```
In [1]:
       #reference:https://realpython.com/python-web-scraping-practical-introduction/
       #python -m pip install beautifulsoup4
       #python -m pip show beautifulsoup4
       from bs4 import BeautifulSoup as bs
       from urllib.request import urlopen
       url2 = "https://www.w3schools.com/html/html_tables.asp"
       url = "http://olympus.realpython.org/profiles/dionysus"
       url1 = "https://pes.edu/"
       #url2 = "https://www.tutorialspoint.com"
       #returns an httpresponse object
       page = urlopen(url2)
       #To extract the HTML from the page, first use the
       #HTTPResponse object's .read() method, which returns a sequence of bytes.
       #Then use .decode() to decode the bytes to a string using UTF-8:
       html = page.read().decode("utf-8")
       soup = bs(html, "html.parser")
       #print(soup.get_text())
       #soup.find_all("img")
       table = soup.find('table')
       print(table)
      Company
      Contact
      Country
      Alfreds Futterkiste
      Maria Anders
      Germany
      >
      Centro comercial Moctezuma
      Francisco Chang
      Mexico
      Ernst Handel
      Roland Mendel
      Austria
      Island Trading
      Helen Bennett
      UK
      Laughing Bacchus Winecellars
      Yoshi Tannamuri
      Canada
      >
      Magazzini Alimentari Riuniti
      Giovanni Rovelli
      Italy
```

```
import requests
url = "http://en.wikipedia.org/wiki/Robots_exclusion_standard"
page=requests.get(url)
page
```

Out[2]:

<Response [200]>

In [3]:

print(page.content)

b'<!DOCTYPE html>\n<html class="client-nojs" lang="en" dir="ltr">\n<head>\n<meta cha rset="UTF-8"/>\n<title>Robots exclusion standard - Wikipedia</title>\n<script>docume nt.documentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSeparatorTran sformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgM onthNames":["","January","February","March","April","May","June","July","August","Se ptember", "October", "November", "December"], "wgRequestId": "fdd9e463-1043-4ee4-a270-c5f e896fffed", "wgCSPNonce": !1, "wgCanonicalNamespace": "", "wgCanonicalSpecialPageName": ! 1,"wgNamespaceNumber":0,"wgPageName":"Robots_exclusion_standard","wgTitle":"Robots e xclusion standard", "wgCurRevisionId":1043553668, "wgRevisionId":1043553668, "wgArticle Id":101673, "wgIsArticle":!0, "wgIsRedirect":!1, "wgAction": "view", "wgUserName":null, "w gUserGroups":["*"],"wgCategories":["CS1 maint: multiple names: authors list","CS1 er rors: missing periodical", "Articles with short description", "Short description match es Wikidata", "Wikipedia pending changes protected pages", "All articles with vague or ambiguous time",\n"Vague or ambiguous time from December 2020","Official website dif ferent in Wikidata and Wikipedia", "Articles with MA identifiers", "World Wide Web", "I nternet bots"],"wgPageContentLanguage":"en","wgPageContentModel":"wikitext","wgRelev antPageName":"Robots_exclusion_standard","wgRelevantArticleId":101673,"wgIsProbablyE ditable":!0,"wgRelevantPageIsProbablyEditable":!0,"wgRestrictionEdit":[],"wgRestrict ionMove":[],"wgFlaggedRevsParams":{"tags":{"status":{"levels":-1}}},"wgStableRevisio nId":1043553668, "wgMediaViewerOnClick":!0, "wgMediaViewerEnabledByDefault":!0, "wgPopu psFlags":10,"wgVisualEditor":{"pageLanguageCode":"en","pageLanguageDir":"ltr","pageV ariantFallbacks":"en"},"wgMFDisplayWikibaseDescriptions":{"search":!0,"nearby":!0,"w atchlist":!0,"tagline":!1},"wgWMESchemaEditAttemptStepOversample":!1,"wgULSCurrentAu tonym":"English","wgNoticeProject":"wikipedia","wgEditSubmitButtonLabelPublish":! 0,"wgULSPosition":"interlanguage","wgULSisCompactLinksEnabled":!0,\n"wgGENewcomerTas ksGuidanceEnabled":!0,"wgGEAskQuestionEnabled":!1,"wgGELinkRecommendationsFrontendEn abled":!1,"wgCentralAuthMobileDomain":!1,"wgWikibaseItemId":"Q80776"};RLSTATE={"ext. globalCssJs.user.styles":"ready","site.styles":"ready","noscript":"ready","user.styl es":"ready","ext.globalCssJs.user":"ready","user":"ready","user.options":"loadin g","ext.flaggedRevs.icons":"ready","oojs-ui-core.styles":"ready","oojs-ui.styles.ind icators":"ready","mediawiki.widgets.styles":"ready","oojs-ui-core.icons":"ready","ex t.cite.styles":"ready","ext.pygments":"ready","skins.vector.styles.legacy":"read y","ext.flaggedRevs.basic":"ready","ext.visualEditor.desktopArticleTarget.noscrip t":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","wikibase.c lient.init":"ready"};RLPAGEMODULES=["ext.cite.ux-enhancements","site","mediawiki.pag e.ready", "mediawiki.toc", "skins.vector.legacy.js", "ext.flaggedRevs.advanced", "ext.ga dget.ReferenceTooltips","ext.gadget.charinsert",\n"ext.gadget.extra-toolbar-button s", "ext.gadget.refToolbar", "ext.gadget.switcher", "mmv.head", "mmv.bootstrap.autostar t", "ext.popups", "ext.visualEditor.desktopArticleTarget.init", "ext.visualEditor.targe tLoader", "ext.eventLogging", "ext.wikimediaEvents", "ext.navigationTiming", "ext.uls.co mpactlinks", "ext.uls.interface", "ext.cx.eventlogging.campaigns", "ext.centralNotice.g eoIP","ext.centralNotice.startUp","ext.centralauth.centralautologin","ext.growthExpe riments.SuggestedEditSession"];</script>\n<script>(RLQ=window.RLQ||[]).push(function in*/mw.user.tokens.set({"patrolToken":"+\\\\","watchToken":"+\\\\","csrfToken":"+ \\\"});\n});});</script>\n<link rel="stylesheet" href="/w/load.php?lang=en&modu les=ext.cite.styles%7Cext.flaggedRevs.basic%2Cicons%7Cext.pygments%2CwikimediaBadge s%7Cext.uls.interlanguage%7Cext.visualEditor.desktopArticleTarget.noscript%7Cmediawi ki.widgets.styles%7Coojs-ui-core.icons%2Cstyles%7Coojs-ui.styles.indicators%7Cskins. vector.styles.legacy%7Cwikibase.client.init&only=styles&skin=vector"/>\n<scr</pre> ipt async="" src="/w/load.php?lang=en&modules=startup&only=scripts&raw=1 &skin=vector"></script>\n<meta name="ResourceLoaderDynamicStyles" content=""/>\n <link rel="stylesheet" href="/w/load.php?lang=en&modules=site.styles&only=st</pre> yles&skin=vector"/>\n<meta name="generator" content="MediaWiki 1.37.0-wmf.23"/>

\n<meta name="referrer" content="origin"/>\n<meta name="referrer" content="origin-wh en-crossorigin"/>\n<meta name="referrer" content="origin-when-cross-origin"/>\n<meta name="format-detection" content="telephone=no"/>\n<meta property="og:title" content</pre> ="Robots exclusion standard - Wikipedia"/>\n<meta property="og:type" content="websit e"/>\n<link rel="preconnect" href="//upload.wikimedia.org"/>\n<link rel="alternate" media="only screen and (max-width: 720px)" href="//en.m.wikipedia.org/wiki/Robots ex clusion_standard"/>\n<link rel="alternate" type="application/x-wiki" title="Edit thi s page" href="/w/index.php?title=Robots_exclusion_standard&action=edit"/>\n<link</pre> rel="edit" title="Edit this page" href="/w/index.php?title=Robots_exclusion_standard &action=edit"/>\n<link rel="apple-touch-icon" href="/static/apple-touch/wikipedi</pre> a.png"/>\n<link rel="shortcut icon" href="/static/favicon/wikipedia.ico"/>\n<link re l="search" type="application/opensearchdescription+xml" href="/w/opensearch_desc.ph p" title="Wikipedia (en)"/>\n<link rel="EditURI" type="application/rsd+xml" href="//</pre> en.wikipedia.org/w/api.php?action=rsd"/>\n<link rel="license" href="//creativecommon</pre> s.org/licenses/by-sa/3.0/"/>\n<link rel="canonical" href="https://en.wikipedia.org/w iki/Robots_exclusion_standard"/>\n<link rel="dns-prefetch" href="//meta.wikimedia.or</pre> g" />\n<link rel="dns-prefetch" href="//login.wikimedia.org"/>\n</head>\n<body class ="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject mw-editable page-Robot s_exclusion_standard rootpage-Robots_exclusion_standard skin-vector action-view skin -vector-legacy"><div id="mw-page-base" class="noprint"></div>\n<div id="mw-head-bas e" class="noprint"></div>\n<div id="content" class="mw-body" role="main">\n\t\n\t<div id="siteNotice"><!-- CentralNotice --></div>\n\t<div class="mw-i ndicators">\n\t<div id="mw-indicator-pp-autoreview" class="mw-indicator"></div>\n\t</div>\n\t<h1 id="firstHeading" class="firstHeading" >Robots exclusio n standard</h1>\n\t<div id="bodyContent" class="vector-body">\n\t\t<div id="siteSub" class="noprint">From Wikipedia, the free encyclopedia</div>\n\t\t<div id="contentSu b"><div id="mw-fr-revisiontag" class="flaggedrevs_short flaggedrevs_stable_synced pl ainlinks noprint"><div class="flaggedrevs_short_basic"> <span id=\'mw-fr-revisiontoggle\' aria-disabled=\'false\' title=\'show/hide details</pre> \' class=\'fr-toggle-arrow oo-ui-widget oo-ui-widget-enabled oo-ui-indicatorElementindicator oo-ui-indicator-down oo-ui-indicatorElement oo-ui-labelElement-invisible o o-ui-indicatorWidget\'></div>\n<div id="mw-fr-revisiondetails-wrapper" style ="position:relative;"><div id="mw-fr-revisiondetails" class="flaggedrevs_short_detai ls" style="display:none">This is the latest accepted revision, reviewed on <i>10 September 2021</i>..</div>\n</div> \n</div>\n</div>\n\t\t<div id="contentSub2"></div>\n\t\t\n\t\t<div id="jump-to-nav"> </div>\n\t\tJump to navigation\n\t\t<a c lass="mw-jump-link" href="#searchInput">Jump to search\n\t\t<div id="mw-contenttext" class="mw-body-content mw-content-ltr" lang="en" dir="ltr"><div class="mw-pars er-output"><div class="shortdescription nomobile noexcerpt noprint searchaux" style ="display:none">Standard used to advise web crawlers and scrapers not to index a web page or site</div>\n\n\nThe robots exclusion stand ard, also known as the robots exclusion protocol or simply robots.txt , is a standard used by website to communicate with web crawlers an d other web robots. The standa rd specifies how to inform the web robot about which areas of the website should not be processed or scanned. Robots are often used by search engines to categorize websi tes. Not all robots cooperate with the standard; email harvesters, spambots, <a href="/w</pre>

iki/Malware" title="Malware">malware and robots that scan for security vulnerabi lities may even start with the portions of the website where they have been told to stay out. The standard can be used in conjunction with Sitemaps, a robot inclusion standard for websites.\n\n<div id ="toc" class="toc" role="navigation" aria-labelledby="mw-toc-heading"><input type="c heckbox" role="button" id="toctogglecheckbox" class="toctogglecheckbox" style="displ ay:none" /><div class="toctitle" lang="en" dir="ltr"><h2 id="mw-toc-heading">Content s</h2><label class="toctogglelabel" for="toctogglecheckb"</pre> ox"></label></div>\n\nclass="toclevel-1 tocsection-1">1 History\n class="toclevel-1 tocsection-2">2</s</pre> pan> Standard\n<li class="toclevel-1 tocsectio n-3">3 Secu rity\n<s pan class="tocnumber">4 Alternatives\n li class="toclevel-1 tocsection-5">5</sp an> Examples\nclass="toclevel-1 tocsection" -6">6 Nonstandard extensions\n\nclass="toclevel-2 tocsection" -7">6.1 Crawl-delay directive\n<li class="toclevel-2 tocsection-8">6.2 Allow directive\n6.3 Sitemap</ li>\nclass="toclevel-2 tocsection-10">6. 4 Host\n<li class="toclevel-2 tocsectio" n-11">6.5 <s pan class="toctext">Universal "*" match\n\n\n<li class="to clevel-1 tocsection-12">7/ span> Meta tags and headers\n<li class="toclev el-1 tocsection-13">8 See also\n9 References \n10 External links\n \n</div>\n\n<h2>History<span class="mw</pre> -editsection">[edit]</h2>\nThe stan dard was proposed by Martijn K oster,^{[1] }^{[2] } when working for Nexor^{[3]} in February 1994^{[4	 3;} on the <i>>www-talk</i> mailing list, the main communication channel for WWW-related activities at the time. Charles Stross claims to have provoked Koster to suggest robots.txt, after he wrote a badly-behaved web crawler that inadvertently caused a denial-of-service attack on Koster\'s server.^{&# 91;5]}\nIt quickly became a de facto standard that present and future web crawlers w ere expected to follow; most complied, including those operated by search engines su ch as WebCrawler, Lycos, and AltaV ista.^{[6]}\nOn July 1, 2019, Google announced the proposa l of the Robots Exclusion Protocol as an official standard under Internet Engine ering Task Force.^{& #91;7]} The draft^{[8]} will now undergo the acceptance process.\n\n<h2><s pan class="mw-headline" id="Standard">Standard<s pan class="mw-editsection-bracket">[<a href="/w/index.php?title=Robots_exclus"]</pre>

ion_standard&action=edit§ion=2" title="Edit section: Standard">edit< span class="mw-editsection-bracket">]</h2>\nWhen a site owner wishe s to give instructions to web robots they place a text file called <style data-mw-de duplicate="TemplateStyles:r886049734">.mw-parser-output .monospaced{font-family:mono space,monospace}</style>robots.txt in the root of th e web site hierarchy (e.g. <link rel="mw-deduplicated-inline-style" href="mw-data:Te mplateStyles:r886049734"/>https://www.example.com/robots.tx t). This text file contains the instructions in a specific format (see exampl es below). Robots that <i>choose</i> to follow the instructions try to fetch this fi le and read the instructions before fetching any other file from the website. If this file doesn\'t exist, web robots assume that the website owner does not wish to place any limitations on crawling the entire site.\nA robots.txt file on a website will function as a request that specifi ed robots ignore specified files or directories when crawling a site. This might be, for example, out of a preference for privacy from search engine results, or the beli ef that the content of the selected directories might be misleading or irrelevant to the categorization of the site as a whole, or out of a desire that an application on ly operates on certain data. Links to pages listed in robots.txt can still appear in search results if they are linked to from a page that is crawled.^{[9]}\n<P>A robots.t xt file covers one origin. For websites with multiple subdomains, each subdomain m ust have its own robots.txt file. If <link rel="mw-deduplicated-inline-style" href ="mw-data:TemplateStyles:r886049734"/>example.com ha d a robots.txt file but <link rel="mw-deduplicated-inline-style" href="mw-data:Templ ateStyles:r886049734"/>a.example.com did not, the ru les that would apply for <link rel="mw-deduplicated-inline-style" href="mw-data:Temp lateStyles:r886049734"/>example.com would not apply to <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r88604973 4"/>a.example.com. In addition, each protocol and po rt needs its own robots.txt file; <link rel="mw-deduplicated-inline-style" href="mwdata:TemplateStyles:r886049734"/>http://example.com/robots. txt does not apply to pages under <link rel="mw-deduplicated-inline-style" hr ef="mw-data:TemplateStyles:r886049734"/>http://example.com: 8080/ or <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyle s:r886049734"/>https://example.com/.\nSome ma jor search engines followi ng this standard include Ask,^{[10]} AOL,^{[11]} Baidu,^{[12]} DuckDuckGo,^{[13]} Google,<sup id="cite_ref-google-webmasters-spec_14-0" cl ass="reference">[14]</sup > Yahoo!,^{[15]} and Yandex.^{[16]} Bing^{<a hr</pre> ef="#cite_note-bing-blog-robots-17">[17]} is still^{[<i>when?</i>]} not fully compatible with the s tandard as it cannot inherit settings from the wildcard (<code>*</code>).^{[18]}\n The volunteering group Archive Team explicitly ignores robots.txt for the most part, viewing it as an obsolete standard that hinders web archival efforts. According to p roject leader Jason Scott, "unchecked, and left alone, the robots.txt file ensures n o mirroring or reference for items that may have general use and meaning beyond the website\'s context."^{<a href="#cite_note-1"</pre> $9">\&\#91;19\&\#93;}\nFor some years, the href="/wiki/Int$ e" title="Internet Archive">Internet Archive did not crawl sites with robots.tx t, but in April 2017, it announced that it would no longer honour directives in the robots.txt files.<sup id="cite_ref-20" class="reference">&#

91;20]</sup> "Over time we have observed that the robots.txt files that are geared toward search engine crawlers do not necessarily serve our archival purpose s".^{[21]} This was in response to entire domains bei ng tagged with robots.txt when the content became obsolete.^{[21&# 93;}\n\n<h2>Security<spa n class="mw-editsection">[edit]</h2> \nDespite the use of the terms "allow" and "disallow", the protocol is purely adv isory and relies on the compliance of the web robot.^{[22]} Malicious web robots are unlikely to honor r obots.txt; some may even use the robots.txt as a guide to find disallowed links and go straight to them. While this is sometimes claimed to be a security risk,^{[23]} this sort of <a href="/wiki/Security_through_obscurity" title="Security through obscurit</pre> y">security through obscurity is discouraged by standards bodies. The National Institute of Standards and Technology (NIST) in the United States specifically recommends against this practice: "System security sh ould not depend on the secrecy of the implementation or its components."^{[24]} In the co ntext of robots.txt files, security through obscurity is not recommended as a securi ty technique.^{[25]}\n\n<h2>Alternatives [edit]</sp an></h2>\nMany robots also pass a special user-agent to the web server when fetching con tent.^{[26]} A web administrator could also configure the server to automatically return failure (or pass alternative content) when it detects a connection using one of the robots.^{[27]}^{[28]}\nSome sites, such as Google, host a <code>humans.txt</code> file that displays information meant for humans to read.^{[29]} Some sites such as <a hr ef="/wiki/GitHub" title="GitHub">GitHub redirect humans.txt to an about page.<su p id="cite_ref-30" class="reference">[30]</sup> \nPreviously Google had a joke file hosted at <code>/killer-robots.txt</code> instructing th e Terminator not to kill the company founders Larry Page and Ser gey Brin.^{[31]}^{[$32\&\#93; } \\ n \\ nExamples$ [<a href="/</pre> w/index.php?title=Robots_exclusion_standard&action=edit§ion=5" title="Ed it section: Examples">edit]</ h2>\nThis example tells all robots that they can visit all files because the wild card <code>*</code> stands for all robots and the <code>Disallow</code> directive ha s no value, meaning no pages are disallowed.\n\nUser-agent: *\nAllow: /\n</ pre>\nThe same result can be accomplished with an empty or missing robots.txt fil agent: *\nDisallow: /\n\nThis example tells all robots not to enter three d irectories:\n\nUser-agent: *\nDisallow: /cgi-bin/\nDisallow: /tmp/\nDisallo w: /junk/\n\nThis example tells all robots to stay away from one specific f ile:\n\nUser-agent: *\nDisallow: /directory/file.html\n\nAll other files in the specified directory will be processed.\nThis example tells a spe cific robot to stay out of a website:\n\nUser-agent: BadBot # replace \'Bad Bot\' with the actual user-agent of the bot\nDisallow: $/\n\nThis example te$

