shopping, gambling, and adult websites. The protocol may vary depending on the company and industry, but this is usually the standard scenario.

Companies can effectively implement content filtering with the use of the Website Categorization API. However, a more secure and cybersecurity-focused approach would also require companies to use additional tools that allow them to glean more information about a particular website.

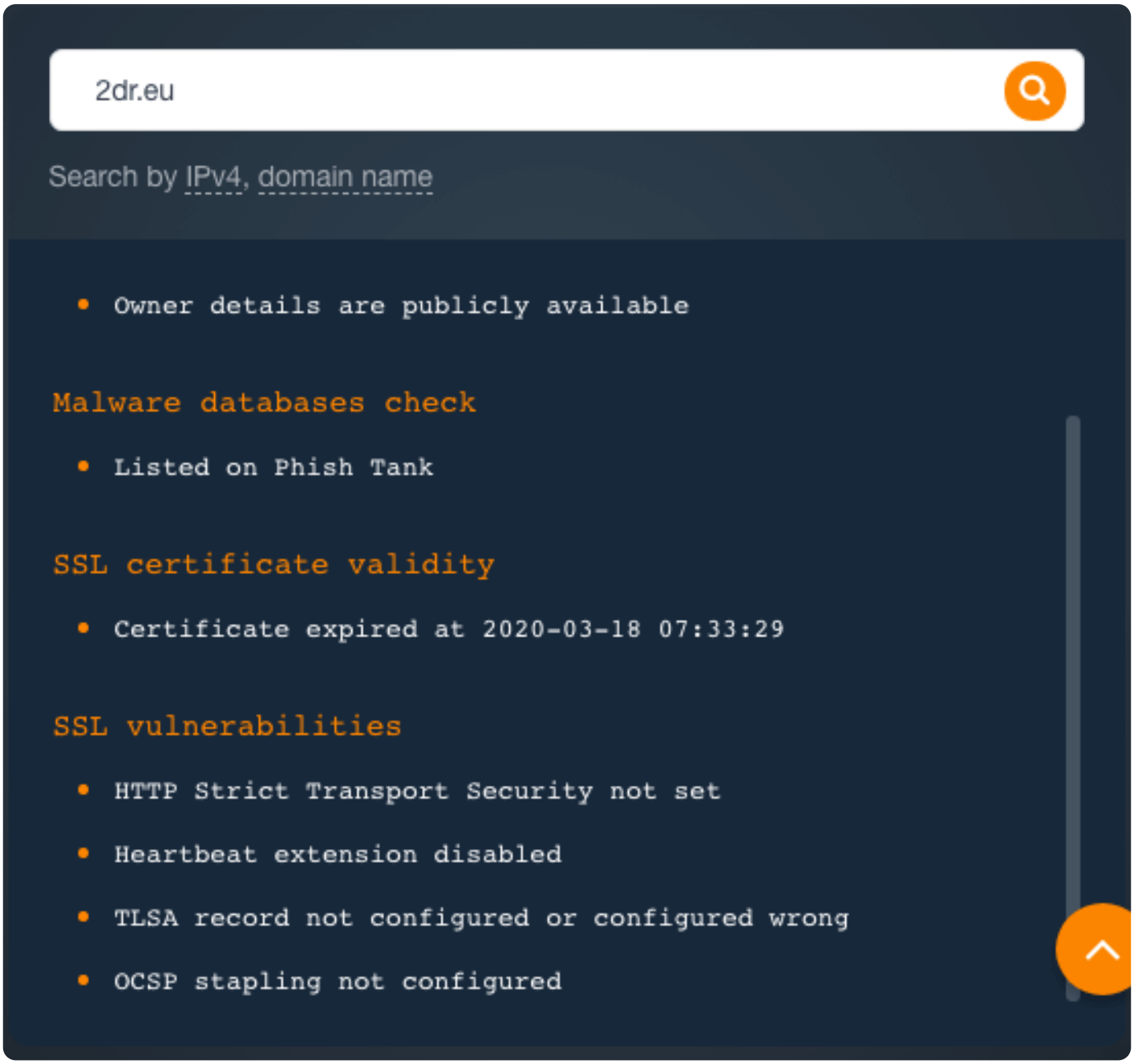
Consider this hypothetical scenario where an office staff member tries to access the following websites. Note that this employee may knowingly type these sites' URLs into his/her browser or be taken to the pages after clicking a link embedded in an email, chat message, or an ad. In other cases, the employee may land on these websites via redirection when he/she opens an email attachment or a transferred file.

