WIKIPEDIA

HTTP 404

The **HTTP 404**, **404 Not Found**, **404**, **404 Error**, **Page Not Found**, **File Not Found**, or **Server Not Found** error message is a Hypertext Transfer Protocol (HTTP) standard response code, in computer network communications, to indicate that the browser was able to communicate with a given server, but the server could not find what was requested. The error may also be used when a server does not wish to disclose whether it has the requested information.[1]

The website hosting server will typically generate a "404 Not Found" web page when a user attempts to follow a broken or dead link; hence the 404 error is one of the most recognizable errors encountered on the World Wide Web.

Contents

Overview

When communicating via HTTP, a server is required to respond to a request, such as a web browser request for a web page, with a numeric response code and an optional, mandatory, or disallowed (based upon the status code) message. In code 404, the first digit indicates a client error, such as a mistyped Uniform Resource Locator (URL). The following two digits indicate the specific error encountered. HTTP's use of three-digit codes is similar to the use of such codes in earlier protocols such as FTP and NNTP. At the HTTP level, a 404 response code is followed by a human-readable "reason phrase". The HTTP specification suggests the phrase "Not Found"[2] and many web servers by default issue an HTML page that includes both the 404 code and the "Not Found" phrase.

A 404 error is often returned when pages have been moved or deleted. In the first case, it is better to employ URL mapping or URL redirection by returning a 301 Moved Permanently response, which can be configured in most server configuration files, or through URL rewriting; in the second case, a 410 Gone should be returned. Because these two options require special server configuration, most websites do not make use of them.

404 errors should not be confused with DNS errors, which appear when the given URL refers to a server name that does not exist. A 404 error indicates that the server itself was found, but that the server was not able to retrieve the requested page.

Soft 404 errors

Some websites report a "not found" error by returning a standard web page with a "200 OK" response code, falsely reporting that the page loaded properly; this is known as a *soft 404*. The term "soft 404" was introduced in 2004 by Ziv Bar-Yossef *et al.*[3]

Soft 404s are problematic for automated methods of discovering whether a link is broken. Some search engines, like Yahoo and Google, use automated processes to detect soft 404s.[4] Soft 404s can occur as a result of configuration errors when using certain HTTP server software, for example with the Apache software, when an Error Document 404 (specified in a .htaccess file) is specified as an absolute path (e.g. http://example.com/error.html) rather than a relative path (/error.html).[5] This can also be done on purpose to force some browsers (like Internet Explorer) to display a customized 404 error message rather than replacing what is served with a browser-specific "friendly" error message (in Internet Explorer, this behavior is triggered when a 404 is served and the received HTML is shorter than a certain length, and can be manually disabled by the user).

There are also "soft 3XX" errors where content is returned with a status 200 but it comes from a redirected page, such as when missing pages are redirected to the domain root/home page.

Proxy servers

Some proxy servers generate a 404 error when a 500-range error code would be more correct. If the proxy server is unable to satisfy a request for a page because of a problem with the remote host (such as hostname resolution failures or refused TCP connections), this should be described as a 5xx Internal Server Error, but might deliver a 404 instead. This can confuse programs that expect and act on specific responses, as they can no longer easily distinguish between an absent web server and a missing web page on a web server that is present.

Intentional 404s

In July 2004, the UK telecom provider BT Group deployed the Cleanfeed content blocking system, which returns a 404 error to any request for content identified as potentially illegal by the Internet Watch Foundation.[6] Other ISPs return a HTTP 403 "forbidden" error in the same circumstances.[7] The practice of employing fake 404 errors as a means to conceal censorship has also been reported in Thailand[8] and Tunisia.[9] In Tunisia, where censorship was severe before the 2011 revolution, people became aware of the nature of the fake 404 errors and created an imaginary character named "Ammar 404" who represents "the invisible censor".[10]

Microsoft Internet Server 404 substatus error codes

The webserver software developed by Microsoft, Microsoft's Internet Information Services (IIS), returns a set of substatus codes with its 404 responses. The substatus codes take the form of decimal numbers appended to the 404 status code. The substatus codes are not officially recognized by IANA and are not returned by non-Microsoft servers.