lls two specific robots not to enter one specific directory:\n\n\npre>User-agent: BadBot # replace \'BadBot\' with the actual user-agent of the bot\nUser-agent: Googl ebot\nDisallow: /private/\n\nExample demonstrating how comments can be use d:\n\n# Comments appear after the "#" symbol at the start of a line, or aft er a directive\nUser-agent: * # match all bots\nDisallow: / # keep them out\n \nIt is also possible to list multiple robots with their own rules. The actual ro bot string is defined by the crawler. A few robot operators, such as Google, support several user-agent strings that allow the operator to deny access to a subset of their services by using specific user-agent s trings.^{[14]}\nExample demonstrat ing multiple user-agents:\n\nUser-agent: googlebot # all Google serv ices\nDisallow: /private/ # disallow this directory\n\nUser-agent: googlebo # only the news service\nDisallow: / # disallow everything # any robot\nDisallow: /something/ # disallo \n\nUser-agent: * w this directory\n\n<h2> Nonstandard extensions n-bracket">[edit]</h2>\n<h3>Crawl-delay directive[edit]</h3>\nThe crawl-d elay value is supported by some crawlers to throttle their visits to the host. Since this value is not part of the standard, its interpretation is dependent on the crawl er reading it. It is used when the multiple burst of visits from bots is slowing dow n the host. Yandex interprets the value as the number of seconds to wait between sub sequent visits.^{[16]} Bing defines crawl-delay as the size of a time window (from 1 to 30 seconds) during which BingBot will access a web site only once.^{[33]} Google provides an interface in it s search console for webmasters, to control the Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot">Googlebot lebot\'s subsequent visits.^{[34]}\n\nUser-agent: bingbot\nAllow : /\nC rawl-delay: 10\n\n<h3>Allow dir ective[</sp an>edit]</h3>\nSome major crawlers support an <code>Allow</code> direct ive, which can counteract a following <code>Disallow</code> directive.^{[35]}^{[36]} This is u seful when one tells robots to avoid an entire directory but still wants some HTML d ocuments in that directory crawled and indexed. While by standard implementation the first matching robots.txt pattern always wins, Google\'s implementation differs in t hat Allow patterns with equal or more characters in the directive path win over a ma tching Disallow pattern.^{[37]} Bing uses either the <code>Allow</code> or <code>Disallow </code> directive, whichever is more specific, based on length, like Google.^{\&\#91;17\&\#93;}\setminus n>In order to be compatible to all robots, if$ one wants to allow single files inside an otherwise disallowed directory, it is nec essary to place the Allow directive(s) first, followed by the Disallow, for exampl e:\n\nAllow: /directory1/myfile.html\nDisallow: /directory1/\n\nTh is example will Disallow anything in /directory1/ except /directory1/myfile.html, si nce the latter will match first. The order is only important to robots that follow t he standard; in the case of the Google or Bing bots, the order is not important.\n</ p>\n<h3>Sitemap[ed it]</h3>\nSome crawlers su pport a <code>Sitemap</code> directive, allowing multiple <a href="/wiki/Sitemaps" t

```
itle="Sitemaps">Sitemaps</a> in the same <samp>robots.txt</samp> in the form <code>S
itemap: <i>full-url</i></code>:<sup id="cite_ref-38" class="reference"><a href="#cit
e_note-38">[38]</a></sup>\n\nSitemap: http://www.example.com/sitema
p.xml\n<h3><span class="mw-headline" id="Host">Host</span><span class="mw-edit</pre>
section"><span class="mw-editsection-bracket">[</span><a href="/w/index.php?title=Ro
bots exclusion standard&action=edit&section=10" title="Edit section: Host">e
dit</a><span class="mw-editsection-bracket">]</span></h3>\nSome crawlers
 (<a href="/wiki/Yandex_Search" title="Yandex Search">Yandex</a>) support a <code>Ho
st</code> directive, allowing websites with multiple mirrors to specify their prefer
red domain:<sup id="cite_ref-39" class="reference"><a href="#cite_note-39">&#91;39&#
93;</a></sup>\n\nting.example.com\n\nThis is not supporte
d by all crawlers.\n\n<h3><span id="Universal_.22.2A.22_match"></span><span clas
s="mw-headline" id="Universal_"*"_match">Universal "*" match</span><span c
lass="mw-editsection"><span class="mw-editsection-bracket">[</span><a href="/w/inde"
x.php?title=Robots_exclusion_standard&action=edit&section=11" title="Edit se
ction: Universal " *" match">edit</a><span class="mw-editsection-bracket">]
</span></span></h3>\nThe <i>Robot Exclusion Standard</i> does not mention the "*"
character in the <code>Disallow:</code> statement.<sup id="cite_ref-40" class="refer
ence"><a href="#cite_note-40">&#91;40&#93;</a></sup>\n\n<h2><span class="mw-head
line" id="Meta_tags_and_headers">Meta tags and headers</span><span class="mw-editsec
tion"><span class="mw-editsection-bracket">[</span><a href="/w/index.php?title=Robot
s_exclusion_standard&action=edit&section=12" title="Edit section: Meta tags
and headers">edit</a><span class="mw-editsection-bracket">]</span></span></h2>\n
In addition to root-level robots.txt files, robots exclusion directives can be appli
ed at a more granular level through the use of <a href="/wiki/Robots_meta_tag" class
="mw-redirect" title="Robots meta tag">Robots meta tags</a> and X-Robots-Tag HTTP he
aders. The robots meta tag cannot be used for non-HTML files such as images, text fi
les, or PDF documents. On the other hand, the X-Robots-Tag can be added to non-HTML
files by using <a href="/wiki/.htaccess" title=".htaccess">.htaccess</a> and <a hre
f="/wiki/Httpd.conf" class="mw-redirect" title="Httpd.conf">httpd.conf</a> files.<su
p id="cite_ref-google-meta_41-0" class="reference"><a href="#cite_note-google-meta-4">
1">[41]</a></sup>\n\n<dl><dt>A "noindex" meta tag</dt></dl>\n<div class
="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr"><<span></s
pan class="p"><</span><span class="nt">meta</span> <span class="na">name</span><s</pre>
pan class="o">=</span><span class="s">&quot;robots&quot;</span> <span class="na">con
tent</span><span class="o">=</span><span class="s">&quot;noindex&quot;</span> <span</pre>
class="p">/></span>\n</div>\n<dl><dt>A "noindex" HTTP response header</dt>
</dl>\n<div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr"><pr
e><span></span>X-Robots-Tag: noindex\n</div>\nThe X-Robots-Tag is only effe
ctive after the page has been requested and the server responds, and the robots meta
tag is only effective after the page has loaded, whereas robots.txt is effective bef
ore the page is requested. Thus if a page is excluded by a robots.txt file, any robo
ts meta tags or X-Robots-Tag headers are effectively ignored because the robot will
not see them in the first place.<sup id="cite_ref-google-meta_41-1" class="referenc
e"><a href="#cite_note-google-meta-41">&#91;41&#93;</a></sup>\n\n<h2><span class
="mw-headline" id="See also">See also</span><span class="mw-editsection"><span class
="mw-editsection-bracket">[</span><a href="/w/index.php?title=Robots_exclusion_stand
ard&action=edit&section=13" title="Edit section: See also">edit</a><span cla
ss="mw-editsection-bracket">]</span></span></h2>\n<style data-mw-deduplicate="Templa
teStyles:r998391716">.mw-parser-output .div-col{margin-top:0.3em;column-width:30em}.
mw-parser-output .div-col-small{font-size:90%}.mw-parser-output .div-col-rules{colum
n-rule:1px solid #aaa}.mw-parser-output .div-col dl,.mw-parser-output .div-col ol,.m
w-parser-output .div-col ul{margin-top:0}.mw-parser-output .div-col li,.mw-parser-ou
tput .div-col dd{page-break-inside:avoid;break-inside:avoid-column}</style><div clas
s="div-col" style="column-width: 30em;">\n<code><a href="/wiki/Ads.txt" titl
e="Ads.txt">ads.txt</a></code>, a standard for listing authorized ad sellers
li><code><a href="/wiki/Security.txt" title="Security.txt">security.txt</a></code>,
a file to describe the process for security researchers to follow in order to repor
t security vulnerabilities\n<a href="/wiki/Automated_Content_Access_Protoco">href="/wiki/Automated_Content_Access_Protoco</a>
l" title="Automated Content Access Protocol">Automated Content Access Protocol</a>
\xe2\x80\x93 a failed proposal to extend robots.txt\n<a href="/wiki/BotSee"/>
r" title="BotSeer">BotSeer</a> \xe2\x80\x93 now inactive search engine for robots.tx
t files\n<a href="/wiki/Distributed_web_crawling" title="Distributed web cr
awling">Distributed web crawling</a>\n<a href="/wiki/Focused_crawler" title
="Focused crawler">Focused crawler</a>\n<a href="/wiki/Internet_Archive" ti
```

tle="Internet Archive">Internet Archive\n\a href="/wiki/National_Digita l_Library_Program" title="National Digital Library Program">National Digital Library Program (NDLP)\na href="/wiki/National_Digital_Information_Infrastruct" ure_and_Preservation_Program" title="National Digital Information Infrastructure and Preservation Program">National Digital Information Infrastructure and Preservation P rogram (NDIIPP)\nNofollow \nPerma.cc\n<a href ="/wiki/Robots_meta_tag" class="mw-redirect" title="Robots meta tag">Meta elements</ a> for search engines\nSitemaps</ a>\nSpider trap\n< li>Web archiving\n< a href="/wiki/Web_crawler" title="Web crawler">Web crawler\nnoindex\n</div>\n<h2>References[edit<span class</pre> ="mw-editsection-bracket">]</h2>\n<style data-mw-deduplicate="Template Styles:r1011085734">.mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;lis t-style-type:decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bo ttom:0;list-style-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30e m}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflistcolumns{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-pars er-output .reflist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mwparser-output .reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .re flist-upper-roman{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha {list-style-type:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type: lower-greek}.mw-parser-output .reflist-lower-roman{list-style-type:lower-roman}</sty</pre> le><div class="reflist reflist-columns references-column-width" style="column-width: 30em;">\n\n^ <style dat a-mw-deduplicate="TemplateStyles:r999302996">.mw-parser-output cite.citation{font-st yle:inherit}.mw-parser-output .citation q{quotes:"\\"""\\"""\'""}.mw-parser-outpu t .id-lock-free a,.mw-parser-output .citation .cs1-lock-free a{background:linear-gra dient(transparent,transparent),url("//upload.wikimedia.org/wikipedia/commons/6/65/Lo ck-green.svg")right 0.1em center/9px no-repeat}.mw-parser-output .id-lock-limited a,.mw-parser-output .id-lock-registration a,.mw-parser-output .citation .cs1-lock-1 imited a,.mw-parser-output .citation .cs1-lock-registration a{background:linear-grad ient(transparent,transparent),url("//upload.wikimedia.org/wikipedia/commons/d/d6/Loc k-gray-alt-2.svg")right 0.1em center/9px no-repeat}.mw-parser-output .id-lock-subscr iption a,.mw-parser-output .citation .cs1-lock-subscription a{background:linear-grad ient(transparent,transparent),url("//upload.wikimedia.org/wikipedia/commons/a/aa/Loc k-red-alt-2.svg")right 0.1em center/9px no-repeat}.mw-parser-output .cs1-subscriptio n,.mw-parser-output .cs1-registration{color:#555}.mw-parser-output .cs1-subscription span,.mw-parser-output .cs1-registration span{border-bottom:1px dotted;cursor:help}. mw-parser-output .cs1-ws-icon a{background:linear-gradient(transparent,transparent), url("//upload.wikimedia.org/wikipedia/commons/4/4c/Wikisource-logo.svg")right 0.1em center/12px no-repeat}.mw-parser-output code.cs1-code{color:inherit;background:inhe rit;border:none;padding:inherit}.mw-parser-output .cs1-hidden-error{display:none;fon t-size:100%}.mw-parser-output .cs1-visible-error{font-size:100%}.mw-parser-output .c s1-maint{display:none;color:#33aa33;margin-left:0.3em}.mw-parser-output .cs1-format {font-size:95%}.mw-parser-output .cs1-kern-left,.mw-parser-output .cs1-kern-wl-left {padding-left:0.2em}.mw-parser-output .cs1-kern-right,.mw-parser-output .cs1-kern-wl -right{padding-right:0.2em}.mw-parser-output .citation .mw-selflink{font-weight:inhe rit}</style><cite class="citation web cs1">"Historical". <i>Greenhills.c o.uk</i>. <a rel="nofollow" class="external text" href="https://web.archive.org/web/</pre> 20170403152037/http://www.greenhills.co.uk/historical.html">Archived from the or iginal on 2017-04-03. Retrieved 2017-03-03.</cite>\n\n^ <link re l="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id

="CITEREFFielding1994" class="citation web cs1">Fielding, Roy (1994). - rel="nof ow" class="external text" href="http://www94.web.cern.ch/WWW94/PapersWWW94/fielding. ps">"Maintaining Distributed Hypertext Infostructures: Welcome to MOMspider\'s Web" (PostScript). <i>First International Conference on the World Wide Web</i>. Geneva. Archived from the original on 2013-09-27. Retrieved September 25, 2013.</cite><spa</pre> n title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal& amp;rft.genre=unknown&rft.jtitle=First+International+Conference+on+the+World+Wid e+Web&rft.atitle=Maintaining+Distributed+Hypertext+Infostructures%3A+Welcome+to+ MOMspider%27s+Web&rft.date=1994&rft.aulast=Fielding&rft.aufirst=Roy& rft_id=http%3A%2F%2Fwww94.web.cern.ch%2FWWW94%2FPapersWWW94%2Ffielding.ps&rfr_id =info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z3988"></s pan>\n\n^ <link rel="mw-deduplicat ed-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"The Web Robots Pages". Robotstxt.org. 1994-06-30. Archived from the original on 2014-01-12. Retrieved 2013-12-29.</ cite>\n\nid="cite_note-4"> ^ \n<link rel</pre> ="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id ="CITEREFKoster1994" class="citation web cs1">Koster, Martijn (25 February 1994). "Important: Spiders, Robots and Web Wa nderers". <i>www-talk mailing list</i>. Archived from the original <sp</pre> an class="cs1-format">(Hypermail arc hived message) on October 29, 2013.</cite>\n li id="cite note-5">^ <link rel="mw-deduplicated-inline-style" href</pre> ="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"How I got here in the end, part five: "things can only get better!"". <i>Charlie\'s Diary </i>. 19 June 2006. Archived from the original on 2013-11-25. Retrieved 19 April 2014</sp an>.</cite>\n\nid="c ite_note-sear_Robo-6">^ <link rel="mw-deduplicated-inlin e-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFBarry Schwartz20 14" class="citation web cs1">Barry Schwartz (30 June 2014). "Robots.txt Celebrates 20 Years Of Blocking Search Eng ines". <i>Search Engine Land</i>. <a rel="nofollow" class="external text" href ="https://web.archive.org/web/20150907000430/http://searchengineland.com/robots-txt-

celebrates-20-years-blocking-search-engines-195479">Archived from the original o n 2015-09-07. Retrieved 2015 -11-19.</cite> \n\n\n\n\n 7">^ <link rel="mw-deduplicated-inline-s tyle" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1"><a re l="nofollow" class="external text" href="https://webmasters.googleblog.com/2019/07/r ep-id.html">"Formalizing the Robots Exclusion Protocol Specification". <i>Offici al Google Webmaster Central Blog</i>. Retrieved < span class="nowrap">2019-07-10.</cite>\n\n^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302 996"/><cite id="CITEREFM._Koster,_Stalworthy_Computing,_Ltd.,_G._Illyes,_H._Zeller,_ L._Harvey,_Google2019" class="citation web cs1">M. Koster, Stalworthy Computing, Lt d., G. Illyes, H. Zeller, L. Harvey, Google (2019-07-01). "Robots Exclu sion Protocol draft-rep-wg-topic-00". <i>IETF Documents</i>. Retrieved 2019-09-09.</cite>CS1 maint: multiple names: authors list (link)\n\n^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302 996"/><cite class="citation web cs1">"Uncrawled URLs in search result s". YouTube. Oct 5, 2009. Archived from the original on 2014-01-06. Retrieved 2013-12-29.</cite>\n\n^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateSty les:r999302996"/><cite class="citation web cs1">"About Ask.com: Webmaste rs". <i>About.ask.com</i>. Retrieved 16 February 2013.</cite>\n\n<li id ="cite_note-about-aol-search-11">^ <link rel="mw -deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="c itation web cs1"><a rel="nofollow" class="external text" href="http://search.aol.co</pre> m/aol/about">"About AOL Search". <i>Search.aol.com</i><span class="reference-acc

essdate">. Retrieved 16 February 2013.</cite><spa n title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal& amp;rft.genre=unknown&rft.jtitle=Search.aol.com&rft.atitle=About+AOL+Search& amp;rft_id=http%3A%2F%2Fsearch.aol.com%2Faol%2Fabout&rfr_id=info%3Asid%2Fen.wiki pedia.org%3ARobots+exclusion+standard" class="Z3988">\n\nid ="cite note-baidu-spider-12">^ <link rel="mw-dedupli cated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation" web cs1">"Baiduspider". <i>Baidu.com</i>. Retrieved 16 February 2013.</cite><span</pre> title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&a mp;rft.genre=unknown&rft.jtitle=Baidu.com&rft.atitle=Baiduspider&rft_id= http%3A%2F%2Fwww.baidu.com%2Fsearch%2Fspider_english.html&rfr_id=info%3Asid%2Fe n.wikipedia.org%3ARobots+exclusion+standard" class="Z3988">\n\n<1 i id="cite_note-duckduckgo-bot-13">^ <link rel="mw -deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="c itation web cs1">"DuckDuckGo Bot". <i>DuckDuckGo.com</i><span class="reference-acc</pre> essdate">. Retrieved 25 April 2017.</cite>\n\nid="cit e_note-google-webmasters-spec-14">^ ^{<i>a</i>} ^{<i></i>} <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r9993 02996"/><cite class="citation web cs1">"Web masters: Robots.txt Specifications". <i>Google Developers</i>. <a rel="nofollow"</pre> class="external text" href="https://web.archive.org/web/20130115214137/https://devel opers.google.com/webmasters/control-crawl-index/docs/robots_txt">Archived from t he original on 2013-01-15. Retrieved 16 February 2013.</cite> \\ (span) \\ (span)$ class="mw-cite-backlink">^ <link rel="mw-deduplicated-inline-style" href</pre> ="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Submitting your websit e to Yahoo! Search". Archived from the or iginal on 2013-01-21. Retrieved 16 February 2013.</cite>\n\nid="cite note-yandex-robots-16">^ ^{<i>a</i>} a> ^{<i>b</i>} <spa n class="reference-text"><link rel="mw-deduplicated-inline-style" href="mw-data:Temp lateStyles:r999302996"/><cite class="citation web cs1">"Using robots.txt"</a</pre> >. <i>Help.yandex.com</i>. Archi ved from the original on 2013-01-25. Retrieve d 16 February 2013.</cite><span title="ctx_ver=Z3")</pre>

9.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unkno wn&rft.jtitle=Help.yandex.com&rft.atitle=Using+robots.txt&rft_id=http%3 A%2F%2Fhelp.yandex.com%2Fwebmaster%2F%3Fid%3D1113851&rfr_id=info%3Asid%2Fen.wiki $\tt pedia.org\%3ARobots+exclusion+standard"~class="Z3988">\n$ ="cite_note-bing-blog-robots-17">^ ^{<i></i>} ^{<i>b</i>} <liin</pre> k rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots Exclusion Protocol: joining together to provide bet ter documentation". <i>Blogs.bing.com</i>. Archived from the original on 2014-08-18. Retrieved 16 February 2013.</cite>\n\n<span cl</pre> ass="mw-cite-backlink">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r99 9302996"/><cite class="citation web cs1">"Ho w to Create a Robots.txt File - Bing Webmaster Tools". <i>>www.bing.com</i>>. Retrieved 2019-02-06</sp an>.</cite>\n \n<a href="#cite_ref-1")</pre> 9">^ <link rel="mw-deduplicated-inline-s tyle" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFJason_Scott" class ="citation web cs1">Jason Scott. <a rel="nofollow" class="external text" href="http://www.archiveteam.org/index.php?t</pre> itle=Robots.txt">"Robots.txt is a suicide note". Archive Team. Archived from the original on 2017-02-18. Retrieved 18 February</s pan> 2017.</cite><span title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%"</pre> 2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Robots.txt+is+a+suicid e+note&rft.pub=Archive+Team&rft.au=Jason+Scott&rft_id=http%3A%2F%2Fwww.a rchiveteam.org%2Findex.php%3Ftitle%3DRobots.txt&rfr id=info%3Asid%2Fen.wikipedi a.org%3ARobots+exclusion+standard" class="Z3988">\n\nid="cite" _note-20">^ <link rel="mw-deduplicated-inline-style" href="mw-dat</pre> a:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots.txt meant for search engines do n\'t work well for web archives | Internet Archive Blogs". <i>blog.archive.org</ i>. 17 April 2017. Archived from the original on 2018 -12-04. Retrieved 2018-12-01 .</cite>\n\n^ <sup><i></

i></sup> ^{<i>b></i>} <link rel="mw-deduplicated-inline-style" href</pre> ="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFJones2017" class="citation ne ws cs1">Jones, Brad (24 April 2017). "The Internet Archive Will Ignore Robots.txt Files to Maintain Accuracy". <i> Digital Trends</i>. <a rel</pre> ="nofollow" class="external text" href="https://web.archive.org/web/20170516130029/h ttps://www.digitaltrends.com/computing/internet-archive-robots-txt/#ixzz4gQYOqpUi">A rchived from the original on 2017-05-16. Retr ieved 8 May 2017.</cite><span title="ctx_ver=Z39.</pre> 88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article &rft.jtitle=Digital+Trends&rft.atitle=The+Internet+Archive+Will+Ignore+Robot s.txt+Files+to+Maintain+Accuracy&rft.date=2017-04-24&rft.aulast=Jones&rf t.aufirst=Brad&rft_id=https%3A%2F%2Fwww.digitaltrends.com%2Fcomputing%2Finternet -archive-robots-txt%2F%23ixzz4gQYOqpUi&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARo bots+exclusion+standard" class="Z3988">\n\nid="cite_note-22"> ^ <span clas</pre> s="reference-text"><link rel="mw-deduplicated-inline-style" href="mw-data:TemplateSt yles:r999302996"/><cite class="citation web cs1">"Block URLs with ro bots.txt: Learn about robots.txt files". Archived from the original on 2015-08-14. Retrieved 2015-08-10.</cite> \n\n<a hr ef="#cite_ref-23">^ <link rel="mw-dedupl icated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation" web cs1">"Robots.txt tells hackers the places you don't want them to look". <i>The Register</i>. Archived from the original on 2015-08-21. Retrieved August 12, 2015.</cite>\n\nid="cite_note-24"><s pan class="mw-cite-backlink">^ <span class</pre> ="reference-text"><link rel="mw-deduplicated-inline-style" href="mw-data:TemplateSty les:r999302996"/><cite id="CITEREFScarfoneJansenTracy2008" class="citation journal c s1">Scarfone, K. A.; Jansen, W.; Tracy, M. (July 2008). "Guide to General Server Security" (PDF). Nat ="mw-redirect" title="Doi (identifier)">doi:10.6028/NIST.SP.800-123. Archived <sp an class="cs1-format">(PDF) from the original on 2011-10-08. Retrieved August 12, 2015.</ci te></sp an> Cite journal requires <cod e class="cs1-code">|journal=</code> (help)\n\nid="cite_note-2">help)$