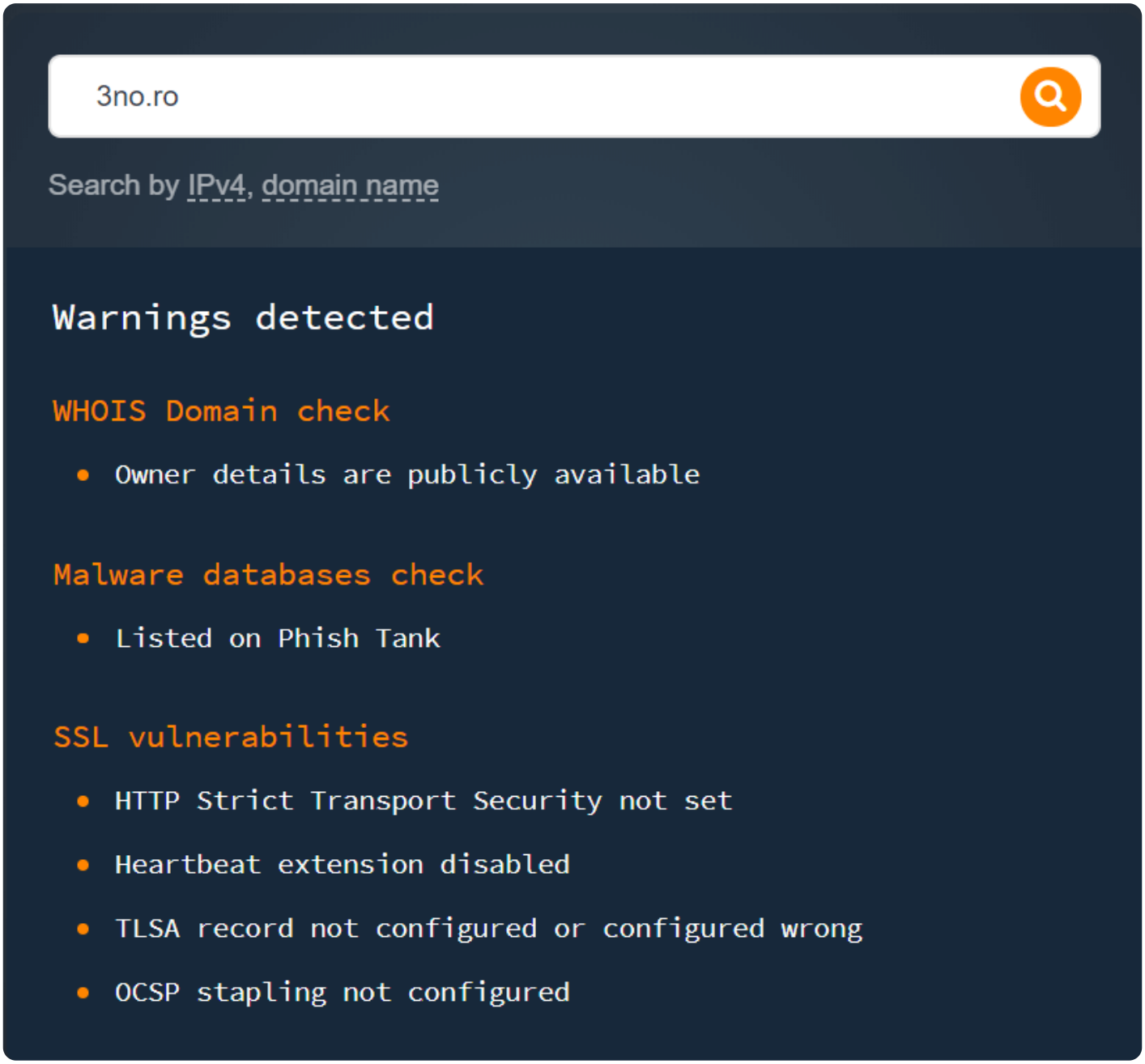
- **2dr[.]eu:** Categorized under "Internet and Telecom," and "Computer and Electronics" by the API.
- **3no[.]ro:** Categorized under "Internet and Telecom," "Health," and "Pets and Animals" by the API.
- **rr[.]co:** Categorized under "Computer and Electronics" by the API.
- **betonline[.]ag:** Categorized under "Gambling" by the API.