Substatus codes

Microsoft's IIS 7.0, IIS 7.5, and IIS 8.0 servers define the following HTTP substatus codes to indicate a more specific cause of a 404 error:

- 404.0 – Not found.
- 404.1 – Site Not Found.
- 404.2 – ISAPI or CGI restriction.
- 404.3 – MIME type restriction.
- 404.4 – No handler configured.
- 404.5 – Denied by request filtering configuration.
- 404.6 – Verb denied.
- 404.7 – File extension denied.
- 404.8 – Hidden namespace.
- 404.9 – File attribute hidden.
- 404.10 – Request header too long.
- 404.11 – Request contains double escape sequence.
- 404.12 – Request contains high-bit characters.
- 404.13 – Content length too large.
- 404.14 – Request URL too long.
- 404.15 – Query string too long.
- 404.16 – DAV request sent to the static file handler.
- 404.17 – Dynamic content mapped to the static file handler via a wildcard MIME mapping.
- 404.18 – Query string sequence denied.
- 404.19 – Denied by filtering rule.
- 404.20 – Too Many URL Segments.

Custom error pages

Web servers can typically be configured to display a customised 404 error page, including a more natural description, the parent site's branding, and sometimes a site map, a search form or 404-page widget. The protocol level phrase, which is hidden from the user, is rarely customized. Internet Explorer, however, will not display custom pages unless they are larger than 512 bytes, opting instead to display a "friendly" error page.[11] Google Chrome included similar functionality, where the 404 is replaced with alternative suggestions generated by Google algorithms, if the page is under 512 bytes in size. Another problem is that if the page does not provide a favicon, and a separate custom 404-page exists, extra traffic and longer loading times will be generated on every page view.[12][13]

The Wikimedia 404 message

Many organizations use 404 error pages as an opportunity to inject humor into what may otherwise be a serious website. For example, Metro UK shows a polar bear on a skateboard, and the web development agency Left Logic has a simple drawing program.[14] During the 2015 UK general election campaign the main political parties all used their 404 pages to either take aim at political opponents or show relevant policies to potential supporters.[15] In Europe, the NotFound project, created by multiple European organizations including Missing Children Europe and Child Focus, encourages site operators to add a snippet of code to serve customized 404 error pages[16] which provide data about missing children.[17]

While many websites send additional information in a 404 error message—such as a link to the homepage of a website or a search box—some also endeavor to find the correct web page the user wanted. Extensions are available for some content management systems (CMSs) to do this.[18]

Tracking 404 errors

A number of tools exist that crawl through a website to find pages that return 404 status codes. These tools can be helpful in finding links that exist within a particular website. The limitation of these tools is that they only find links within one particular website, and ignore 404s resulting from links on other websites. As a result, these tools miss out on 83% of the 404s on websites.[19] One way around this is to find 404 errors by analyzing external links.[20]

One of the most effective ways to discover 404 errors is by using Google Search Console, Google Analytics or crawling software.

Another common method is tracking traffic to 404 pages using log file analysis.[21] This can be useful to understand more about what 404s users reached on the site. Another method of tracking traffic to 404 pages is using JavaScript-based traffic tracking tools.[22]

See also

- Blue screen of death
- Funky caching
- Link rot
- List of HTTP status codes

References

1. "RFC 7231, HTTP/1.1 Semantics and Content, Section 6.5.4 404 Not Found" (http://tools.ietf.org/html/rfc7231#section-6.5.4). ietf.org. June 2014. Retrieved 13 December 2018.
2. "RFC 7231, HTTP/1.1 Semantics and Content, Section 6.5.4 404 Not Found" (http://tools.ietf.org/html/rfc7231#section-6.5.4). ietf.org. June 2014. Retrieved 13 December 2018.
3. Ziv Bar-Yossef; Andrei Z. Broder; Ravi Kumar; Andrew Tompkins (2004). *Sic Transit Gloria Telae: Towards an Understanding of the Web's Decay*. *Proceedings of the 13th International Conference on World Wide Web (WWW)*. pp. 328–337. doi:10.1145/988672.988716 (https://doi.org/10.1145%2F988672.988716). ISBN 978-1581138443.
4. "Why is your crawler asking for strange URLs that have never existed on my site?" (https://web.archive.org/web/20140715000803/http://help.yahoo.com/l/nl/yahoo/ysearch/slurp/slurp-10.html). Yahoo Ysearch Help page. Archived from the original (http://help.yahoo.com/l/nl/yahoo/ysearch/slurp/slurp-10.html) on 15 July 2014. Retrieved 4 September 2013.
5. "Farewell to soft 404s" (http://googlewebmastercentral.blogspot.com/2008/08/farewell-to-soft-404s.html). Google Official Blog. Retrieved 20 September 2008.
6. "LINX Public Affairs » Cleanfeed: the facts" (https://web.archive.org/web/20110513071912/https://publicaffairs.linx.net/news/?p=154). Publicaffairs.linx.net. 10 September 2004. Archived from the original (https://publicaffairs.linx.net/news/?p=154) on 13 May 2011. Retrieved 6 March 2011.
7. "DEMON – Error 403" (https://airbrake.io/blog/http-errors/403-forbidden-error). Retrieved 14 June 2012.