5">^ <link rel="mw-deduplicated-inline-style" href="mw-data:Templat eStyles:r999302996"/><cite id="CITEREFSverre_H._Huseby2004" class="citation book cs 1">Sverre H. Huseby (2004). <i>Innocent Code: A Security Wake-Up Call for Web Programmers</i>. John Wiley & Dp. & 160;91\xe2\x80\x939 2. I SBN <bdi>9780470857472</bdi>. Archived from the original on 2016-04-01. Retrieved 2015-08-12. </cite>\n\n<1 i id="cite_note-26">^</ b> <link rel="mw-deduplicated-inline-style" href ="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"List of User-Agents (Sp iders, Robots, Browser)". User-agents.org. Archive d from the original on 2014-01-07. Retrieved 2013-12-29.</cite><span title="ctx_ver=Z39.88-20"</pre> 04&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rf t.btitle=List+of+User-Agents+%28Spiders%2C+Robots%2C+Browser%29&rft.pub=User-age nts.org&rft_id=http%3A%2F%2Fwww.user-agents.org%2F&rfr_id=info%3Asid%2Fen.wi kipedia.org%3ARobots+exclusion+standard" class="Z3988">\n ="cite_note-27">^</ span> <link rel="mw-deduplicated-inline-style" href="mw -data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Acc ess Control - Apache HTTP Server". Httpd.apache.org. Archived from the original on 2013-12-29. Retrieved 2013-12-29. </cite>\n\nid="cite_note-28">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><</pre> cite class="citation web cs1">"Deny Strings for Filtering Rules : The Official M icrosoft IIS Site". Iis.net. 2013-11-06. Archived from the original on 2014-01-01. Retrieved 2013-12-29.</cite>\n\nid="cite note-29">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r9993 02996"/><cite class="citation web cs1">"Google humans.txt". Retrieved October 3, 2019.</cite><sp an title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&am

```
p;rft.genre=unknown&rft.btitle=Google+humans.txt&rft_id=https%3A%2F%2Fwww.go
ogle.com%2Fhumans.txt&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+st
andard" class="Z3988"></span></span>\n\nid="cite_note-30"><span class="mw-c
ite-backlink"><b><a href="#cite_ref-30">^</a></b></span> <span class="reference-tex
t"><link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r99930299
6"/><cite class="citation web cs1"><a rel="nofollow" class="external text" href="htt
ps://github.com/humans.txt">"Github humans.txt"</a><span class="reference-accessdat
e">. Retrieved <span class="nowrap">October 3,</span> 2019</span>.</cite><span title
="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.ge
nre=unknown&rft.btitle=Github+humans.txt&rft_id=https%3A%2F%2Fgithub.com%2Fh
umans.txt&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" clas
s="Z3988"></span></span>\n\n<span class="mw-cite-backlin
k"><b><a href="#cite_ref-31">^</a></b></span> <span class="reference-text"><link rel
="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id
="CITEREFNewman2014" class="citation web cs1">Newman, Lily Hay (2014-07-03). <a rel
="nofollow" class="external text" href="https://slate.com/technology/2014/07/a-kille
r-robots-txt-google-easter-egg.html">"Is This a Google Easter Egg or Proof That Skyn
et Is Actually Plotting World Domination?"</a>. <i>Slate Magazine</i><span class="re
ference-accessdate">. Retrieved <span class="nowrap">2019-10-03</span></span>.</cite
><span title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajou</pre>
rnal&rft.genre=unknown&rft.jtitle=Slate+Magazine&rft.atitle=Is+This+a+Go
ogle+Easter+Egg+or+Proof+That+Skynet+Is+Actually+Plotting+World+Domination%3F&rf
t.date=2014-07-03&rft.aulast=Newman&rft.aufirst=Lily+Hay&rft_id=https%3
A%2F%2Fslate.com%2Ftechnology%2F2014%2F07%2Fa-killer-robots-txt-google-easter-egg.ht
ml&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z398
8"></span></span>\n\nid="cite_note-32"><span class="mw-cite-backlink"><b><a
href="#cite_ref-32">^</a></b></span> <span class="reference-text"><link rel="mw-dedu
plicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citati
on web cs1"><a rel="nofollow" class="external text" href="https://www.google.com/kil
ler-robots.txt">"/killer-robots.txt"</a>. 2018-01-10. <a rel="nofollow" class="exter
nal text" href="https://web.archive.org/web/20180110160916/https://www.google.com/ki
ller-robots.txt">Archived</a> from the original on 2018-01-10<span class="reference-
accessdate">. Retrieved <span class="nowrap">2018-05-25</span></span>.</cite><span t
itle="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rf
t.genre=unknown&rft.btitle=%2Fkiller-robots.txt&rft.date=2018-01-10&rft_
id=https%3A%2F%2Fwww.google.com%2Fkiller-robots.txt&rfr_id=info%3Asid%2Fen.wikip
edia.org%3ARobots+exclusion+standard" class="Z3988"></span></span>\n\nid="c
ite_note-bing-crawl-delay-33"><span class="mw-cite-backlink"><b><a href="#cite_ref-b</a>
ing-crawl-delay_33-0">^</a></b></span> <span class="reference-text"><link rel="mw-de
duplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="cita
tion web cs1"><a rel="nofollow" class="external text" href="https://blogs.bing.com/w
ebmaster/2012/05/03/to-crawl-or-not-to-crawl-that-is-bingbots-question/">"To crawl o
r not to crawl, that is BingBot\'s question"</a>. 3 May 2012. <a rel="nofollow" clas
s="external text" href="https://web.archive.org/web/20160203142822/https://blogs.bin
g.com/webmaster/2012/05/03/to-crawl-or-not-to-crawl-that-is-bingbots-question/">Arch
ived</a> from the original on 2016-02-03<span class="reference-accessdate">. Retriev
ed <span class="nowrap">9 February</span> 2016</span>.</cite><span title="ctx_ver=Z3"
9.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&
amp;rft.btitle=To+crawl+or+not+to+crawl%2C+that+is+BingBot%27s+question&rft.date
=2012-05-03&rft_id=https%3A%2F%2Fblogs.bing.com%2Fwebmaster%2F2012%2F05%2F03%2Ft
o-crawl-or-not-to-crawl-that-is-bingbots-question%2F&rfr_id=info%3Asid%2Fen.wiki
pedia.org%3ARobots+exclusion+standard" class="Z3988"></span></span>\n\n
="cite note-34"><span class="mw-cite-backlink"><b><a href="#cite ref-34">^</a></b></
span> <span class="reference-text"><link rel="mw-deduplicated-inline-style" href="mw
-data:TemplateStyles:r999302996"/><cite class="citation web cs1"><a rel="nofollow" c
lass="external text" href="https://support.google.com/webmasters/answer/48620?hl=e
n">"Change Googlebot crawl rate - Search Console Help"</a>. <i>support.google.com</i
>. <a rel="nofollow" class="external text" href="https://web.archive.org/web/2018111
8205747/https://support.google.com/webmasters/answer/48620?hl=en">Archived</a> from
the original on 2018-11-18<span class="reference-accessdate">. Retrieved <span clas
s="nowrap">22 October</span> 2018</span>.</cite><span title="ctx_ver=Z39.88-2004&am
p;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.j
title=support.google.com&rft.atitle=Change+Googlebot+crawl+rate+-+Search+Console
+Help&rft_id=https%3A%2F%2Fsupport.google.com%2Fwebmasters%2Fanswer%2F48620%3Fh
1%3Den&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class
```

```
="Z3988"></span></span>\n\nid="cite_note-35"><span class="mw-cite-backlin
k"><b><a href="#cite_ref-35">^</a></b></span> <span class="reference-text"><link rel
="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite clas
s="citation web cs1"><a rel="nofollow" class="external text" href="https://www.googl
e.com/support/webmasters/bin/answer.py?hl=en&answer=156449&from=40364">"Webm
aster Help Center - How do I block Googlebot?"</a>. <a rel="nofollow" class="externa
1 text" href="https://web.archive.org/web/20100801082008/http://www.google.com/suppo
rt/webmasters/bin/answer.py?hl=en&answer=156449&from=40364">Archived</a> fro
m the original on 2010-08-01<span class="reference-accessdate">. Retrieved <span cla
ss="nowrap">2007-11-20</span></span>.</cite><span title="ctx_ver=Z39.88-2004&amp;rft
_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=We
bmaster+Help+Center+-+How+do+I+block+Googlebot%3F&rft_id=https%3A%2F%2Fwww.googl
e.com%2Fsupport%2Fwebmasters%2Fbin%2Fanswer.py%3Fhl%3Den%26answer%3D156449%26from%3D
40364&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z
3988"></span></span>\n\ndi id="cite_note-36"><span class="mw-cite-backlink"><b
><a href="#cite_ref-36">^</a></b></span> <span class="reference-text"><link rel="mw-</pre>
deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="ci
tation web cs1"><a rel="nofollow" class="external text" href="http://help.yahoo.com/
1/us/yahoo/search/webcrawler/slurp-02.html">"How do I prevent my site or certain sub
directories from being crawled? - Yahoo Search Help"</a>. <a rel="nofollow" class="e
xternal text" href="https://web.archive.org/web/20071013014215/http://help.yahoo.co
m/l/us/yahoo/search/webcrawler/slurp-02.html">Archived</a> from the original on 2007
-10-13<span class="reference-accessdate">. Retrieved <span class="nowrap">2007-11-20
</span></span>.</cite><span title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%2F
fmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=How+do+I+prevent+my+site
+or+certain+subdirectories+from+being+crawled%3F+-+Yahoo+Search+Help&rft_id=htt
p%3A%2F%2Fhelp.yahoo.com%2F1%2Fus%2Fyahoo%2Fsearch%2Fwebcrawler%2Fslurp-02.html&
rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z3988"></sp
an></span>\n\n<span class="mw-cite-backlink"><b><a href
="#cite_ref-37">^</a></b></span> <span class="reference-text"><link rel="mw-deduplic
ated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation w
eb cs1"><a rel="nofollow" class="external text" href="http://blog.semetrical.com/goo
gles-secret-approach-to-robots-txt/">"Google\'s Hidden Interpretation of Robots.txt"
</a>. <a rel="nofollow" class="external text" href="https://web.archive.org/web/2010
1120023619/http://blog.semetrical.com/googles-secret-approach-to-robots-txt/">Archiv
ed</a> from the original on 2010-11-20<span class="reference-accessdate">. Retrieved
<span class="nowrap">2010-11-15</span></span>.</cite><span title="ctx_ver=Z39.88-200">
4&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.
btitle=Google%27s+Hidden+Interpretation+of+Robots.txt&rft_id=http%3A%2F%2Fblog.s
emetrical.com%2Fgoogles-secret-approach-to-robots-txt%2F&rfr_id=info%3Asid%2Fen.
wikipedia.org%3ARobots+exclusion+standard" class="Z3988"></span></span>\n
id="cite_note-38"><span class="mw-cite-backlink"><b><a href="#cite_ref-38">^</a></b
></span> <span class="reference-text"><link rel="mw-deduplicated-inline-style" href
="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1"><a rel="nofollo
w" class="external text" href="https://web.archive.org/web/20090305061841/http://yse
archblog.com/2007/04/11/webmasters-can-now-auto-discover-with-sitemaps/">"Yahoo! Sea
rch Blog - Webmasters can now auto-discover with Sitemaps"</a>. Archived from <a rel
="nofollow" class="external text" href="http://ysearchblog.com/2007/04/11/webmasters
-can-now-auto-discover-with-sitemaps/">the original</a> on 2009-03-05<span class="re
ference-accessdate">. Retrieved <span class="nowrap">2009-03-23</span></span>.</cite
><span title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Aboo</pre>
k&rft.genre=unknown&rft.btitle=Yahoo%21+Search+Blog+-+Webmasters+can+now+aut
o-discover+with+Sitemaps&rft_id=http%3A%2F%2Fysearchblog.com%2F2007%2F04%2F11%2F
webmasters-can-now-auto-discover-with-sitemaps%2F&rfr_id=info%3Asid%2Fen.wikiped
ia.org%3ARobots+exclusion+standard" class="Z3988"></span></span>\n\nid="cit
e note-39"><span class="mw-cite-backlink"><b><a href="#cite ref-39">^</a></b></span>
<span class="reference-text"><link rel="mw-deduplicated-inline-style" href="mw-data:</pre>
TemplateStyles:r999302996"/><cite class="citation web cs1"><a rel="nofollow" class
="external text" href="http://help.yandex.com/webmaster/?id=1113851">"Yandex - Using
robots.txt"</a>. <a rel="nofollow" class="external text" href="https://web.archive.o
rg/web/20130509230548/http://help.yandex.com/webmaster/?id=1113851">Archived</a> fro
m the original on 2013-05-09<span class="reference-accessdate">. Retrieved <span cla
ss="nowrap">2013-05-13</span></span>.</cite><span title="ctx_ver=Z39.88-2004&amp;rft
_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Ya
ndex+-+Using+robots.txt&rft_id=http%3A%2F%2Fhelp.yandex.com%2Fwebmaster%2F%3Fid%
```

3D1113851&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" clas s="Z3988">\n\nid="cite_note-40">^ <link rel ="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite clas s="citation web cs1">"Robots.txt Specifications". <i> Google Developers</i>. Retrieved February 15, 2020.</cite>\n \n^ <a href="#</pre> cite_ref-google-meta_41-0">^{<i></i>} ^{<i><i>b></i>} <link</pre> rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots_meta_ tag and X-Robots-Tag HTTP header specifications - Webmasters \xe2\x80\x94 Google De velopers". Archived from the original on 2013-08-08. Retrieved 2013-08-17.</cite><spa n title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&am p;rft.genre=unknown&rft.btitle=Robots+meta+tag+and+X-Robots-Tag+HTTP+header+spec ifications+-+Webmasters+%E2%80%94+Google+Developers&rft_id=https%3A%2F%2Fdevelop ers.google.com%2Fwebmasters%2Fcontrol-crawl-index%2Fdocs%2Frobots_meta_tag&rfr_i d=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z3988"></ span>\n\n</div>\n<h2>Externa l links[</s</pre> pan><a href="/w/index.php?title=Robots_exclusion_standard&action=edit&sectio</pre> n=15" title="Edit section: External links">edit]</h2>\n< a rel="nofollow" class="external text" href="http://www.robotstxt.org">Official webs ite\n\a rel="nofollow" class="external text" href="htt p://www.robotstxt.org/db.html">Robots Database (list of bot names) rel="nofollow" class="external text" href="https://sitechecker.pro/robots-tester/"> Robots.txt checker tool\na rel="nofollow" class="external text" href ="https://www.bing.com/webmaster/help/how-to-create-a-robots-txt-file-cb7c31ec">Micr osoft bingbot official documentation\n<div class="noprint metadata nav box" role="navigation" aria-label="Portals" style="font-weight:bold;padding:0.4em 2e m"><img alt="" src="//upload.wikimedia.o</pre> rg/wikipedia/commons/thumb/f/f9/Crystal_Clear_app_linneighborhood.svg/21px-Crystal_C lear app linneighborhood.svg.png" decoding="async" width="21" height="21" srcset="// upload.wikimedia.org/wikipedia/commons/thumb/f/f9/Crystal_Clear_app_linneighborhood. svg/32px-Crystal_Clear_app_linneighborhood.svg.png 1.5x, //upload.wikimedia.org/wiki pedia/commons/thumb/f/f9/Crystal_Clear_app_linneighborhood.svg/42px-Crystal_Clear_ap p_linneighborhood.svg.png 2x" data-file-width="407" data-file-height="407" /></s pan>Internet portal</spa n></div>\n<div role="navigation" class="navbox authority-control" aria-lab elledby="Authority_control_frameless_|text-top_|10px_|alt =Edit_this_at_Wikidata_|link=https://www.wikidata.org/wiki/Q80776#i dentifiers|Edit_this_at_Wikidata" style="padding:3px"> inks hlist navbox-inner" style="border-spacing:0;background:transparent;color:inheri t"><th id="Authority_control_frameless_&#124;text-top_&#124;10px_ |alt=Edit_this_at_Wikidata_|link=https://www.wikidata.org/ wiki/Q80776#identifiers|Edit_this_at_Wikidata" scope="row" class="navbox-gr oup" style="width:1%">Authority control <img alt="Edit this at Wikidata" src="//upload.wi kimedia.org/wikipedia/en/thumb/8/8a/00js_UI_icon_edit-ltr-progressive.svg/10px-00js_ UI_icon_edit-ltr-progressive.svg.png" decoding="async" width="10" height="10" style ="vertical-align: text-top" srcset="//upload.wikimedia.org/wikipedia/en/thumb/8/8a/0

```
Ojs_UI_icon_edit-ltr-progressive.svg/15px-00js_UI_icon_edit-ltr-progressive.svg.png
1.5x, //upload.wikimedia.org/wikipedia/en/thumb/8/8a/00js_UI_icon_edit-ltr-progress
ive.svg/20px-00js_UI_icon_edit-ltr-progressive.svg.png 2x" data-file-width="20" data
-file-height="20" /></a><td class="navbox-list navbox-odd" style="text-align:le
ft;border-left-width:2px;border-left-style:solid;width:100%;padding:0px"><div style
="padding:0em 0.25em">\n<span class="uid"><a rel="nofollow" class="external"
text" href="https://academic.microsoft.com/v2/detail/163978704">Microsoft Academic
</a></span>\n</div></div>\n<!-- \nNewPP limit re
port\nParsed by mw1395\nCached time: 20210919053121\nCache expiry: 1814400\nReduced
expiry: false\nComplications: [vary\xe2\x80\x90revision\xe2\x80\x90sha1]\nCPU time
usage: 0.604 seconds\nReal time usage: 0.727 seconds\nPreprocessor visited node cou
nt: 2467/1000000\nPost\xe2\x80\x90expand include size: 78042/2097152 bytes\nTemplate
argument size: 1921/2097152 bytes\nHighest expansion depth: 12/40\nExpensive parser
function count: 7/500\nUnstrip recursion depth: 1/20\nUnstrip post\xe2\x80\x90expan
d size: 135669/5000000 bytes\nLua time usage: 0.351/10.000 seconds\nLua memory usag
e: 6446910/52428800 bytes\nNumber of Wikibase entities loaded: 1/400\n-->\n<!--\nTra
nsclusion expansion time report (%,ms,calls,template)\n100.00% 661.793
1\n 45.61% 301.824
                        1 Template:Reflist\n 33.71% 223.116
                                                                 38 Template:Cite w
eb\n 12.94%
              85.627
                         1 Template:Short_description\n 10.33%
                                                                68.389
                                                                            1 Temp
late:Pp-pc1\n 7.63%
                      50.497
                                  1 Template:Pagetype\n 7.26%
                                                                 48.071
                                                                            1 Temp
                    41.950
late:When\n 6.34%
                                1 Template:Fix\n 5.07%
                                                          33.554
                                                                      1 Template:Au
thority_control\n 3.66%
                                      2 Template:Category_handler\n-->\n\n<!-- Save</pre>
                         24.227
d in parser cache with key enwiki:stable-pcache:idhash:101673-0!canonical and timest
amp 20210919053121 and revision id 1043553668. Serialized with JSON.\n -->\n</div><n
oscript><img src="//en.wikipedia.org/wiki/Special:CentralAutoLogin/start?type=1x1" a
lt="" title="" width="1" height="1" style="border: none; position: absolute;" /></no</pre>
script>\n<div class="printfooter">Retrieved from "<a dir="ltr" href="https://en.wiki</pre>
pedia.org/w/index.php?title=Robots_exclusion_standard&oldid=1043553668">https://
en.wikipedia.org/w/index.php?title=Robots_exclusion_standard&oldid=1043553668</a
>"</div></div>\n\t\<div id="catlinks" class="catlinks" data-mw="interface"><div id
="mw-normal-catlinks" class="mw-normal-catlinks"><a href="/wiki/Help:Category" title
="Help:Category">Categories</a>: <a href="/wiki/Category:World_Wide_Web" tit
le="Category:World Wide Web">World Wide Web</a><a href="/wiki/Category:Inte"
rnet_bots" title="Category:Internet bots">Internet bots</a></div><div id</pre>
="mw-hidden-catlinks" class="mw-hidden-catlinks mw-hidden-cats-hidden">Hidden catego
ries: <a href="/wiki/Category:CS1_maint:_multiple_names:_authors_list" title</a>
="Category:CS1 maint: multiple names: authors list">CS1 maint: multiple names: autho
rs list</a>a href="/wiki/Category:CS1_errors:_missing_periodical" title="C
ategory:CS1 errors: missing periodical">CS1 errors: missing periodical</a>
a href="/wiki/Category:Articles_with_short_description" title="Category:Articles wit
h short description">Articles with short description</a><a href="/wiki/Cate"
gory:Short_description_matches_Wikidata" title="Category:Short description matches W
ikidata">Short description matches Wikidata</a><a href="/wiki/Category:Wiki
pedia_pending_changes_protected_pages" title="Category:Wikipedia pending changes pro
tected pages">Wikipedia pending changes protected pages</a><a href="/wiki/C"></a>
ategory:All articles with vague or ambiguous time" title="Category:All articles with
vague or ambiguous time">All articles with vague or ambiguous time</a><a hr
ef="/wiki/Category:Vague_or_ambiguous_time_from_December_2020" title="Category:Vague
or ambiguous time from December 2020">Vague or ambiguous time from December 2020</a>
<a href="/wiki/Category:Official_website_different_in_Wikidata_and_Wikipedi"
a" title="Category:Official website different in Wikidata and Wikipedia">Official we
bsite different in Wikidata and Wikipedia</a><a href="/wiki/Category:Articl">href="/wiki/Category:Articl">href="/wiki/Category:Articl">href="/wiki/Category:Articl">href="/wiki/Category:Articl</a>
es with MA identifiers" title="Category:Articles with MA identifiers">Articles with
MA identifiers</a></div></div>\n\t</div>\n</div\\n<div id=\'mw-data-after
-content\'>\n\t<div class="read-more-container"></div>\n</div>\n\n<div id="mw-naviga
tion">\n\t<h2>Navigation menu</h2>\n\t<div id="mw-head">\n\t\t<nav id="p-personal" c
lass="mw-portlet mw-portlet-personal vector-user-menu-legacy vector-menu" aria-label
ledby="p-personal-label" role="navigation" \n\t >\n\t<h3 id="p-personal-label" class
="vector-menu-heading"> <span>Personal tools</span>\n\t</h3>\n\t<div class="vector-m
enu-content">\n\t\t\n\t\<li id="pt-anonuserpag
e" class="mw-list-item">Not logged inid="pt-anontalk" class="mw-list-item">
<a href="/wiki/Special:MyTalk" title="Discussion about edits from this IP address
 [n]" accesskey="n">Talk</a><a hr
ef="/wiki/Special:MyContributions" title="A list of edits made from this IP address
[y]" accesskey="y">Contributions</a>id="pt-createaccount" class="mw-list-i
```