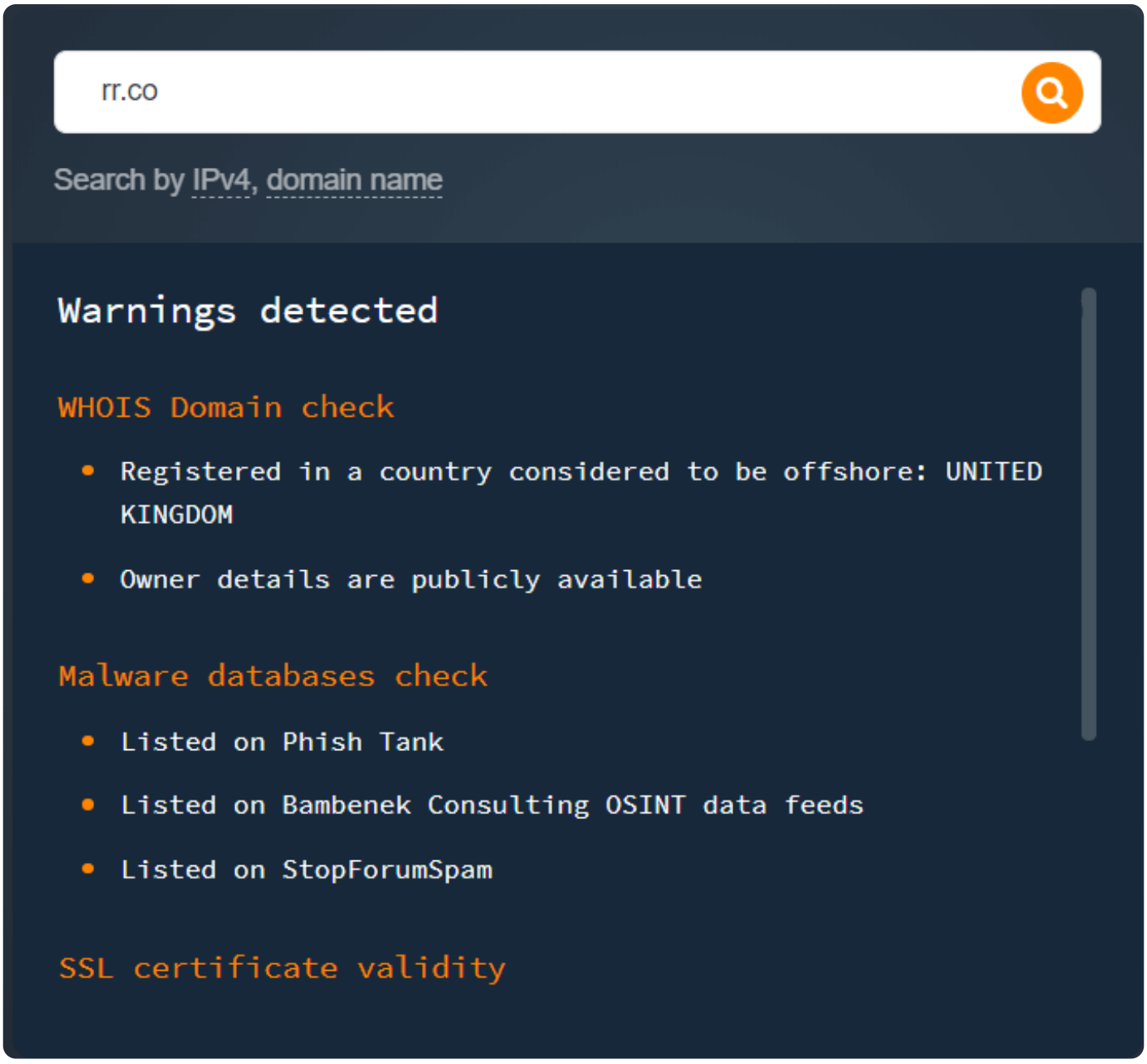
The standard content filtering protocol would immediately block the employee's access to betonline[.]ag since it is considered a gambling site. The other three websites, on the other hand,