8. Sambandaraksa, Don (18 February 2009). "The old fake '404 Not Found' routine - Dead link" (http://www.bangkokpost.com/tech/technews/11872/the-old-fake-404-not-found-routine). *Bangkok Post*. Retrieved 12 September 2010.

9. Noman, Helmi (12 September 2008). "Tunisian journalist sues government agency for blocking Facebook, claims damage for the use of 404 error message instead of 403" (http://opennet.net/node/950). Open Net Initiative. Retrieved 21 November 2010.

10. "Anti-censorship movement in Tunisia: creativity, courage and hope!" (http://advocacy.globalvoicesonline.org/2010/05/27/anti-censorship-movement-in-tunisia-creativity-courage-and-hope/). Global Voices Advocacy. 27 May 2010. Retrieved 28 August 2010.

11. "Friendly HTTP Error Pages" (https://web.archive.org/web/20101202044514/http://blogs.msdn.com/b/ieinternals/archive/2010/08/19/http-error-pages-in-internet-explorer.aspx). msdn.com. 18 August 2010. Archived from the original (http://blogs.msdn.com/b/ieinternals/archive/2010/08/19/http-error-pages-in-internet-explorer.aspx) on 2 December 2010. Retrieved 14 June 2012.

12. Heng, Christopher (7 September 2008). "What is Favicon.ico and How to Create a Favicon Icon for Your Website" (http://www.thesitewizard.com/archive/favicon.shtml). *The Site Wizard*. Retrieved 23 February 2011.

13. "The Dastardly "favicon.ico not found" Error" (http://internetfolks.com/dastardly-favicon-ico-found-error/). *Internet Folks*. 3 August 1999.

14. "From skateboarding bears to missing children: The power of the 404 Not Found error page" (http://metro.co.uk/2013/04/16/from-skateboarding-bears-to-missing-children-the-power-of-the-404-not-found-error-page-3591031/). Metro. 6 June 2011. Retrieved 16 April 2013.

15. "The political Page 404 war" (http://www.bbc.co.uk/newsbeat/article/32480181/the-political-page-404-war). *BBC Newsbeat*. 27 April 2015. Retrieved 18 May 2018.

16. "Notfound.org" (https://web.archive.org/web/20140902081543/http://notfound.org/notfound). *notfound*. notfound. Archived from the original (http://notfound.org/notfound) on 2 September 2014.

17. "Missing children messages go on 404 error pages" (https://www.bbc.com/news/technology-19741350). *BBC News*. 27 September 2012. Retrieved 20 September 2014.

18. Swenson, Sahala (19 August 2008). "Make your 404 pages more useful" (http://googlewebmastercentral.blogspot.com/2008/08/make-your-404-pages-more-useful.html). *Official Google Webmaster Central Blog*. Google, Inc. Retrieved 28 August 2009.

19. "Sources Leading To 404s" (https://springtrax.com/key-stats-about-404-errors/default.htm#section4). SpringTrax. Retrieved 11 February 2013.

20. Cushing, Anne (2 April 2013). "A Data-Centric Approach To Identifying 404 Pages Worth Saving" (http://searchengineland.com/a-data-centric-approach-to-identifying-404-pages-worth-saving-152844). Search Engine Land. Retrieved 7 June 2013.

21. "Tracking and Preventing 404 Errors" (http://www.404errorpages.com/#tracking). 404errorpages.com. Retrieved 7 June 2013.

22. "Understand 404 Errors" (https://springtrax.com/how-springtrax-helps/understand-your-websites-404-errors.htm). SpringTrax.com. Retrieved 7 June 2013.

External links

- A More Useful 404 (http://alistapart.com/article/amoreuseful404)
- 404 Not Found (http://tools.ietf.org/html/rfc7231#section-6.5.4) of the *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content* specification, at the Internet Engineering Task Force
- ErrorDocument Directive (http://httpd.apache.org/docs/2.4/mod/core.html#errordocument) – instructions on custom error pages for the Apache 2.0 web server
- 404: Not Found (http://www.ovff.org/pegasus/songs/404-not-found.html) – an award-winning song about the error code