tem">Create accountid="pt-login" class="mw-list-item">Log in\n\t\t\n\t</div>\n</nav>\n\n\t\t<div id="left-navigation"> \n\t\t<nav id="p-namespaces" class="mw-portlet mw-portlet-namespaces vector-menu v</pre> ector-menu-tabs" aria-labelledby="p-namespaces-label" role="navigation" \n\t >\n\t<h 3 id="p-namespaces-label" class="vector-menu-heading"> Namespaces\n\t h3>\n\t<div class="vector-menu-content">\n\t\t\n\t\t<ul class="vector-menu-content-l ist">Article<li id="ca-talk" class="mw-list-item">Ta $lk\n\t</div>\n</nav>\n\t\t\rank id="p-variants" class="mw-por"$ tlet mw-portlet-variants emptyPortlet vector-menu-dropdown-noicon vector-menu vector -menu-dropdown" aria-labelledby="p-variants-label" role="navigation" \n\t >\n\t<inpu t type="checkbox"\n\t\tid="p-variants-checkbox"\n\t\trole="button"\n\t\taria-haspopu p="true"\n\t\tdata-event-name="ui.dropdown-p-variants"\n\t\tclass=" vector-menu-chec kbox" aria-labelledby="p-variants-label" />\n\t<h3 id="p-variants-label" class="vect or-menu-heading"> Variants\n\t\t\expanded\n\t\t\tcollapsed \n\t</h3>\n\t<div class="vector-menu-content">\n\t\t\n\t\t<ul class="vector-m enu-content-list">\n\t\\t\n\t</div>\n<\n\t\\t<div id="right-na" vigation">\n\t\t<nav id="p-views" class="mw-portlet mw-portlet-views vector-menu v ector-menu-tabs" aria-labelledby="p-views-label" role="navigation" \n\t >\n\t<h3 id ="p-views-label" class="vector-menu-heading"> Views\n\t</h3>\n\t<div cl ass="vector-menu-content">\n\t\t\n\t\t<li id="c a-view" class="selected mw-list-item">Read id="ca-edit" class="mw-list-item">Edit id="ca-history" class="mw-list-item">View history\n\t\t\n\t</div>\n</nav>\n\n\t\t\cap id="p-caction"$ s" class="mw-portlet mw-portlet-cactions emptyPortlet vector-menu-dropdown-noicon ve ctor-menu vector-menu-dropdown" aria-labelledby="p-cactions-label" role="navigation" title="More options"\n\t >\n\t<input type="checkbox"\n\t\tid="p-cactions-checkbox"\n</pre> \t\trole="button"\n\t\taria-haspopup="true"\n\t\tdata-event-name="ui.dropdown-p-cact ions"\n\t\tclass=" vector-menu-checkbox" aria-labelledby="p-cactions-label" />\n\t<h</pre> 3 id="p-cactions-label" class="vector-menu-heading"> More\n\t\t\expanded\n\t\t\collapsed/\square\n\t</h3>\n\t<div class="vector-menu-conten"</pre> t">\n\t\t\n\t\</div>\n</nav>\n \n\t\t<div id="p-search" role="search" class="vector-search-box">\n\t<div>\n\t\t<h</pre> 3 >\n\t\t<label for="searchInput">Search</label>\n\t\t</h3>\n\t\t<form action="/w/</pre> index.php" id="searchform">\n\t\t\cdiv id="simpleSearch" data-search-loc="header-na vigation">\n\t\t\t<input type="search" name="search" placeholder="Search Wikipedi</pre> a" autocapitalize="sentences" title="Search Wikipedia [f]" accesskey="f" id="searchI nput"/>\n\t\t\t<input type="hidden" name="title" value="Special:Search"/>\n\t\t\t \t<input type="submit" name="fulltext" value="Search" title="Search Wikipedia for th</pre> is text" id="mw-searchButton" class="searchButton mw-fallbackSearchButton"/>\n\t\t\t \t<input type="submit" name="go" value="Go" title="Go to a page with this exact name if it exists" id="searchButton" class="searchButton"/>\n\t\t</div>\n\t\t</form>\n \t</div>\n</div>\n\t\div>\n\t\div>\n\t\cdiv id="mw-panel">\n\t<div id="p-log o" role="banner">\n\t\t\n\t</div>\n\t<nav id="p-navigation" class="mw-portlet mw -portlet-navigation vector-menu vector-menu-portal portal" aria-labelledby="p-naviga tion-label" role="navigation" \n\t >\n\t<h3 id="p-navigation-label" class="vector-me nu-heading"> Navigation\n\t</h3>\n\t<div class="vector-menu-content">\n \t\t\n\t\cul class="vector-menu-content-list"><li id="n-mainpage-description" class ="mw-list-item">Main pageid="n-contents" class="mw-list-item">Contents<li id="n-curre ntevents" class="mw-list-item">Current eventsid="n-randompage" class="mw-li

```
st-item"><a href="/wiki/Special:Random" title="Visit a randomly selected article
 [x]" accesskey="x">Random article</a>id="n-aboutsite" class="mw-list-ite
m"><a href="/wiki/Wikipedia:About" title="Learn about Wikipedia and how it works">Ab
out Wikipedia</a>id="n-contactpage" class="mw-list-item"><a href="//en.wiki"
pedia.org/wiki/Wikipedia:Contact_us" title="How to contact Wikipedia">Contact us</a>
id="n-sitesupport" class="mw-list-item"><a href="https://donate.wikimedia.o"><a href="https://donate.wikimedia.o">><a href
rg/wiki/Special:FundraiserRedirector?utm_source=donate&utm_medium=sidebar&ut
m_campaign=C13_en.wikipedia.org&uselang=en" title="Support us by donating to the
Wikimedia Foundation">Donate</a>\n\t\t\n\t</div>\n</nav>\n\n\t<nav id="p-i
nteraction" class="mw-portlet mw-portlet-interaction vector-menu vector-menu-portal
 portal" aria-labelledby="p-interaction-label" role="navigation" \n\t <\n\t<h3 id="p</pre>
-interaction-label" class="vector-menu-heading"> <span>Contribute</span>\n\
\t<div class="vector-menu-content">\n\t\t\n\t\t
<a href="/wiki/Help:Contents" title="Guidance o</pre>
n how to use and edit Wikipedia">Help</a><li id="n-introduction" class="mw-list
-item"><a href="/wiki/Help:Introduction" title="Learn how to edit Wikipedia">Learn t
o edit</a><a href="/wiki/Wikipedia:Commu
nity_portal" title="The hub for editors">Community portal</a>id="n-recentch
anges" class="mw-list-item"><a href="/wiki/Special:RecentChanges" title="A list of r
ecent changes to Wikipedia [r]" accesskey="r">Recent changes</a>id="n-uploa"
d" class="mw-list-item"><a href="/wiki/Wikipedia:File_Upload_Wizard" title="Add imag
es or other media for use on Wikipedia">Upload file</a>\n\t\\n\t</div>\n
\verb|\| <| \text{nav} | \text{ncnav id="p-tb" class="mw-portlet mw-portlet-tb vector-menu-portlet-tb vector-menu-portlet-tb
1 portal" aria-labelledby="p-tb-label" role="navigation" \n\t <\n\t<h3 id="p-tb-labe"</pre>
1" class="vector-menu-heading"> <span>Tools</span>\n\t</h3>\n\t<div class="vector-me</pre>
nu-content">\n\t\t\n\t\t<li id="t-whatlinksher
e" class="mw-list-item"><a href="/wiki/Special:WhatLinksHere/Robots_exclusion_standa
rd" title="List of all English Wikipedia pages containing links to this page [j]" ac
cesskey="j">What links here</a><li id="t-recentchangeslinked" class="mw-list-it
em"><a href="/wiki/Special:RecentChangesLinked/Robots_exclusion_standard" rel="nofol
low" title="Recent changes in pages linked from this page [k]" accesskey="k">Related
changes</a>id="t-upload" class="mw-list-item"><a href="/wiki/Wikipedia:File
_Upload_Wizard" title="Upload files [u]" accesskey="u">Upload file</a><li id="t
-specialpages" class="mw-list-item"><a href="/wiki/Special:SpecialPages" title="A li
st of all special pages [q]" accesskey="q">Special pages</a>id="t-permalin"
k" class="mw-list-item"><a href="/w/index.php?title=Robots_exclusion_standard&amp;ol
did=1043553668" title="Permanent link to this revision of this page">Permanent link
</a>id="t-info" class="mw-list-item"><a href="/w/index.php?title=Robots_exc"><a href="/w/index.php?title=Robots_exc">/w/index.php?title=Robots_exc</a><a href="/w/index.php?title=Robots_exc">/w/inde
lusion standard&action=info" title="More information about this page">Page infor
mation</a>id="t-cite" class="mw-list-item"><a href="/w/index.php?title=Spec</pre>
ial:CiteThisPage&page=Robots_exclusion_standard&id=1043553668&wpFormIden
tifier=titleform" title="Information on how to cite this page">Cite this page</a></l
i><a href="https://www.wikidata.org/wiki/Sp
ecial:EntityPage/Q80776" title="Structured data on this page hosted by Wikidata [g]"
accesskey="g">Wikidata item</a>\n\t\t\n\t</div>\n</nav>\n<nav id="p-coll-p
rint export" class="mw-portlet mw-portlet-coll-print export vector-menu vector-menu-
portal portal" aria-labelledby="p-coll-print_export-label" role="navigation" \n\t >
\n\t<h3 id="p-coll-print_export-label" class="vector-menu-heading"> <span>Print/expo
rt</span>\n\t</h3>\n\t<div class="vector-menu-content">\n\t\t\n\t\t<ul class="vector
-menu-content-list"><a href="/w/in
dex.php?title=Special:DownloadAsPdf&page=Robots_exclusion_standard&action=sh
ow-download-screen" title="Download this page as a PDF file">Download as PDF</a></li
>id="t-print" class="mw-list-item"><a href="/w/index.php?title=Robots exclusion"</pre>
standard&printable=yes" title="Printable version of this page [p]" accesskey
="p">Printable version</a>\n\t\t\n\t</div>\n</nav>\n\n\t<nav id="p-lang" c
lass="mw-portlet mw-portlet-lang vector-menu vector-menu-portal portal" aria-labelle
dby="p-lang-label" role="navigation" \n\t >\n\t<h3 id="p-lang-label" class="vector-m
enu-heading"> <span>Languages</span>\n\t</h3>\n\t<div class="vector-menu-content">\n
\t\t\n\t\cul class="vector-menu-content-list">
ki-ar mw-list-item"><a href="https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D9%8A%D8%A7%
D8%B1_%D8%A7%D8%B3%D8%AA%D8%A8%D8%B9%D8%A7%D8%AF_%D8%A7%D9%84%D8%B1%D9%88%D8%A8%D9%8
3\xd8\xaa\xd8\xa8\xd8\xb9\xd8\xa7\xd8\xaf \xd8\xa7\xd9\x84\xd8\xb1\xd9\x88\xd8\xa8\x
d9\x88\xd8\xaa\xd8\xaa \xe2\x80\x93 Arabic" lang="ar" hreflang="ar" class="i
nterlanguage-link-target">\xd8\xa7\xd9\x84\xd8\xb1\xd8\xa8\xd9\x8a\xd8\xa9</
```

a>Boar ischclass="interlanguage-link interwiki-ca mw-list-item">Catal\xc3\xa0<li class="interlanguage-link interwi ki-cs mw-list-item">\xc4\x8ce\xc5\xa1tina ="interlanguage-link interwiki-da mw-list-item">Dansk<li class="int erlanguage-link interwiki-de mw-list-item">Deutsch<li class="int erlanguage-link interwiki-es mw-list-item">Espa\xc3\xb1ol<li class="interlanguage-link interwiki-fa mw-list-ite m">\xd9\x81\xd8\xa7\xd8\xb1\xd8\xb3\xdb\x8c<li class="int erlanguage-link interwiki-fr mw-list-item">Fran\xc3 \xa7ais\xed \x95\x9c\xea\xb5\xad\xec\x96\xb4<li class="interlanguage-link interwiki-id mw-list-item"> Bahasa IndonesiaItaliano<li class="interlanguage-link interwiki-he mw-lis t-item">\xd7\xa2\xd7\x91\xd7\xa8\xd7\x99\xd7\xaa class="interlanguage-link interwiki-nl mw-list-item">Nederlands<li class="interlanguage-link interwiki-ja mw-list-item">\xe6\x97\xa5\xe6\x9 c\xac\xe8\xaa\x9eclass="interlanguage-link interwiki-pl mw-list-item">< a href="https://pl.wikipedia.org/wiki/Robots Exclusion Protocol" title="Robots Exclu sion Protocol \xe2\x80\x93 Polish" lang="pl" hreflang="pl" class="interlanguage-link -target">PolskiPortugu\xc3\xaas<li class="in terlanguage-link interwiki-ru mw-list-item"><a href="https://ru.wikipedia.org/wiki/% D0%A1%D1%82%D0%B0%D0%BD%D0%B4%D0%B0%D1%80%D1%82_%D0%B8%D1%81%D0%BA%D0%BB%D1%8E%D1%8 7%D0%B5%D0%BD%D0%B8%D0%B9_%D0%B4%D0%BB%D1%8F_%D1%80%D0%BE%D0%B1%D0%BE%D1%82%D0%BE%D 0%B2" title="\xd0\xa1\xd1\x82\xd0\xb0\xd0\xbd\xd0\xb4\xd0\xb4\xd1\x82 \xd0\x

b8\xd1\x81\xd0\xba\xd0\xbb\xd1\x8e\xd1\x87\xd0\xb5\xd0\xbd\xd0\xb8\xd0\xb9 \xd0\xb4 $\xd0\xbb\xd1\x8f \xd0\xbe\xd0\xbe\xd1\x82\xd0\xbe\xd$ 93 Russian" lang="ru" hreflang="ru" class="interlanguage-link-target">\xd0\xa0\xd1\x $83\xd1\x81\xd1\xd0\xb8\xd0\xb9\/a>\/li>\clic class="interlanguage-link interlanguage-link interlanguage-li$ erwiki-simple mw-list-item">Simple English</ li>\xd0\xa1\xd1\x80\xd0\xbf\xd1\x81\xd0\xba\xd0\xb8 / srpski<li class="int for example of the content of the conten erlanguage-link interwiki-fi mw-list-item">Suomi<li class="interlan guage-link interwiki-sv mw-list-item"><a href="https://sv.wikipedia.org/wiki/Robots_</pre> Exclusion_Standard" title="Robots Exclusion Standard \xe2\x80\x93 Swedish" lang="sv" hreflang="sv" class="interlanguage-link-target">Svenska<li class="interlang uage-link interwiki-tr mw-list-item">T\xc3\xbcrk\xc3\xa7e\xd0\xa3\xd0\xba\xd1\x80\xd0\xb0\xd1\x 97\xd0\xbd\xd1\x81\xd1\x8c\xd0\xba\xd0\xb0<li class="interlanguage-link int erwiki-zh mw-list-item">\xe4\xb8\xad\xe6\x96\x87\n\t\t<div class="after-portlet after-por tlet-lang">Edit links</div>\n\t</div>\n\n</</pre> div>\n\n</div>\n<footer id="footer" class="mw-footer" role="contentinfo" >\n\t<ul id ="footer-info" >\n\tid="footer-info-lastmod"> This page was last edited on 10 Se ptember 2021, at 17:48 (UTC).\n\t<!i i</pre> d="footer-info-copyright">Text is available under the Creative Commons Attribution-ShareAlike License;\nadditiona l terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia\xc2\xae is a registered tradema rk of the Wikimedia Foundation, Inc., a non-profit organization.\n\n\n\t\n\t -places-privacy">Privacy policy\n\t<li id="footer-place s-about">About Wikipedia er" title="Wikipedia:General disclaimer">Disclaimers\n\t<li id="footer-plac es-contact">Contact Wikipedia \n\tMobile view\n\t pers"> Developers\n\tid="footer-places-statslink">Statistics\n\t<li id="footer-places-cookiestat ement">Cookie state ment\n\n\n\t\n\t<li id="footer-c opyrightico">\n\t<a hr ef="https://www.mediawiki.org/">\n\n\n</footer>\n\n<script>(RLQ=wi ndow.RLQ||[]).push(function(){mw.config.set({"wgPageParseReport":{"limitreport":{"cp