could be permitted, since they do not belong to the blacklisted categories. As such, the office staff can access the first three domains. But are these websites safe? Let's find out by adding one more step—a domain reputation check.

In our hypothetical example, the first two domain names, 2dr[.]eu and 3no[.]ro, are actually listed as phishing sites on PhishTank. Rr[.]co, on the other hand, is not only listed on PhishTank but also on the Bambenek Consulting OSINT data feeds and StopForumSpam. A quick domain reputation check using Domain Reputation API would reveal as much.


2dr.eu
Search by IPv4, domain name
Owner details are publicly available
Malware databases check
Listed on Phish Tank
SSL certificate validity
Certificate expired at 2020-03-18 07:33:29
SSL vulnerabilities
HTTP Strict Transport Security not set
Heartbeat extension disabled
TLSA record not configured or configured wrong
OCSP stapling not configured

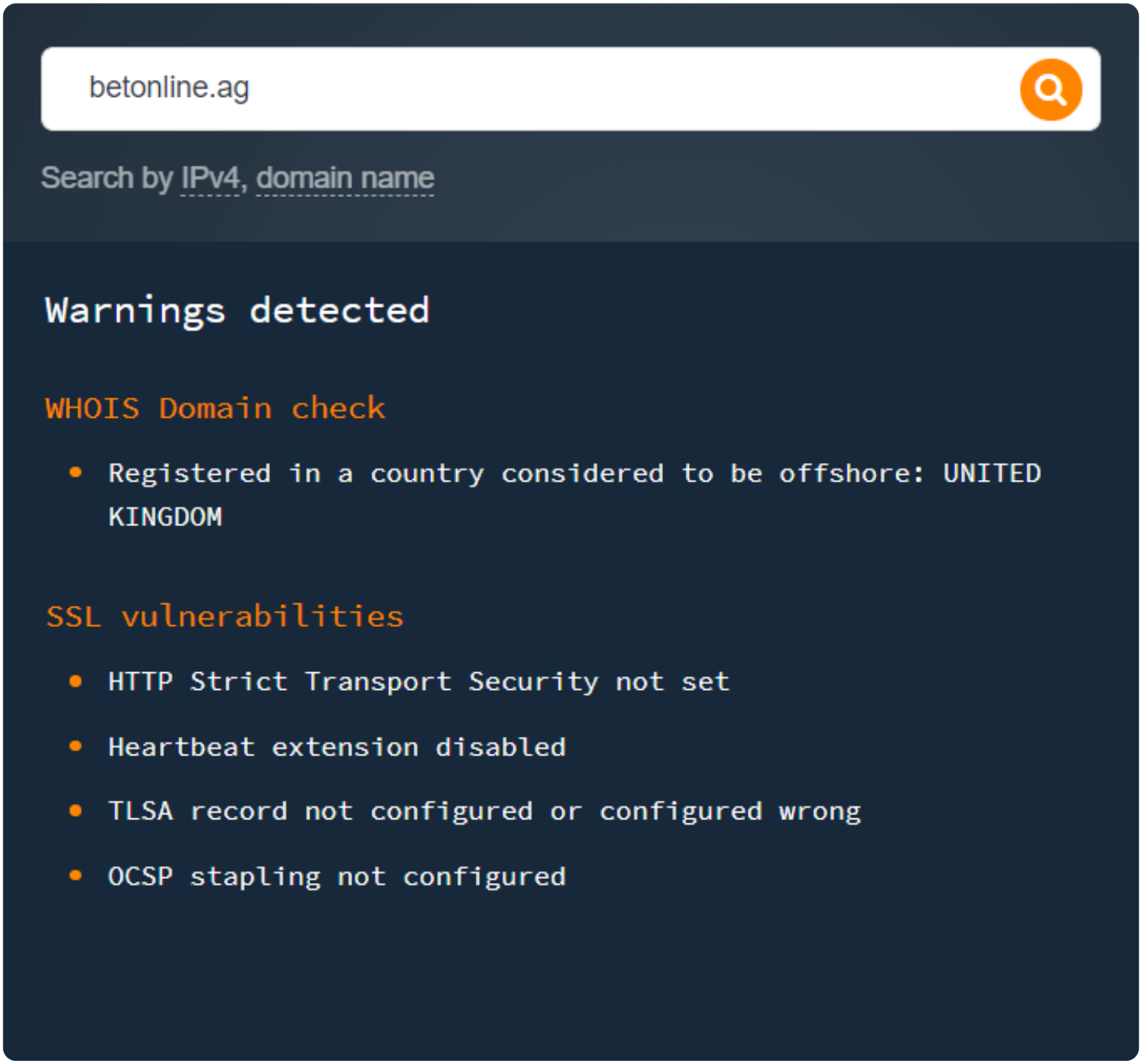

3no.ro
Search by IPv4, domain name
Warnings detected
WHOIS Domain check
Owner details are publicly available
Malware databases check
Listed on Phish Tank
SSL vulnerabilities
HTTP Strict Transport Security not set
Heartbeat extension disabled
TLSA record not configured or configured wrong
OCSP stapling not configured



The three websites are hosted on known malicious domains on various malware data feeds. As such, although they fall under whitelisted categories, they are not safe to visit. As an additional layer of protection to blacklisting website categories, it may also be a good idea to block access to websites that are listed on any threat feed aided by a domain reputation checker.