utime":"0.604","walltime":"0.727","ppvisitednodes":{"value":2467,"limit":1000000},"p ostexpandincludesize":{"value":78042,"limit":2097152},"templateargumentsize":{"valu e":1921,"limit":2097152},"expansiondepth":{"value":12,"limit":40},"expensivefunction count":{"value":7,"limit":500},"unstrip-depth":{"value":1,"limit":20},"unstrip-siz e":{"value":135669,"limit":5000000},"entityaccesscount":{"value":1,"limit":400},"tim 1 -total"," 45.61% 301.824 ingprofile":["100.00% 661.793 1 Template:Refl ist"," 33.71% 223.116 38 Template:Cite_web"," 12.94% 85.627 1 Template:S hort_description"," 10.33% 1 Template:Pp-pc1"," 7.63% 68.389 50.497 1 Template:When"," 6.34% Template:Pagetype"," 7.26% 48.071 41.950 1 1 Template:Authority_control"," 3.66% Template:Fix"," 5.07% 33.554 2 Template:Category_handler"]},"scribunto":{"limitreport-timeusage":{"valu e":"0.351","limit":"10.000"},"limitreport-memusage":{"value":6446910,"limit":5242880 0}},"cachereport":{"origin":"mw1395","timestamp":"20210919053121","ttl":1814400,"tra nsientcontent":false}}});;/);</script>\n<script type="application/ld+json">{"@contex t":"https:\\/\/schema.org","@type":"Article","name":"Robots exclusion standard","ur l":"https:\\/\/en.wikipedia.org\\/wiki\\/Robots_exclusion_standard","sameAs":"htt p:\\/\\/www.wikidata.org\\/entity\\/080776","mainEntity":"http:\\/\\/www.wikidata.or g\\/entity\\/Q80776","author":{"@type":"Organization","name":"Contributors to Wikime dia projects"}, "publisher": {"@type": "Organization", "name": "Wikimedia Foundation, In c.","logo":{"@type":"ImageObject","url":"https:\\/\\/www.wikimedia.org\\/static\\/im ages\\/wmf-hor-googpub.png"}},"datePublished":"2002-10-09T14:11:28Z","headline":"sta ndard used to advise web crawlers and scrapers not to index a web page or site"}</sc ript>\n<script>(RLQ=window.RLQ||[]).push(function(){mw.config.set({"wgBackendRespons eTime":160, "wgHostname": "mw1436"});});</script>\n</body></html>'

```
In [4]:
```

```
soup=bs(page.content,'html.parser')
print(soup.prettify())
```

```
<!DOCTYPE html>
<html class="client-nojs" dir="ltr" lang="en">
<head>
  <meta charset="utf-8"/>
  <title>
   Robots exclusion standard - Wikipedia
  </title>
  <script>
```

document.documentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSepa ratorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateForma t":"dmy","wgMonthNames":["","January","February","March","April","May","June","Jul y","August","September","October","November","December"],"wgRequestId":"fdd9e463-104 3-4ee4-a270-c5fe896fffed","wgCSPNonce":!1,"wgCanonicalNamespace":"","wgCanonicalSpec ialPageName":!1,"wgNamespaceNumber":0,"wgPageName":"Robots_exclusion_standard","wgTitle":"Robots exclusion standard","wgCurRevisionId":1043553668,"wgRevisionId":1043553668,"wgArticleId":101673,"wgIsArticle":!0,"wgIsRedirect":!1,"wgAction":"view","wgUse rName":null,"wgUserGroups":["*"],"wgCategories":["CS1 maint: multiple names: authors list","CS1 errors: missing periodical","Articles with short description","Short description matches Wikidata","Wikipedia pending changes protected pages","All articles with vague or ambiguous time",

"Vague or ambiguous time from December 2020", "Official website different in Wikidata and Wikipedia", "Articles with MA identifiers", "World Wide Web", "Internet bots"], "wgP ageContentLanguage": "en", "wgPageContentModel": "wikitext", "wgRelevantPageName": "Robot s_exclusion_standard", "wgRelevantArticleId":101673, "wgIsProbablyEditable":!0, "wgRele vantPageIsProbablyEditable":!0, "wgRestrictionEdit":[], "wgRestrictionMove":[], "wgFlag gedRevsParams": {"tags": {"status": {"levels":-1}}}, "wgStableRevisionId":1043553668, "wg MediaViewerOnClick":!0, "wgMediaViewerEnabledByDefault":!0, "wgPopupsFlags":10, "wgVisu alEditor": {"pageLanguageCode": "en", "pageLanguageDir": "ltr", "pageVariantFallbacks": "e n"}, "wgMFDisplayWikibaseDescriptions": {"search":!0, "nearby":!0, "watchlist":!0, "tagli ne":!1}, "wgWMESchemaEditAttemptStepOversample":!1, "wgULSCurrentAutonym": "English", "wgNoticeProject": "wikipedia", "wgEditSubmitButtonLabelPublish":!0, "wgULSPosition": "int erlanguage", "wgULSisCompactLinksEnabled":!0,

"wgGENewcomerTasksGuidanceEnabled":!0,"wgGEAskQuestionEnabled":!1,"wgGELinkRecommend
ationsFrontendEnabled":!1,"wgCentralAuthMobileDomain":!1,"wgWikibaseItemId":"Q8077
6"};RLSTATE={"ext.globalCssJs.user.styles":"ready","site.styles":"ready","noscrip
t":"ready","user.styles":"ready","ext.globalCssJs.user":"ready","user":"ready","user

```
r.options":"loading","ext.flaggedRevs.icons":"ready","oojs-ui-core.styles":"read
y","oojs-ui.styles.indicators":"ready","mediawiki.widgets.styles":"ready","oojs-ui-c
ore.icons":"ready","ext.cite.styles":"ready","ext.pygments":"ready","skins.vector.st
yles.legacy":"ready","ext.flaggedRevs.basic":"ready","ext.visualEditor.desktopArticl
eTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"rea
dy","wikibase.client.init":"ready"};RLPAGEMODULES=["ext.cite.ux-enhancements","sit
e", "mediawiki.page.ready", "mediawiki.toc", "skins.vector.legacy.js", "ext.flaggedRevs.
advanced","ext.gadget.ReferenceTooltips","ext.gadget.charinsert",
"ext.gadget.extra-toolbar-buttons", "ext.gadget.refToolbar", "ext.gadget.switcher", "mm
v.head", "mmv.bootstrap.autostart", "ext.popups", "ext.visualEditor.desktopArticleTarge
t.init","ext.visualEditor.targetLoader","ext.eventLogging","ext.wikimediaEvents","ex
t.navigationTiming","ext.uls.compactlinks","ext.uls.interface","ext.cx.eventlogging.
campaigns","ext.centralNotice.geoIP","ext.centralNotice.startUp","ext.centralauth.ce
ntralautologin","ext.growthExperiments.SuggestedEditSession"];
  </script>
  <script>
   (RLQ=window.RLQ||[]).push(function(){mw.loader.implement("user.options@1hzgi",fun
ction($,jQuery,require,module){/*@nomin*/mw.user.tokens.set({"patrolToken":"+\\","wa
tchToken":"+\\","csrfToken":"+\\"});
});});
  </script>
  <link href="/w/load.php?lang=en&amp;modules=ext.cite.styles%7Cext.flaggedRevs.basi</pre>
c%2Cicons%7Cext.pygments%2CwikimediaBadges%7Cext.uls.interlanguage%7Cext.visualEdito
r.desktopArticleTarget.noscript%7Cmediawiki.widgets.styles%7Coojs-ui-core.icons%2Cst
yles%7Coojs-ui.styles.indicators%7Cskins.vector.styles.legacy%7Cwikibase.client.init
&only=styles&skin=vector" rel="stylesheet"/>
  <script async="" src="/w/load.php?lang=en&amp;modules=startup&amp;only=scripts&am</pre>
p;raw=1&skin=vector">
  </script>
  <meta content="" name="ResourceLoaderDynamicStyles"/>
  <link href="/w/load.php?lang=en&amp;modules=site.styles&amp;only=styles&amp;skin=v</pre>
ector" rel="stylesheet"/>
  <meta content="MediaWiki 1.37.0-wmf.23" name="generator"/>
  <meta content="origin" name="referrer"/>
  <meta content="origin-when-crossorigin" name="referrer"/>
  <meta content="origin-when-cross-origin" name="referrer"/>
  <meta content="telephone=no" name="format-detection"/>
  <meta content="Robots exclusion standard - Wikipedia" property="og:title"/>
  <meta content="website" property="og:type"/>
  <link href="//upload.wikimedia.org" rel="preconnect"/>
  <link href="//en.m.wikipedia.org/wiki/Robots_exclusion_standard" media="only scree</pre>
n and (max-width: 720px)" rel="alternate"/>
  <link href="/w/index.php?title=Robots exclusion standard&amp;action=edit" rel="alt</pre>
ernate" title="Edit this page" type="application/x-wiki"/>
  <link href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit" rel="edi</pre>
t" title="Edit this page"/>
  <link href="/static/apple-touch/wikipedia.png" rel="apple-touch-icon"/>
  <link href="/static/favicon/wikipedia.ico" rel="shortcut icon"/>
  <link href="/w/opensearch_desc.php" rel="search" title="Wikipedia (en)" type="appl</pre>
ication/opensearchdescription+xml"/>
  <link href="//en.wikipedia.org/w/api.php?action=rsd" rel="EditURI" type="applicati</pre>
on/rsd+xml"/>
  <link href="//creativecommons.org/licenses/by-sa/3.0/" rel="license"/>
  <link href="https://en.wikipedia.org/wiki/Robots_exclusion_standard" rel="canonica"</pre>
  <link href="//meta.wikimedia.org" rel="dns-prefetch"/>
  <link href="//login.wikimedia.org" rel="dns-prefetch"/>
 </head>
 <body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject mw-editabl
e page-Robots exclusion standard rootpage-Robots exclusion standard skin-vector acti
on-view skin-vector-legacy">
  <div class="noprint" id="mw-page-base">
  </div>
  <div class="noprint" id="mw-head-base">
  </div>
```

```
<div class="mw-body" id="content" role="main">
   <a id="top">
   </a>
   <div id="siteNotice">
    <!-- CentralNotice -->
   </div>
   <div class="mw-indicators">
    <div class="mw-indicator" id="mw-indicator-pp-autoreview">
     <a href="/wiki/Wikipedia:Protection_policy#pending" title="All edits by unregis</pre>
tered and new users are subject to review prior to becoming visible to unregistered
users">
      <img alt="Page protected with pending changes" data-file-height="512" data-fil</pre>
e-width="512" decoding="async" height="20" src="//upload.wikimedia.org/wikipedia/en/
thumb/b/b7/Pending-protection-shackle.svg/20px-Pending-protection-shackle.svg.png" s
rcset="//upload.wikimedia.org/wikipedia/en/thumb/b/b7/Pending-protection-shackle.sv
g/30px-Pending-protection-shackle.svg.png 1.5x, //upload.wikimedia.org/wikipedia/en/
thumb/b/b7/Pending-protection-shackle.svg/40px-Pending-protection-shackle.svg.png 2
x" width="20"/>
     </a>
    </div>
   </div>
   <h1 class="firstHeading" id="firstHeading">
    Robots exclusion standard
   <div class="vector-body" id="bodyContent">
    <div class="noprint" id="siteSub">
     From Wikipedia, the free encyclopedia
    </div>
    <div id="contentSub">
     <div class="flaggedrevs_short flaggedrevs_stable_synced plainlinks noprint" id</pre>
="mw-fr-revisiontag">
      <div class="flaggedrevs_short_basic">
       <span aria-disabled="false" class="flaggedrevs-icon oo-ui-widget oo-ui-widget</pre>
-enabled oo-ui-iconElement-icon oo-ui-icon-articleSearch oo-ui-iconElement oo-ui-lab
elElement-invisible oo-ui-iconWidget" title="Changes must be reviewed before being d
isplayed on this page.">
       </span>
       <span aria-disabled="false" class="fr-toggle-arrow oo-ui-widget oo-ui-widget-</pre>
enabled oo-ui-indicatorElement-indicator oo-ui-indicator-down oo-ui-indicatorElement
oo-ui-labelElement-invisible oo-ui-indicatorWidget" id="mw-fr-revisiontoggle" title
="show/hide details">
       </span>
      </div>
      <div id="mw-fr-revisiondetails-wrapper" style="position:relative;">
       <div class="flaggedrevs_short_details" id="mw-fr-revisiondetails" style="disp</pre>
lay:none">
        This is the
        <a href="/wiki/Wikipedia:Pending changes" title="Wikipedia:Pending changes">
         latest accepted revision
        </a>
        <a class="external text" href="https://en.wikipedia.org/w/index.php?title=Sp</pre>
ecial:Log&type=review&page=Robots_exclusion_standard">
         reviewed
        </a>
        on
        <i>>
         10 September 2021
        </i>
       </div>
      </div>
     </div>
    </div>
    <div id="contentSub2">
```

```
</div>
    <div id="jump-to-nav">
    <a class="mw-jump-link" href="#mw-head">
    Jump to navigation
    </a>
    <a class="mw-jump-link" href="#searchInput">
     Jump to search
    </a>
    <div class="mw-body-content mw-content-ltr" dir="ltr" id="mw-content-text" lang</pre>
="en">
     <div class="mw-parser-output">
      <div class="shortdescription nomobile noexcerpt noprint searchaux" style="disp</pre>
lay:none">
      Standard used to advise web crawlers and scrapers not to index a web page or
site
      </div>
      >
       The
       <b>
        robots exclusion standard
       </b>
       , also known as the
        robots exclusion protocol
       </b>
      or simply
       <b>
        robots.txt
       </b>
       , is a standard used by
       <a href="/wiki/Website" title="Website">
       websites
       </a>
       to communicate with
       <a href="/wiki/Web_crawler" title="Web crawler">
       web crawlers
       </a>
       and other
       <a href="/wiki/Internet_bot" title="Internet bot">
        web robots
       </a>
       . The standard specifies how to inform the web robot about which areas of the
website should not be processed or scanned. Robots are often used by
       <a class="mw-redirect" href="/wiki/Web search engine" title="Web search engin</pre>
e">
        search engines
       to categorize websites. Not all robots cooperate with the standard;
       <a class="mw-redirect" href="/wiki/Email_address_harvesting" title="Email add</pre>
ress harvesting">
        email harvesters
       </a>
       <a class="mw-redirect" href="/wiki/Spambots" title="Spambots">
       spambots
       </a>
       <a href="/wiki/Malware" title="Malware">
       malware
       and robots that scan for security vulnerabilities may even start with the por
tions of the website where they have been told to stay out. The standard can be used
```

```
in conjunction with
     <a href="/wiki/Sitemaps" title="Sitemaps">
     </a>
     , a robot inclusion standard for websites.
     <div aria-labelledby="mw-toc-heading" class="toc" id="toc" role="navigation">
     <input class="toctogglecheckbox" id="toctogglecheckbox" role="button" style</pre>
="display:none" type="checkbox"/>
     <div class="toctitle" dir="ltr" lang="en">
      <h2 id="mw-toc-heading">
       Contents
      </h2>
      <span class="toctogglespan">
       <label class="toctogglelabel" for="toctogglecheckbox">
       </label>
      </span>
     </div>
     <l
      <a href="#History">
        <span class="tocnumber">
         1
        </span>
        <span class="toctext">
        History
        </span>
       </a>
      <a href="#Standard">
        <span class="tocnumber">
         2
        </span>
        <span class="toctext">
         Standard
        </span>
       </a>
      <a href="#Security">
        <span class="tocnumber">
         3
        </span>
        <span class="toctext">
         Security
        </span>
       </a>
      <a href="#Alternatives">
        <span class="tocnumber">
         4
        </span>
        <span class="toctext">
         Alternatives
        </span>
       </a>
      <a href="#Examples">
        <span class="tocnumber">
         5
        </span>
        <span class="toctext">
```

```
Examples
 </span>
</a>
<a href="#Nonstandard extensions">
 <span class="tocnumber">
  6
 </span>
 <span class="toctext">
  Nonstandard extensions
 </span>
</a>
<l
 <a href="#Crawl-delay_directive">
  <span class="tocnumber">
   6.1
  </span>
   <span class="toctext">
   Crawl-delay directive
  </span>
  </a>
 <a href="#Allow_directive">
  <span class="tocnumber">
   6.2
   </span>
   <span class="toctext">
   Allow directive
  </span>
  </a>
 <a href="#Sitemap">
  <span class="tocnumber">
   6.3
   </span>
  <span class="toctext">
   Sitemap
  </span>
  </a>
 <a href="#Host">
   <span class="tocnumber">
   6.4
   </span>
   <span class="toctext">
   Host
   </span>
  </a>
 <a href='#Universal "*" match'>
  <span class="tocnumber">
   6.5
   </span>
   <span class="toctext">
   Universal "*" match
  </span>
  </a>
```

```
<a href="#Meta_tags_and_headers">
        <span class="tocnumber">
         7
        </span>
        <span class="toctext">
         Meta tags and headers
        </span>
       </a>
       <a href="#See_also">
        <span class="tocnumber">
         8
        </span>
        <span class="toctext">
         See also
        </span>
       </a>
       <a href="#References">
        <span class="tocnumber">
         9
        </span>
        <span class="toctext">
         References
        </span>
       </a>
       <a href="#External_links">
        <span class="tocnumber">
         10
        </span>
        <span class="toctext">
         External links
        </span>
       </a>
      </div>
     <h2>
      <span class="mw-headline" id="History">
      History
      </span>
      <span class="mw-editsection">
      <span class="mw-editsection-bracket">
       [
      </span>
      <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=1" title="Edit section: History">
       edit
      </a>
      <span class="mw-editsection-bracket">
      </span>
      </span>
     </h2>
     >
     The standard was proposed by
      <a href="/wiki/Martijn_Koster" title="Martijn Koster">
      Martijn Koster
      </a>
```

```
<sup class="reference" id="cite_ref-1">
        <a href="#cite note-1">
         [1]
        </a>
       </sup>
       <sup class="reference" id="cite_ref-2">
        <a href="#cite_note-2">
         [2]
        </a>
       </sup>
       when working for
       <a href="/wiki/Nexor" title="Nexor">
       </a>
       <sup class="reference" id="cite_ref-:0_3-0">
        <a href="#cite note-:0-3">
         [3]
        </a>
       </sup>
       in February 1994
       <sup class="reference" id="cite_ref-4">
        <a href="#cite_note-4">
         [4]
        </a>
       </sup>
       on the
       <i>>
        www-talk
       </i>
       mailing list, the main communication channel for WWW-related activities at th
       <a href="/wiki/Charles_Stross" title="Charles Stross">
       Charles Stross
       </a>
       claims to have provoked Koster to suggest robots.txt, after he wrote a badly-
behaved web crawler that inadvertently caused a
       <a href="/wiki/Denial-of-service_attack" title="Denial-of-service attack">
        denial-of-service attack
       </a>
       on Koster's server.
       <sup class="reference" id="cite_ref-5">
        <a href="#cite note-5">
         [5]
        </a>
       </sup>
      >
       It quickly became a
       <a href="/wiki/De_facto_standard" title="De facto standard">
       de facto standard
       </a>
       that present and future web crawlers were expected to follow; most complied,
including those operated by search engines such as
       <a href="/wiki/WebCrawler" title="WebCrawler">
       WebCrawler
       </a>
       <a href="/wiki/Lycos" title="Lycos">
        Lycos
       </a>
       , and
       <a href="/wiki/AltaVista" title="AltaVista">
        AltaVista
       </a>
```

```
<sup class="reference" id="cite_ref-sear_Robo_6-0">
        <a href="#cite_note-sear_Robo-6">
         [6]
        </a>
       </sup>
      >
       On July 1, 2019, Google announced the proposal of the Robots Exclusion Protoc
ol as an official standard under
       <a href="/wiki/Internet_Engineering_Task_Force" title="Internet Engineering T</pre>
ask Force">
        Internet Engineering Task Force
       </a>
       <sup class="reference" id="cite_ref-7">
        <a href="#cite note-7">
         [7]
        </a>
       </sup>
       The draft
       <sup class="reference" id="cite_ref-8">
        <a href="#cite_note-8">
         [8]
        </a>
       </sup>
       will now undergo the acceptance process.
      <h2>
       <span class="mw-headline" id="Standard">
        Standard
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
         </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=2" title="Edit section: Standard">
         edit
        </a>
        <span class="mw-editsection-bracket">
        </span>
       </span>
      </h2>
      >
       When a site owner wishes to give instructions to web robots they place a text
       <style data-mw-deduplicate="TemplateStyles:r886049734">
        .mw-parser-output .monospaced{font-family:monospace, monospace}
       </style>
       <span class="monospaced">
        robots.txt
       </span>
       in the root of the web site hierarchy (e.g.
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
vle"/>
       <span class="monospaced">
        https://www.example.com/robots.txt
       ). This text file contains the instructions in a specific format (see example
s below). Robots that
       <i>>
        choose
       </i>
```

```
to follow the instructions try to fetch this file and read the instructions b
efore fetching any other file from the
       <a href="/wiki/Website" title="Website">
        website
       </a>
       . If this file doesn't exist, web robots assume that the website owner does n
ot wish to place any limitations on crawling the entire site.
      >
       A robots.txt file on a website will function as a request that specified robo
ts ignore specified files or directories when crawling a site. This might be, for ex
ample, out of a preference for privacy from search engine results, or the belief tha
t the content of the selected directories might be misleading or irrelevant to the c
ategorization of the site as a whole, or out of a desire that an application only op
erates on certain data. Links to pages listed in robots.txt can still appear in sear
ch results if they are linked to from a page that is crawled.
       <sup class="reference" id="cite ref-9">
        <a href="#cite_note-9">
         [9]
        </a>
       </sup>
      >
       A robots.txt file covers one
       <a class="mw-redirect" href="/wiki/Same_origin_policy" title="Same origin pol</pre>
icy">
        origin
       . For websites with multiple subdomains, each subdomain must have its own rob
ots.txt file. If
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        example.com
       </span>
       had a robots.txt file but
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        a.example.com
       </span>
       did not, the rules that would apply for
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        example.com
       </span>
       would not apply to
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        a.example.com
       </span>
       . In addition, each protocol and port needs its own robots.txt file;
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        http://example.com/robots.txt
       </span>
       does not apply to pages under
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        http://example.com:8080/
       </span>
```

```
or
       <link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st</pre>
yle"/>
       <span class="monospaced">
        https://example.com/
       </span>
      >
       Some major search
       <a href="/wiki/Facebook_Paper" title="Facebook Paper">
        engines
       </a>
       following this standard include Ask,
       <sup class="reference" id="cite_ref-ask-webmasters_10-0">
        <a href="#cite_note-ask-webmasters-10">
         [10]
        </a>
       </sup>
       AOL,
       <sup class="reference" id="cite_ref-about-aol-search_11-0">
        <a href="#cite_note-about-aol-search-11">
         [11]
        </a>
       </sup>
       Baidu,
       <sup class="reference" id="cite_ref-baidu-spider_12-0">
        <a href="#cite note-baidu-spider-12">
         [12]
        </a>
       </sup>
       DuckDuckGo,
       <sup class="reference" id="cite_ref-duckduckgo-bot_13-0">
        <a href="#cite_note-duckduckgo-bot-13">
        </a>
       </sup>
       Google,
       <sup class="reference" id="cite ref-google-webmasters-spec 14-0">
        <a href="#cite_note-google-webmasters-spec-14">
         [14]
        </a>
       </sup>
       Yahoo!,
       <sup class="reference" id="cite_ref-yahoo-search-is-bing_15-0">
        <a href="#cite note-yahoo-search-is-bing-15">
         [15]
        </a>
       </sup>
       and Yandex.
       <sup class="reference" id="cite_ref-yandex-robots_16-0">
        <a href="#cite_note-yandex-robots-16">
         [16]
        </a>
       </sup>
       <sup class="reference" id="cite ref-bing-blog-robots 17-0">
        <a href="#cite note-bing-blog-robots-17">
         [17]
        </a>
       </sup>
       is still
       <sup class="noprint Inline-Template" style="white-space:nowrap;">
        Γ
        <i>>
```

```
<a href="/wiki/Wikipedia:Manual_of_Style/Dates_and_numbers#Chronological_it</pre>
ems" title="Wikipedia:Manual of Style/Dates and numbers">
          <span title="The time period mentioned near this tag is ambiguous. (Decemb</pre>
er 2020)">
           when?
          </span>
         </a>
        </i>
        1
       </sup>
       not fully compatible with the standard as it cannot inherit settings from the
wildcard (
       <code>
       </code>
       ).
       <sup class="reference" id="cite ref-18">
        <a href="#cite_note-18">
         [18]
        </a>
       </sup>
      >
       The volunteering group Archive Team explicitly ignores robots.txt for the mos
t part, viewing it as an obsolete standard that hinders web archival efforts. Accord
ing to project leader Jason Scott, "unchecked, and left alone, the robots.txt file e
nsures no mirroring or reference for items that may have general use and meaning bey
ond the website's context."
       <sup class="reference" id="cite_ref-19">
        <a href="#cite_note-19">
         [19]
        </a>
       </sup>
      >
       For some years, the
       <a href="/wiki/Internet_Archive" title="Internet Archive">
       Internet Archive
       did not crawl sites with robots.txt, but in April 2017, it announced that it
would no longer honour directives in the robots.txt files.
       <sup class="reference" id="cite_ref-20">
        <a href="#cite_note-20">
         [20]
        </a>
       </sup>
       "Over time we have observed that the robots.txt files that are geared toward
search engine crawlers do not necessarily serve our archival purposes".
       <sup class="reference" id="cite ref-Internet Archive 21-0">
        <a href="#cite note-Internet Archive-21">
         [21]
        </a>
       </sup>
       This was in response to entire domains being tagged with robots.txt when the
content became obsolete.
       <sup class="reference" id="cite ref-Internet Archive 21-1">
        <a href="#cite note-Internet Archive-21">
         [21]
        </a>
       </sup>
      <h2>
       <span class="mw-headline" id="Security">
        Security
       </span>
```

```
<span class="mw-editsection">
        <span class="mw-editsection-bracket">
         Γ
        </span>
        <a href="/w/index.php?title=Robots exclusion standard&amp;action=edit&amp;se</pre>
ction=3" title="Edit section: Security">
         edit
        </a>
        <span class="mw-editsection-bracket">
        </span>
       </span>
      </h2>
      >
       Despite the use of the terms "allow" and "disallow", the protocol is purely a
dvisory and relies on the compliance of the
       <a class="mw-redirect" href="/wiki/Web robot" title="Web robot">
        web robot
       </a>
       <sup class="reference" id="cite_ref-22">
        <a href="#cite_note-22">
         [22]
        </a>
       </sup>
       Malicious web robots are unlikely to honor robots.txt; some may even use the
robots.txt as a guide to find disallowed links and go straight to them. While this i
s sometimes claimed to be a security risk,
       <sup class="reference" id="cite_ref-23">
        <a href="#cite_note-23">
         [23]
        </a>
       </sup>
       this sort of
       <a href="/wiki/Security_through_obscurity" title="Security through obscurit</pre>
y">
        security through obscurity
       </a>
       is discouraged by standards bodies. The
       <a href="/wiki/National_Institute_of_Standards_and_Technology" title="Nationa</pre>
1 Institute of Standards and Technology">
        National Institute of Standards and Technology
       </a>
       (NIST) in the United States specifically recommends against this practice: "S
ystem security should not depend on the secrecy of the implementation or its compone
nts."
       <sup class="reference" id="cite ref-24">
        <a href="#cite note-24">
         [24]
        </a>
       </sup>
       In the context of robots.txt files, security through obscurity is not recomme
nded as a security technique.
       <sup class="reference" id="cite ref-25">
        <a href="#cite_note-25">
         [25]
        </a>
       </sup>
      <h2>
       <span class="mw-headline" id="Alternatives">
        Alternatives
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
```

```
</span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=4" title="Edit section: Alternatives">
        edit
        </a>
        <span class="mw-editsection-bracket">
        ]
        </span>
       </span>
      </h2>
      >
      Many robots also pass a special
       <a class="mw-redirect" href="/wiki/User-agent" title="User-agent">
       user-agent
       </a>
      to the web server when fetching content.
       <sup class="reference" id="cite ref-26">
       <a href="#cite_note-26">
         [26]
       </a>
       </sup>
       A web administrator could also configure the server to automatically return f
       <a href="/wiki/Cloaking" title="Cloaking">
       pass alternative content
       </a>
       ) when it detects a connection using one of the robots.
       <sup class="reference" id="cite_ref-27">
       <a href="#cite_note-27">
         [27]
       </a>
       </sup>
       <sup class="reference" id="cite_ref-28">
        <a href="#cite_note-28">
         [28]
        </a>
       </sup>
      >
      Some sites, such as
       <a href="/wiki/Google" title="Google">
       Google
       </a>
       , host a
       <code>
       humans.txt
       </code>
       file that displays information meant for humans to read.
       <sup class="reference" id="cite ref-29">
       <a href="#cite_note-29">
        [29]
       </a>
       </sup>
       Some sites such as
       <a href="/wiki/GitHub" title="GitHub">
       GitHub
       </a>
       redirect humans.txt to an about page.
       <sup class="reference" id="cite ref-30">
       <a href="#cite_note-30">
         [30]
        </a>
       </sup>
```

```
>
      Previously Google had a joke file hosted at
        /killer-robots.txt
       </code>
       instructing
       <a href="/wiki/Terminator_(character)" title="Terminator (character)">
       the Terminator
       </a>
       not to kill the company founders
       <a href="/wiki/Larry_Page" title="Larry Page">
        Larry Page
       </a>
       and
       <a href="/wiki/Sergey_Brin" title="Sergey Brin">
       Sergey Brin
       </a>
       <sup class="reference" id="cite_ref-31">
        <a href="#cite_note-31">
         [31]
        </a>
       </sup>
       <sup class="reference" id="cite_ref-32">
        <a href="#cite_note-32">
         [32]
        </a>
      </sup>
      <span class="mw-headline" id="Examples">
       Examples
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=5" title="Edit section: Examples">
        edit
        </a>
        <span class="mw-editsection-bracket">
         ]
        </span>
       </span>
      </h2>
      < g>>
      This example tells all robots that they can visit all files because the wildc
ard
       <code>
        *
       </code>
       stands for all robots and the
       <code>
       Disallow
      directive has no value, meaning no pages are disallowed.
      User-agent: *
Allow: /
The same result can be accomplished with an empty or missing robots.txt file.
      >
```

```
This example tells all robots to stay out of a website:
     User-agent: *
Disallow: /
This example tells all robots not to enter three directories:
     User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /junk/
>
      This example tells all robots to stay away from one specific file:
     User-agent: *
Disallow: /directory/file.html
All other files in the specified directory will be processed.
     >
      This example tells a specific robot to stay out of a website:
     User-agent: BadBot # replace 'BadBot' with the actual user-agent of the b
ot
Disallow: /
>
      This example tells two specific robots not to enter one specific directory:
     User-agent: BadBot # replace 'BadBot' with the actual user-agent of the b
ot
User-agent: Googlebot
Disallow: /private/
>
      Example demonstrating how comments can be used:
     # Comments appear after the "#" symbol at the start of a line, or after a
User-agent: * # match all bots
Disallow: / # keep them out
>
      It is also possible to list multiple robots with their own rules. The actual
robot string is defined by the crawler. A few robot operators, such as
      <a href="/wiki/Google" title="Google">
       Google
      </a>
      , support several user-agent strings that allow the operator to deny access t
o a subset of their services by using specific user-agent strings.
      <sup class="reference" id="cite_ref-google-webmasters-spec_14-1">
       <a href="#cite_note-google-webmasters-spec-14">
        [14]
       </a>
      </sup>
     Example demonstrating multiple user-agents:
     User-agent: googlebot
                                      # all Google services
Disallow: /private/
                           # disallow this directory
```

```
User-agent: googlebot-news # only the news service
Disallow: /
                             # disallow everything
User-agent: *
                             # any robot
Disallow: /something/
                             # disallow this directory
<span class="mw-headline" id="Nonstandard_extensions">
       Nonstandard extensions
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
         </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=6" title="Edit section: Nonstandard extensions">
        edit
        </a>
        <span class="mw-editsection-bracket">
       </span>
       </span>
      </h2>
      <h3>
       <span class="mw-headline" id="Crawl-delay_directive">
       Crawl-delay directive
       </span>
       <span class="mw-editsection">
       <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=7" title="Edit section: Crawl-delay directive">
        edit
        </a>
        <span class="mw-editsection-bracket">
        ]
        </span>
       </span>
      </h3>
       The crawl-delay value is supported by some crawlers to throttle their visits
to the host. Since this value is not part of the standard, its interpretation is dep
endent on the crawler reading it. It is used when the multiple burst of visits from
bots is slowing down the host. Yandex interprets the value as the number of seconds
to wait between subsequent visits.
       <sup class="reference" id="cite ref-yandex-robots 16-1">
        <a href="#cite note-yandex-robots-16">
         [16]
        </a>
       </sup>
       Bing defines crawl-delay as the size of a time window (from 1 to 30 seconds)
during which BingBot will access a web site only once.
       <sup class="reference" id="cite_ref-bing-crawl-delay_33-0">
        <a href="#cite_note-bing-crawl-delay-33">
         [33]
       </a>
       </sup>
       Google provides an interface in its
       <a href="/wiki/Google Search Console" title="Google Search Console">
       search console
       </a>
       for webmasters, to control the
       <a href="/wiki/Googlebot" title="Googlebot">
       Googlebot
```

```
</a>
       's subsequent visits.
       <sup class="reference" id="cite ref-34">
        <a href="#cite_note-34">
         [34]
        </a>
       </sup>
      User-agent: bingbot
Crawl-delay: 10
<span class="mw-headline" id="Allow_directive">
       Allow directive
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=8" title="Edit section: Allow directive">
         edit
        </a>
        <span class="mw-editsection-bracket">
        </span>
       </span>
      </h3>
      >
      Some major crawlers support an
       <code>
       Allow
       </code>
       directive, which can counteract a following
       <code>
       Disallow
       </code>
       directive.
       <sup class="reference" id="cite_ref-35">
        <a href="#cite_note-35">
         [35]
        </a>
       </sup>
       <sup class="reference" id="cite_ref-36">
        <a href="#cite note-36">
         [36]
        </a>
       </sup>
       This is useful when one tells robots to avoid an entire directory but still w
ants some HTML documents in that directory crawled and indexed. While by standard im
plementation the first matching robots.txt pattern always wins, Google's implementat
ion differs in that Allow patterns with equal or more characters in the directive pa
th win over a matching Disallow pattern.
       <sup class="reference" id="cite_ref-37">
        <a href="#cite note-37">
         [37]
        </a>
       </sup>
       Bing uses either the
       <code>
        Allow
       </code>
       or
       <code>
```

```
Disallow
       </code>
       directive, whichever is more specific, based on length, like Google.
       <sup class="reference" id="cite_ref-bing-blog-robots_17-1">
        <a href="#cite note-bing-blog-robots-17">
         [17]
        </a>
       </sup>
      >
      In order to be compatible to all robots, if one wants to allow single files i
nside an otherwise disallowed directory, it is necessary to place the Allow directiv
e(s) first, followed by the Disallow, for example:
      Allow: /directory1/myfile.html
Disallow: /directory1/
>
       This example will Disallow anything in /directory1/ except /directory1/myfil
e.html, since the latter will match first. The order is only important to robots tha
t follow the standard; in the case of the Google or Bing bots, the order is not impo
rtant.
      <h3>
       <span class="mw-headline" id="Sitemap">
       Sitemap
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=9" title="Edit section: Sitemap">
        edit
        </a>
        <span class="mw-editsection-bracket">
        ]
        </span>
       </span>
      </h3>
      Some crawlers support a
       <code>
       Sitemap
       </code>
       directive, allowing multiple
       <a href="/wiki/Sitemaps" title="Sitemaps">
       Sitemaps
       </a>
       in the same
       <samp>
       robots.txt
       </samp>
       in the form
       <code>
        Sitemap:
        <i>>
         full-url
        </i>
       </code>
       <sup class="reference" id="cite_ref-38">
        <a href="#cite_note-38">
         [38]
        </a>
```

```
</sup>
      Sitemap: http://www.example.com/sitemap.xml
       <span class="mw-headline" id="Host">
       Host
       </span>
       <span class="mw-editsection">
       <span class="mw-editsection-bracket">
        Γ
       </span>
       <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=10" title="Edit section: Host">
        edit
       </a>
       <span class="mw-editsection-bracket">
        1
       </span>
       </span>
      </h3>
      >
      Some crawlers (
      <a href="/wiki/Yandex_Search" title="Yandex Search">
       Yandex
       </a>
       ) support a
       <code>
       Host
       </code>
      directive, allowing websites with multiple mirrors to specify their preferred
domain:
       <sup class="reference" id="cite ref-39">
       <a href="#cite_note-39">
        [39]
       </a>
       </sup>
      Host: hosting.example.com
>
      This is not supported by all crawlers.
      <h3>
      <span id="Universal_.22.2A.22_match">
       <span class="mw-headline" id='Universal "*" match'>
       Universal "*" match
       </span>
       <span class="mw-editsection">
       <span class="mw-editsection-bracket">
        [
       </span>
       <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=11" title='Edit section: Universal "*" match'>
        edit
        </a>
       <span class="mw-editsection-bracket">
        ]
       </span>
      </span>
      </h3>
      >
      The
       <i>>
       Robot Exclusion Standard
```

```
</i>
       does not mention the "*" character in the
       Disallow:
       </code>
       statement.
       <sup class="reference" id="cite ref-40">
       <a href="#cite_note-40">
         [40]
       </a>
       </sup>
      <h2>
       <span class="mw-headline" id="Meta_tags_and_headers">
       Meta tags and headers
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=12" title="Edit section: Meta tags and headers">
        edit
        </a>
        <span class="mw-editsection-bracket">
        </span>
       </span>
      </h2>
      >
      In addition to root-level robots.txt files, robots exclusion directives can b
e applied at a more granular level through the use of
       <a class="mw-redirect" href="/wiki/Robots_meta_tag" title="Robots meta tag">
       Robots meta tags
       </a>
       and X-Robots-Tag HTTP headers. The robots meta tag cannot be used for non-HTM
L files such as images, text files, or PDF documents. On the other hand, the X-Robot
s-Tag can be added to non-HTML files by using
       <a href="/wiki/.htaccess" title=".htaccess">
        .htaccess
       </a>
       and
       <a class="mw-redirect" href="/wiki/Httpd.conf" title="Httpd.conf">
       httpd.conf
       </a>
      files.
       <sup class="reference" id="cite ref-google-meta 41-0">
       <a href="#cite note-google-meta-41">
         [41]
       </a>
       </sup>
      <d1>
       A "noindex" meta tag
       </dt>
      </dl>
      <div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr">
       <span></span><span class="p">&lt;</span><span class="nt">meta</span> <sp
an class="na">name</span><span class="o">=</span><span class="s">"robots"</span> <sp
an class="na">content</span><span class="o">=</span><span class="s">"noindex"</span>
<span class="p">/&gt;</span>
</div>
      <d1>
```

```
<dt>
       A "noindex" HTTP response header
       </dt>
      </dl>
      <div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr">
       </span></span>X-Robots-Tag: noindex
</div>
      >
       The X-Robots-Tag is only effective after the page has been requested and the
server responds, and the robots meta tag is only effective after the page has loade
d, whereas robots.txt is effective before the page is requested. Thus if a page is e
xcluded by a robots.txt file, any robots meta tags or X-Robots-Tag headers are effec
tively ignored because the robot will not see them in the first place.
       <sup class="reference" id="cite_ref-google-meta_41-1">
       <a href="#cite_note-google-meta-41">
         [41]
        </a>
       </sup>
      <h2>
       <span class="mw-headline" id="See_also">
       See also
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=13" title="Edit section: See also">
        edit
        </a>
        <span class="mw-editsection-bracket">
         ]
        </span>
       </span>
      </h2>
      <style data-mw-deduplicate="TemplateStyles:r998391716">
       .mw-parser-output .div-col{margin-top:0.3em;column-width:30em}.mw-parser-outp
ut .div-col-small{font-size:90%}.mw-parser-output .div-col-rules{column-rule:1px sol
id #aaa}.mw-parser-output .div-col dl,.mw-parser-output .div-col ol,.mw-parser-outpu
t .div-col ul{margin-top:0}.mw-parser-output .div-col li,.mw-parser-output .div-col
dd{page-break-inside:avoid;break-inside:avoid-column}
      </style>
      <div class="div-col" style="column-width: 30em;">
       <l
        <
          <a href="/wiki/Ads.txt" title="Ads.txt">
          ads.txt
          </a>
         </code>
         , a standard for listing authorized ad sellers
        <1i>>
          <a href="/wiki/Security.txt" title="Security.txt">
          security.txt
          </a>
         </code>
         , a file to describe the process for security researchers to follow in orde
r to report security vulnerabilities
        <
         <a href="/wiki/Automated_Content_Access_Protocol" title="Automated Content</pre>
```

```
Access Protocol">
         Automated Content Access Protocol
        - a failed proposal to extend robots.txt
       <
        <a href="/wiki/BotSeer" title="BotSeer">
         BotSeer
        </a>
        - now inactive search engine for robots.txt files
       <
        <a href="/wiki/Distributed_web_crawling" title="Distributed web crawling">
         Distributed web crawling
        </a>
       <
        <a href="/wiki/Focused_crawler" title="Focused crawler">
         Focused crawler
        </a>
       <
        <a href="/wiki/Internet_Archive" title="Internet Archive">
         Internet Archive
        </a>