Some may argue that since the domain reputation check already reveals many things about a website, the categorization process could be skipped altogether. However, that is not the case. In our hypothetical scenario, for instance, the office staff member would be immediately be prevented from accessing betonline[.]ag since the domain falls under “Gambling.” If you ignore the categorization step but rely on a domain reputation check, access to the said website could be granted due to the following reasons:

- Betonline[.]ag is not listed on any malware data feed and so it would be deemed safe to visit.
- Although the API detected that the domain is registered in an offshore country, the country (United Kingdom) is not considered a hotbed for threats.



In the end, the employee could access the gambling site if organizations fail to set up website categorization and solely rely only on domain reputation. Website categorization automatically weeds out sites that though are safe to access can hamper employee productivity and use up much-needed resources, which could negatively impact the company's bottom line in the long term.

Bottom line: Check Website Reputation & Trust with the Right Intelligence Sources

Fortifying one's network against e-commerce formjacking, phishing, ransomware, and other cyber attacks requires careful scrutiny of who website visitors are and where they originate. Organizations need to know which websites to trust and which ones should be avoided.

Security solutions are great at preventing malicious files from being executed on vulnerable systems, but not all can distinguish between safe from damaging traffic. Bolstering their capability to distinguish between malicious and non malicious site visits is possible with Website Categorization API and other domain and IP data feeds and APIs — providing for more proactive defense.

Indeed, by preventing malicious individuals from interacting with your network and dropping unwanted files into gaping holes, organizations can effectively stop attacks before they even take root.