       <
        <a href="/wiki/National_Digital_Library_Program" title="National Digital Li</pre>
brary Program">
         National Digital Library Program
        </a>
        (NDLP)
       <1i>>
        <a href="/wiki/National_Digital_Information_Infrastructure_and_Preservation</pre>
_Program" title="National Digital Information Infrastructure and Preservation Progra
m">
         National Digital Information Infrastructure and Preservation Program
        </a>
        (NDIIPP)
       <
        <a href="/wiki/Nofollow" title="Nofollow">
         Nofollow
        </a>
       <
        <a href="/wiki/Perma.cc" title="Perma.cc">
         Perma.cc
        </a>
       <a class="mw-redirect" href="/wiki/Robots meta tag" title="Robots meta ta</pre>
g">
         Meta elements
        </a>
        for search engines
       <
        <a href="/wiki/Sitemaps" title="Sitemaps">
         Sitemaps
        </a>
       <
        <a href="/wiki/Spider_trap" title="Spider trap">
```

```
Spider trap
         </a>
        <
         <a href="/wiki/Web archiving" title="Web archiving">
         Web archiving
         </a>
        <1i>>
         <a href="/wiki/Web crawler" title="Web crawler">
         Web crawler
         </a>
        <1i>>
         <a href="/wiki/Noindex" title="Noindex">
         noindex
         </a>
       </div>
      <h2>
       <span class="mw-headline" id="References">
       References
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=14" title="Edit section: References">
        edit
        </a>
        <span class="mw-editsection-bracket">
         ]
        </span>
       </span>
      </h2>
      <style data-mw-deduplicate="TemplateStyles:r1011085734">
       .mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;list-style-type:
decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bottom:0;list-s
tyle-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30em}.mw-parser-
output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-columns{margi
n-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .re
flist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mw-parser-output
.reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .reflist-upper-ro
man{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha{list-style-ty
pe:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type:lower-greek}.m
w-parser-output .reflist-lower-roman{list-style-type:lower-roman}
      </style>
      <div class="reflist reflist-columns references-column-width" style="column-wid")</pre>
th: 30em;">
       id="cite note-1">
         <span class="mw-cite-backlink">
          <h>>
           <a href="#cite ref-1">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <style data-mw-deduplicate="TemplateStyles:r999302996">
           .mw-parser-output cite.citation{font-style:inherit}.mw-parser-output .cit
ation q{quotes:"\"""\""""\"}.mw-parser-output .id-lock-free a,.mw-parser-output .c
itation .cs1-lock-free a{background:linear-gradient(transparent,transparent),url("//
```

```
upload.wikimedia.org/wikipedia/commons/6/65/Lock-green.svg")right 0.1em center/9px n
o-repeat}.mw-parser-output .id-lock-limited a,.mw-parser-output .id-lock-registratio
n a,.mw-parser-output .citation .cs1-lock-limited a,.mw-parser-output .citation .cs1
-lock-registration a{background:linear-gradient(transparent,transparent),url("//uplo
ad.wikimedia.org/wikipedia/commons/d/d6/Lock-gray-alt-2.svg")right 0.1em center/9px
no-repeat \ .mw-parser-output .id-lock-subscription a, .mw-parser-output .citation .cs1
-lock-subscription a{background:linear-gradient(transparent,transparent),url("//uplo
ad.wikimedia.org/wikipedia/commons/a/aa/Lock-red-alt-2.svg")right 0.1em center/9px n
o-repeat}.mw-parser-output .cs1-subscription,.mw-parser-output .cs1-registration{col
or:#555}.mw-parser-output .cs1-subscription span,.mw-parser-output .cs1-registration
span{border-bottom:1px dotted;cursor:help}.mw-parser-output .cs1-ws-icon a{backgroun
d:linear-gradient(transparent,transparent),url("//upload.wikimedia.org/wikipedia/com
mons/4/4c/Wikisource-logo.svg")right 0.1em center/12px no-repeat}.mw-parser-output c
ode.cs1-code{color:inherit;background:inherit;border:none;padding:inherit}.mw-parser
-output .cs1-hidden-error{display:none;font-size:100%}.mw-parser-output .cs1-visible
-error{font-size:100%}.mw-parser-output .cs1-maint{display:none;color:#33aa33;margin
-left:0.3em}.mw-parser-output .cs1-format{font-size:95%}.mw-parser-output .cs1-kern-
left,.mw-parser-output .cs1-kern-wl-left{padding-left:0.2em}.mw-parser-output .cs1-k
ern-right,.mw-parser-output .cs1-kern-wl-right{padding-right:0.2em}.mw-parser-output
.citation .mw-selflink{font-weight:inherit}
          </style>
          <cite class="citation web cs1">
           <a class="external text" href="http://www.greenhills.co.uk/historical.htm</pre>
l" rel="nofollow">
            "Historical"
           </a>
           <i>>
            Greenhills.co.uk
           </i>
           <a class="external text" href="https://web.archive.org/web/2017040315203</pre>
7/http://www.greenhills.co.uk/historical.html" rel="nofollow">
           Archived
           </a>
           from the original on 2017-04-03
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            2017-03-03
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Greenhills.co.uk&am
p;rft.atitle=Historical&rft_id=http%3A%2F%2Fwww.greenhills.co.uk%2Fhistorical.ht
ml&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite note-2">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite_ref-2">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1" id="CITEREFFielding1994">
           Fielding, Roy (1994).
           <a class="external text" href="http://www94.web.cern.ch/WWW94/PapersWWW9</pre>
```

```
4/fielding.ps" rel="nofollow">
            "Maintaining Distributed Hypertext Infostructures: Welcome to MOMspide
r's Web"
           </a>
           <span class="cs1-format">
            (PostScript)
           </span>
           <i>>
            First International Conference on the World Wide Web
           </i>
           . Geneva.
           <a class="external text" href="https://web.archive.org/web/2013092709365</pre>
8/http://www94.web.cern.ch/WWW94/PapersWWW94/fielding.ps" rel="nofollow">
           Archived
           </a>
           from the original on 2013-09-27
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            September 25,
            </span>
            2013
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=First+International
+Conference+on+the+World+Wide+Web&rft.atitle=Maintaining+Distributed+Hypertext+I
nfostructures%3A+Welcome+to+MOMspider%27s+Web&rft.date=1994&rft.aulast=Field
ing&rft.aufirst=Roy&rft_id=http%3A%2F%2Fwww94.web.cern.ch%2FWWW94%2FPapersWW
W94%2Ffielding.ps&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standa
rd">
          </span>
         </span>
        <span class="mw-cite-backlink">
           <a href="#cite_ref-:0_3-0">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="http://www.robotstxt.org/orig.html#status"</pre>
rel="nofollow">
            "The Web Robots Pages"
           </a>
           . Robotstxt.org. 1994-06-30.
           <a class="external text" href="https://web.archive.org/web/2014011209063</pre>
3/http://www.robotstxt.org/orig.html#status" rel="nofollow">
           Archived
           </a>
           from the original on 2014-01-12
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             2013-12-29
            </span>
           </span>
```

```
</cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=The+Web+Robots+Pages&a
mp;rft.pub=Robotstxt.org&rft.date=1994-06-30&rft_id=http%3A%2F%2Fwww.robotst
xt.org%2Forig.html%23status&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclus
ion+standard">
          </span>
        </span>
        id="cite note-4">
        <span class="mw-cite-backlink">
           <a href="#cite_ref-4">
          </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1" id="CITEREFKoster1994">
          Koster, Martijn (25 February 1994).
           <a class="external text" href="https://web.archive.org/web/2013102920035</pre>
0/http://inkdroid.org/tmp/www-talk/4113.html" rel="nofollow">
            "Important: Spiders, Robots and Web Wanderers"
           </a>
           <i>>
           www-talk mailing list
          </i>
           . Archived from
          <a class="external text" href="http://inkdroid.org/tmp/www-talk/4113.htm</pre>
l" rel="nofollow">
           the original
           </a>
           <span class="cs1-format">
            <a href="/wiki/Hypermail" title="Hypermail">
            Hypermail
            </a>
            archived message)
           </span>
          on October 29, 2013.
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=www-talk+mailing+li
st&rft.atitle=Important%3A+Spiders%2C+Robots+and+Web+Wanderers&rft.date=1994
-02-25&rft.aulast=Koster&rft.aufirst=Martijn&rft id=http%3A%2F%2Finkdroi
d.org%2Ftmp%2Fwww-talk%2F4113.html&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots
+exclusion+standard">
          </span>
        </span>
        id="cite note-5">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite ref-5">
          </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
```

```
<a class="external text" href="http://www.antipope.org/charlie/blog-stati</pre>
c/2009/06/how_i_got_here_in_the_end_part_3.html" rel="nofollow">
            "How I got here in the end, part five: "things can only get better!
            <span class="cs1-kern-right">
            </span>
           </a>
           <i>>
           Charlie's Diary
           </i>
           . 19 June 2006.
           <a class="external text" href="https://web.archive.org/web/2013112522091</pre>
3/http://www.antipope.org/charlie/blog-static/2009/06/how_i_got_here_in_the_end_part
3.html" rel="nofollow">
           Archived
           </a>
           from the original on 2013-11-25
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            19 April
           </span>
           2014
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Charlie%27s+Diary&a
mp;rft.atitle=How+I+got+here+in+the+end%2C+part+five%3A+%22things+can+only+get+bette
r%21%22&rft.date=2006-06-19&rft_id=http%3A%2F%2Fwww.antipope.org%2Fcharlie%2
Fblog-static%2F2009%2F06%2Fhow_i_got_here_in_the_end_part_3.html&rfr_id=info%3As
id%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        <span class="mw-cite-backlink">
          <b>
           <a href="#cite_ref-sear_Robo_6-0">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1" id="CITEREFBarry Schwartz2014">
           Barry Schwartz (30 June 2014).
           <a class="external text" href="http://searchengineland.com/robots-txt-cel</pre>
ebrates-20-years-blocking-search-engines-195479" rel="nofollow">
            "Robots.txt Celebrates 20 Years Of Blocking Search Engines"
           </a>
           <i>>
           Search Engine Land
           </i>
           <a class="external text" href="https://web.archive.org/web/2015090700043</pre>
0/http://searchengineland.com/robots-txt-celebrates-20-years-blocking-search-engines
-195479" rel="nofollow">
           Archived
           </a>
           from the original on 2015-09-07
```

```
<span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2015-11-19
           </span>
          </span>
         </cite>
         <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Search+Engine+Land&
amp;rft.atitle=Robots.txt+Celebrates+20+Years+Of+Blocking+Search+Engines&rft.dat
e=2014-06-30&rft.au=Barry+Schwartz&rft_id=http%3A%2F%2Fsearchengineland.com%
2Frobots-txt-celebrates-20-years-blocking-search-engines-195479&rfr_id=info%3Asi
d%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <h>>
          <a href="#cite ref-7">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="https://webmasters.googleblog.com/2019/07/</pre>
rep-id.html" rel="nofollow">
           "Formalizing the Robots Exclusion Protocol Specification"
          </a>
          <i>>
           Official Google Webmaster Central Blog
          </i>
          <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            2019-07-10
           </span>
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Official+Google+Web
master+Central+Blog&rft.atitle=Formalizing+the+Robots+Exclusion+Protocol+Specifi
cation&rft id=https%3A%2F%2Fwebmasters.googleblog.com%2F2019%2F07%2Frep-id.html&
amp;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       id="cite note-8">
        <span class="mw-cite-backlink">
         <b>
          <a href="#cite ref-8">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1" id="CITEREFM._Koster,_Stalworthy_Computing,</pre>
```

```
_Ltd.,_G._Illyes,_H._Zeller,_L._Harvey,_Google2019">
           M. Koster, Stalworthy Computing, Ltd., G. Illyes, H. Zeller, L. Harvey, G
oogle (2019-07-01).
           <a class="external text" href="https://tools.ietf.org/html/draft-rep-wg-t</pre>
opic-00" rel="nofollow">
            "Robots Exclusion Protocol draft-rep-wg-topic-00"
           </a>
           <i>>
            IETF Documents
           </i>
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             2019-09-09
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=IETF+Documents&
rft.atitle=Robots+Exclusion+Protocol+draft-rep-wg-topic-00&rft.date=2019-07-01&a
mp;rft.au=M.+Koster%2C+Stalworthy+Computing%2C+Ltd.%2C+G.+Illyes%2C+H.+Zeller%2C+L.+
Harvey%2C+Google&rft_id=https%3A%2F%2Ftools.ietf.org%2Fhtml%2Fdraft-rep-wg-topic
-00&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
          <span class="cs1-maint citation-comment">
           CS1 maint: multiple names: authors list (
           <a href="/wiki/Category:CS1_maint:_multiple_names:_authors_list" title="C</pre>
ategory:CS1 maint: multiple names: authors list">
            link
           </a>
           )
          </span>
         </span>
        <span class="mw-cite-backlink">
           <a href="#cite_ref-9">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://www.youtube.com/watch?v=KBdEwpRQRD</pre>
0#t=196s" rel="nofollow">
            "Uncrawled URLs in search results"
           </a>
           . YouTube. Oct 5, 2009.
           <a class="external text" href="https://web.archive.org/web/2014010622250</pre>
0/http://www.youtube.com/watch?v=KBdEwpRQRD0#t=196s" rel="nofollow">
            Archived
           </a>
           from the original on 2014-01-06
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             2013-12-29
            </span>
           </span>
```

```
</cite>
         <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Uncrawled+URLs+in+sear
ch+results&rft.pub=YouTube&rft.date=2009-10-05&rft_id=https%3A%2F%2Fwww.
youtube.com%2Fwatch%3Fv%3DKBdEwpRQRD0%23t%3D196s&rfr id=info%3Asid%2Fen.wikipedi
a.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <h>>
          <a href="#cite_ref-ask-webmasters_10-0">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="http://about.ask.com/docs/about/webmaster</pre>
s.shtml" rel="nofollow">
           "About Ask.com: Webmasters"
          </a>
          <i>>
           About.ask.com
          </i>
          <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=About.ask.com&r
ft.atitle=About+Ask.com%3A+Webmasters&rft_id=http%3A%2F%2Fabout.ask.com%2Fdocs%2
Fabout%2Fwebmasters.shtml&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusio
n+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <b>
          <a href="#cite ref-about-aol-search 11-0">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="http://search.aol.com/aol/about" rel="nofo</pre>
11ow">
           "About AOL Search"
          </a>
          <i>>
           Search.aol.com
```

```
</i>
          <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
         </cite>
         <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Search.aol.com&
rft.atitle=About+AOL+Search&rft_id=http%3A%2F%2Fsearch.aol.com%2Faol%2Fabout&am
p;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
          <a href="#cite ref-baidu-spider 12-0">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="http://www.baidu.com/search/spider_englis</pre>
h.html" rel="nofollow">
           "Baiduspider"
          </a>
          <i>>
           Baidu.com
          </i>
          <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Baidu.com&rft.a
title=Baiduspider&rft id=http%3A%2F%2Fwww.baidu.com%2Fsearch%2Fspider english.ht
ml&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <b>
          <a href="#cite ref-duckduckgo-bot 13-0">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
```

```
<a class="external text" href="https://duckduckgo.com/duckduckbot" rel="n</pre>
ofollow">
            "DuckDuckGo Bot"
           </a>
           <i>>
           DuckDuckGo.com
           </i>
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            25 April
           </span>
           2017
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=DuckDuckGo.com&
rft.atitle=DuckDuckGo+Bot&rft_id=https%3A%2F%2Fduckduckgo.com%2Fduckduckbot&
rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
        </span>
        <span class="mw-cite-backlink">
          <a href="#cite ref-google-webmasters-spec 14-0">
          <sup>
            <i>>
             <b>
             а
             </b>
            </i>
          </sup>
          </a>
          <a href="#cite_ref-google-webmasters-spec_14-1">
          <sup>
            <i>>
             <b>
             h
             </b>
            </i>
          </sup>
          </a>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://developers.google.com/webmasters/c</pre>
ontrol-crawl-index/docs/robots_txt" rel="nofollow">
            "Webmasters: Robots.txt Specifications"
           </a>
           <i>>
           Google Developers
          </i>
           <a class="external text" href="https://web.archive.org/web/2013011521413</pre>
7/https://developers.google.com/webmasters/control-crawl-index/docs/robots txt" rel
="nofollow">
           Archived
           </a>
          from the original on 2013-01-15
```

```
<span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Google+Developers&a
mp;rft.atitle=Webmasters%3A+Robots.txt+Specifications&rft_id=https%3A%2F%2Fdevel
opers.google.com%2Fwebmasters%2Fcontrol-crawl-index%2Fdocs%2Frobots_txt&rfr_id=i
nfo%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <h>>
          <a href="#cite ref-yahoo-search-is-bing 15-0">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="http://help.yahoo.com/kb/index?page=conten</pre>
t&y=PROD_SRCH&locale=en_US&id=SLN2217&impressions=true" rel="nofollo
w">
           "Submitting your website to Yahoo! Search"
          </a>
          <a class="external text" href="https://web.archive.org/web/2013012103580</pre>
1/http://help.yahoo.com/kb/index?page=content&y=PROD_SRCH&locale=en_US&i
d=SLN2217&impressions=true" rel="nofollow">
           Archived
          </a>
          from the original on 2013-01-21
          <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
         </cite>
         <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Submitting+your+websit
e+to+Yahoo%21+Search&rft id=http%3A%2F%2Fhelp.yahoo.com%2Fkb%2Findex%3Fpage%3Dco
ntent%26y%3DPROD_SRCH%26locale%3Den_US%26id%3DSLN2217%26impressions%3Dtrue&rfr_i
d=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
       <span class="mw-cite-backlink">
         <a href="#cite_ref-yandex-robots_16-0">
          <sup>
           <i>>
            <b>
```

```
а
             </b>
            </i>
          </sup>
          </a>
          <a href="#cite ref-yandex-robots 16-1">
          <sup>
            <i>>
             <b>
             b
             </b>
            </i>
          </sup>
          </a>
         </span>
        <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="http://help.yandex.com/webmaster/?id=11138</pre>
51" rel="nofollow">
            "Using robots.txt"
          </a>
           <i>>
           Help.yandex.com
          </i>
          <a class="external text" href="https://web.archive.org/web/2013012504001</pre>
7/http://help.yandex.com/webmaster/?id=1113851" rel="nofollow">
           Archived
           </a>
          from the original on 2013-01-25
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            16 February
           </span>
           2013
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Help.yandex.com&am
p;rft.atitle=Using+robots.txt&rft_id=http%3A%2F%2Fhelp.yandex.com%2Fwebmaster%2
F%3Fid%3D1113851&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
          </span>
        </span>
        <span class="mw-cite-backlink">
          <a href="#cite_ref-bing-blog-robots_17-0">
          <sup>
            <i>>
             <b>
              а
             </b>
            </i>
          </sup>
          </a>
          <a href="#cite_ref-bing-blog-robots_17-1">
           <sup>
            <i>>
```

```
<b>
              b
             </b>
            </i>
           </sup>
          </a>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://blogs.bing.com/webmaster/2008/06/0</pre>
3/robots-exclusion-protocol-joining-together-to-provide-better-documentation/" rel
="nofollow">
            "Robots Exclusion Protocol: joining together to provide better documenta
tion"
           </a>
           <i>>
            Blogs.bing.com
           </i>
           <a class="external text" href="https://web.archive.org/web/2014081802541</pre>
2/http://blogs.bing.com/webmaster/2008/06/03/robots-exclusion-protocol-joining-toget
her-to-provide-better-documentation/" rel="nofollow">
            Archived
           </a>
           from the original on 2014-08-18
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             16 February
            </span>
            2013
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Blogs.bing.com&
rft.atitle=Robots+Exclusion+Protocol%3A+joining+together+to+provide+better+documenta
tion&rft_id=https%3A%2F%2Fblogs.bing.com%2Fwebmaster%2F2008%2F06%2F03%2Frobots-e
xclusion-protocol-joining-together-to-provide-better-documentation%2F&rfr id=inf
o%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite note-18">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite ref-18">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://www.bing.com/webmaster/help/how-to</pre>
-create-a-robots-txt-file-cb7c31ec" rel="nofollow">
            "How to Create a Robots.txt File - Bing Webmaster Tools"
           </a>
           <i>>
            www.bing.com
```

```
</i>
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2019-02-06
           </span>
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=www.bing.com&rf
t.atitle=How+to+Create+a+Robots.txt+File+-+Bing+Webmaster+Tools&rft_id=https%3A%
2F%2Fwww.bing.com%2Fwebmaster%2Fhelp%2Fhow-to-create-a-robots-txt-file-cb7c31ec&
rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
        <span class="mw-cite-backlink">
         <h>>
          <a href="#cite ref-19">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1" id="CITEREFJason_Scott">
          <a href="/wiki/Jason_Scott" title="Jason Scott">
           Jason Scott
          </a>
          <a class="external text" href="http://www.archiveteam.org/index.php?title</pre>
=Robots.txt" rel="nofollow">
            "Robots.txt is a suicide note"
          </a>
           . Archive Team.
          <a class="external text" href="https://web.archive.org/web/2017021804452</pre>
7/http://www.archiveteam.org/index.php?title=Robots.txt" rel="nofollow">
           Archived
           </a>
          from the original on 2017-02-18
           <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            18 February
           </span>
           2017
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Robots.txt+is+a+suicid
e+note&rft.pub=Archive+Team&rft.au=Jason+Scott&rft_id=http%3A%2F%2Fwww.a
rchiveteam.org%2Findex.php%3Ftitle%3DRobots.txt&rfr id=info%3Asid%2Fen.wikipedi
a.org%3ARobots+exclusion+standard">
         </span>
        </span>
        id="cite_note-20">
         <span class="mw-cite-backlink">
         <b>
           <a href="#cite ref-20">
```

```
</a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="https://blog.archive.org/2017/04/17/robots</pre>
-txt-meant-for-search-engines-dont-work-well-for-web-archives/" rel="nofollow">
            "Robots.txt meant for search engines don't work well for web archives |
Internet Archive Blogs"
          </a>
           <i>>
           blog.archive.org
          </i>
           . 17 April 2017.
           <a class="external text" href="https://web.archive.org/web/2018120413002</pre>
8/http://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-w
ell-for-web-archives/" rel="nofollow">
           Archived
           </a>
           from the original on 2018-12-04
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2018-12-01
           </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=blog.archive.org&am
p;rft.atitle=Robots.txt+meant+for+search+engines+don%27t+work+well+for+web+archives
+%7C+Internet+Archive+Blogs&rft.date=2017-04-17&rft_id=https%3A%2F%2Fblog.ar
chive.org%2F2017%2F04%2F17%2Frobots-txt-meant-for-search-engines-dont-work-well-for-
web-archives%2F&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
          </span>
         </span>
        <span class="mw-cite-backlink">
          <a href="#cite_ref-Internet_Archive_21-0">
          <sup>
            <i>>
             <h>>
             а
             </b>
            </i>
           </sup>
          </a>
          <a href="#cite_ref-Internet_Archive_21-1">
           <sup>
            <i>>
             <b>
             h
             </b>
            </i>
          </sup>
          </a>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
```

```
-style"/>
          <cite class="citation news cs1" id="CITEREFJones2017">
           Jones, Brad (24 April 2017).
           <a class="external text" href="https://www.digitaltrends.com/computing/in</pre>
ternet-archive-robots-txt/#ixzz4gQYOqpUi" rel="nofollow">
            "The Internet Archive Will Ignore Robots.txt Files to Maintain Accuracy"
           </a>
           <i>>
            <a href="/wiki/Digital_Trends" title="Digital Trends">
            Digital Trends
            </a>
           </i>
           <a class="external text" href="https://web.archive.org/web/2017051613002</pre>
9/https://www.digitaltrends.com/computing/internet-archive-robots-txt/#ixzz4gQYOqpU
i" rel="nofollow">
           Archived
           </a>
           from the original on 2017-05-16
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            8 May
            </span>
            2017
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.jtitle=Digital+Trends&
rft.atitle=The+Internet+Archive+Will+Ignore+Robots.txt+Files+to+Maintain+Accuracy&am
p;rft.date=2017-04-24&rft.aulast=Jones&rft.aufirst=Brad&rft_id=https%3A%
2F%2Fwww.digitaltrends.com%2Fcomputing%2Finternet-archive-robots-txt%2F%23ixzz4gQY0q
pUi&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite note-22">
         <span class="mw-cite-backlink">
          <h>>
           <a href="#cite_ref-22">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://support.google.com/webmasters/answ</pre>
er/6062608" rel="nofollow">
            "Block URLs with robots.txt: Learn about robots.txt files"
           </a>
           <a class="external text" href="https://web.archive.org/web/2015081401340</pre>
0/https://support.google.com/webmasters/answer/6062608" rel="nofollow">
           Archived
           </a>
           from the original on 2015-08-14
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             2015-08-10
            </span>
```

```
</span>
         </cite>
         <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Block+URLs+with+robot
s.txt%3A+Learn+about+robots.txt+files&rft id=https%3A%2F%2Fsupport.google.com%2F
webmasters%2Fanswer%2F6062608&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+excl
usion+standard">
         </span>
        </span>
        <span class="mw-cite-backlink">
          <a href="#cite_ref-23">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
         <cite class="citation web cs1">
          <a class="external text" href="https://www.theregister.co.uk/2015/05/19/r</pre>
obotstxt/" rel="nofollow">
           "Robots.txt tells hackers the places you don't want them to look"
          </a>
          <i>>
           The Register
          </i>
          <a class="external text" href="https://web.archive.org/web/2015082106375</pre>
9/http://www.theregister.co.uk/2015/05/19/robotstxt/" rel="nofollow">
           Archived
           </a>
          from the original on 2015-08-21
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            August 12,
           </span>
           2015
          </span>
         </cite>
         <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=The+Register&rf
t.atitle=Robots.txt+tells+hackers+the+places+you+don%27t+want+them+to+look&rft i
d=https%3A%2F%2Fwww.theregister.co.uk%2F2015%2F05%2F19%2Frobotstxt%2F&rfr id=inf
o%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
         </span>
        </span>
        <span class="mw-cite-backlink">
          <a href="#cite ref-24">
          </a>
         </b>
        </span>
        <span class="reference-text">
         <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
```

```
<cite class="citation journal cs1" id="CITEREFScarfoneJansenTracy2008">
           Scarfone, K. A.; Jansen, W.; Tracy, M. (July 2008).
           <a class="external text" href="http://csrc.nist.gov/publications/nistpub</pre>
s/800-123/SP800-123.pdf" rel="nofollow">
            "Guide to General Server Security"
           <span class="cs1-format">
            (PDF)
           </span>
           . National Institute of Standards and Technology.
           <a class="mw-redirect" href="/wiki/Doi_(identifier)" title="Doi (identifi</pre>
er)">
           doi
           </a>
           <a class="external text" href="https://doi.org/10.6028%2FNIST.SP.800-123"</pre>
rel="nofollow">
            10.6028/NIST.SP.800-123
           </a>
           <a class="external text" href="https://web.archive.org/web/2011100811541</p>
2/http://csrc.nist.gov/publications/nistpubs/800-123/SP800-123.pdf" rel="nofollow">
           Archived
           </a>
           <span class="cs1-format">
            (PDF)
           </span>
           from the original on 2011-10-08
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            August 12,
            </span>
            2015
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=Guide+to+General+Se
rver+Security&rft.date=2008-07&rft_id=info%3Adoi%2F10.6028%2FNIST.SP.800-123
&rft.aulast=Scarfone&rft.aufirst=K.+A.&rft.au=Jansen%2C+W.&rft.au=Tr
acy%2C+M.&rft id=http%3A%2F%2Fcsrc.nist.gov%2Fpublications%2Fnistpubs%2F800-123%
2FSP800-123.pdf&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
          </span>
          <span class="cs1-hidden-error error citation-comment">
           Cite journal requires
           <code class="cs1-code">
            |iournal=
           </code>
           (
           <a href="/wiki/Help:CS1_errors#missing_periodical" title="Help:CS1 error</pre>
s">
           help
           </a>
          </span>
         </span>
        id="cite note-25">
         <span class="mw-cite-backlink">
          <h>>
           <a href="#cite_ref-25">
           </a>
```

```
</b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation book cs1" id="CITEREFSverre H. Huseby2004">
           Sverre H. Huseby (2004).
           <a class="external text" href="https://books.google.com/books?id=RjVjgPQs</pre>
KogC&pg=PA92" rel="nofollow">
            <i>>
            Innocent Code: A Security Wake-Up Call for Web Programmers
            </i>
           </a>
           . John Wiley & Sons. pp. 91-92.
           <a class="mw-redirect" href="/wiki/ISBN_(identifier)" title="ISBN (identi</pre>
fier)">
            ISBN
           </a>
           <a href="/wiki/Special:BookSources/9780470857472" title="Special:BookSour
ces/9780470857472">
            <hdi>>
            9780470857472
            </bdi>
           </a>
           <a class="external text" href="https://web.archive.org/web/2016040119343</pre>
7/https://books.google.com/books?id=RjVjgPQsKogC&pg=PA92" rel="nofollow">
           Archived
           </a>
           from the original on 2016-04-01
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2015-08-12
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=book&rft.btitle=Innocent+Code%3A+A+Securi
ty+Wake-Up+Call+for+Web+Programmers&rft.pages=91-92&rft.pub=John+Wiley+%26+S
ons&rft.date=2004&rft.isbn=9780470857472&rft.au=Sverre+H.+Huseby&rft
_id=https%3A%2F%2Fbooks.google.com%2Fbooks%3Fid%3DRjVjgPQsKogC%26pg%3DPA92&rfr_i
d=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite note-26">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite_ref-26">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="http://www.user-agents.org/" rel="nofollo"</pre>
w">
            "List of User-Agents (Spiders, Robots, Browser)"
           </a>
           . User-agents.org.
           <a class="external text" href="https://web.archive.org/web/2014010715420</pre>
```

```
5/http://user-agents.org/" rel="nofollow">
           Archived
           </a>
           from the original on 2014-01-07
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            2013-12-29
            </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=List+of+User-Agents+%2
8Spiders%2C+Robots%2C+Browser%29&rft.pub=User-agents.org&rft_id=http%3A%2F%2
Fwww.user-agents.org%2F&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+
standard">
          </span>
         </span>
        id="cite note-27">
         <span class="mw-cite-backlink">
          <h>>
          <a href="#cite_ref-27">
          </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="https://httpd.apache.org/docs/2.2/howto/ac</pre>
cess.html" rel="nofollow">
           "Access Control - Apache HTTP Server"
          </a>
           . Httpd.apache.org.
           <a class="external text" href="https://web.archive.org/web/2013122911083</pre>
1/http://httpd.apache.org/docs/2.2/howto/access.html" rel="nofollow">
           Archived
           </a>
           from the original on 2013-12-29
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2013-12-29
           </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Access+Control+-+Apach
e+HTTP+Server&rft.pub=Httpd.apache.org&rft id=https%3A%2F%2Fhttpd.apache.or
g%2Fdocs%2F2.2%2Fhowto%2Faccess.html&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobo
ts+exclusion+standard">
         </span>
         </span>
        id="cite note-28">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite_ref-28">
          </a>
          </b>
```

```
</span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://www.iis.net/configreference/syste</pre>
m.webserver/security/requestfiltering/filteringrules/filteringrule/denystrings" rel
="nofollow">
            "Deny Strings for Filtering Rules : The Official Microsoft IIS Site"
          </a>
           . Iis.net. 2013-11-06.
           <a class="external text" href="https://web.archive.org/web/2014010111273</pre>
0/http://www.iis.net/configreference/system.webserver/security/requestfiltering/filt
eringrules/filteringrule/denystrings" rel="nofollow">
           Archived
           </a>
           from the original on 2014-01-01
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            2013-12-29
           </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Deny+Strings+for+Filte
ring+Rules+%3A+The+Official+Microsoft+IIS+Site&rft.pub=Iis.net&rft.date=2013
-11-06&rft_id=http%3A%2F%2Fwww.iis.net%2Fconfigreference%2Fsystem.webserver%2Fse
curity%2Frequestfiltering%2Ffilteringrules%2Ffilteringrule%2Fdenystrings&rfr_id=
info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
        </span>
        <span class="mw-cite-backlink">
           <a href="#cite ref-29">
          </a>
          </h>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="https://www.google.com/humans.txt" rel="no</pre>
follow">
            "Google humans.txt"
           <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            October 3,
            </span>
            2019
          </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Google+humans.txt&
rft_id=https%3A%2F%2Fwww.google.com%2Fhumans.txt&rfr_id=info%3Asid%2Fen.wikipedi
a.org%3ARobots+exclusion+standard">
          </span>
         </span>
```

```
id="cite note-30">
         <span class="mw-cite-backlink">
          <b>
           <a href="#cite ref-30">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://github.com/humans.txt" rel="nofoll</pre>
ow">
            "Github humans.txt"
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            October 3,
            </span>
            2019
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Github+humans.txt&
rft_id=https%3A%2F%2Fgithub.com%2Fhumans.txt&rfr_id=info%3Asid%2Fen.wikipedia.or
g%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite_note-31">
         <span class="mw-cite-backlink">
          <h>>
           <a href="#cite ref-31">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1" id="CITEREFNewman2014">
           Newman, Lily Hay (2014-07-03).
           <a class="external text" href="https://slate.com/technology/2014/07/a-kil</pre>
ler-robots-txt-google-easter-egg.html" rel="nofollow">
            "Is This a Google Easter Egg or Proof That Skynet Is Actually Plotting W
orld Domination?"
           </a>
           <i>>
           Slate Magazine
           </i>
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2019-10-03
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Slate+Magazine&
```

```
rft.atitle=Is+This+a+Google+Easter+Egg+or+Proof+That+Skynet+Is+Actually+Plotting+Wor
ld+Domination%3F&rft.date=2014-07-03&rft.aulast=Newman&rft.aufirst=Lily+
Hay&rft id=https%3A%2F%2Fslate.com%2Ftechnology%2F2014%2F07%2Fa-killer-robots-tx
t-google-easter-egg.html&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion
+standard">
          </span>
        </span>
        id="cite note-32">
         <span class="mw-cite-backlink">
          <a href="#cite_ref-32">
          </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="https://www.google.com/killer-robots.txt"</pre>
rel="nofollow">
           "/killer-robots.txt"
          </a>
           . 2018-01-10.
          <a class="external text" href="https://web.archive.org/web/2018011016091</pre>
6/https://www.google.com/killer-robots.txt" rel="nofollow">
           Archived
          </a>
          from the original on 2018-01-10
           <span class="reference-accessdate">
           . Retrieved
           <span class="nowrap">
            2018-05-25
           </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=%2Fkiller-robots.txt&a
mp;rft.date=2018-01-10&rft_id=https%3A%2F%2Fwww.google.com%2Fkiller-robots.txt&a
mp;rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
        </span>
        <span class="mw-cite-backlink">
          <h>>
           <a href="#cite ref-bing-crawl-delay 33-0">
          </a>
          </b>
         </span>
        <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="https://blogs.bing.com/webmaster/2012/05/0</pre>
3/to-crawl-or-not-to-crawl-that-is-bingbots-question/" rel="nofollow">
           "To crawl or not to crawl, that is BingBot's question"
          </a>
           . 3 May 2012.
           <a class="external text" href="https://web.archive.org/web/2016020314282</p>
2/https://blogs.bing.com/webmaster/2012/05/03/to-crawl-or-not-to-crawl-that-is-bingb
```

ots-question/" rel="nofollow">

```
Archived
           </a>
           from the original on 2016-02-03
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            9 February
            </span>
            2016
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=To+crawl+or+not+to+cra
wl%2C+that+is+BingBot%27s+question&rft.date=2012-05-03&rft_id=https%3A%2F%2F
blogs.bing.com%2Fwebmaster%2F2012%2F05%2F03%2Fto-crawl-or-not-to-crawl-that-is-bingb
ots-question%2F&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
          </span>
         </span>
        id="cite_note-34">
         <span class="mw-cite-backlink">
           <a href="#cite_ref-34">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://support.google.com/webmasters/answ</pre>
er/48620?hl=en" rel="nofollow">
            "Change Googlebot crawl rate - Search Console Help"
           </a>
           <i>>
            support.google.com
           </i>
           <a class="external text" href="https://web.archive.org/web/2018111820574</pre>
7/https://support.google.com/webmasters/answer/48620?hl=en" rel="nofollow">
           Archived
           </a>
           from the original on 2018-11-18
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            22 October
            </span>
            2018
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=support.google.com&
amp;rft.atitle=Change+Googlebot+crawl+rate+-+Search+Console+Help&rft id=https%3
A%2F%2Fsupport.google.com%2Fwebmasters%2Fanswer%2F48620%3Fh1%3Den&rfr id=info%3A
sid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite_note-35">
```

```
<span class="mw-cite-backlink">
          <b>
           <a href="#cite ref-35">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://www.google.com/support/webmasters/</pre>
bin/answer.py?hl=en&answer=156449&from=40364" rel="nofollow">
            "Webmaster Help Center - How do I block Googlebot?"
           </a>
           <a class="external text" href="https://web.archive.org/web/2010080108200</pre>
8/http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=156449&am
p;from=40364" rel="nofollow">
            Archived
           </a>
           from the original on 2010-08-01
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2007-11-20
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Webmaster+Help+Center+
-+How+do+I+block+Googlebot%3F&rft_id=https%3A%2F%2Fwww.google.com%2Fsupport%2Fwe
bmasters%2Fbin%2Fanswer.py%3Fhl%3Den%26answer%3D156449%26from%3D40364&rfr_id=inf
o%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite note-36">
         <span class="mw-cite-backlink">
          <h>>
           <a href="#cite_ref-36">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="http://help.yahoo.com/l/us/yahoo/search/we</pre>
bcrawler/slurp-02.html" rel="nofollow">
            "How do I prevent my site or certain subdirectories from being crawled?
 - Yahoo Search Help"
           </a>
           <a class="external text" href="https://web.archive.org/web/2007101301421</pre>
5/http://help.yahoo.com/l/us/yahoo/search/webcrawler/slurp-02.html" rel="nofollow">
           Archived
           </a>
           from the original on 2007-10-13
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
             2007-11-20
```

```
</span>
           </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=How+do+I+prevent+my+si
te+or+certain+subdirectories+from+being+crawled%3F+-+Yahoo+Search+Help&rft id=ht
tp%3A%2F%2Fhelp.yahoo.com%2F1%2Fus%2Fyahoo%2Fsearch%2Fwebcrawler%2Fslurp-02.html&am
p;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
         </span>
        id="cite_note-37">
         <span class="mw-cite-backlink">
           <a href="#cite ref-37">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="http://blog.semetrical.com/googles-secret-</pre>
approach-to-robots-txt/" rel="nofollow">
            "Google's Hidden Interpretation of Robots.txt"
           </a>
           <a class="external text" href="https://web.archive.org/web/2010112002361</pre>
9/http://blog.semetrical.com/googles-secret-approach-to-robots-txt/" rel="nofollow">
           Archived
           </a>
           from the original on 2010-11-20
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            2010-11-15
            </span>
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Google%27s+Hidden+Inte
rpretation+of+Robots.txt&rft_id=http%3A%2F%2Fblog.semetrical.com%2Fgoogles-secre
t-approach-to-robots-txt%2F&rfr id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclus
ion+standard">
          </span>
         </span>
        id="cite note-38">
         <span class="mw-cite-backlink">
           <a href="#cite_ref-38">
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://web.archive.org/web/2009030506184</pre>
1/http://ysearchblog.com/2007/04/11/webmasters-can-now-auto-discover-with-sitemaps/
rel="nofollow">
```

```
"Yahoo! Search Blog - Webmasters can now auto-discover with Sitemaps"
           </a>
           . Archived from
           <a class="external text" href="http://ysearchblog.com/2007/04/11/webmaste</pre>
rs-can-now-auto-discover-with-sitemaps/" rel="nofollow">
           the original
           </a>
          on 2009-03-05
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2009-03-23
            </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx ver=Z39.88-2004&amp;rft val fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Yahoo%21+Search+Blog+-
+Webmasters+can+now+auto-discover+with+Sitemaps&rft_id=http%3A%2F%2Fysearchblog.
com%2F2007%2F04%2F11%2Fwebmasters-can-now-auto-discover-with-sitemaps%2F&rfr_id=
info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
        </span>
        id="cite note-39">
        <span class="mw-cite-backlink">
          <a href="#cite ref-39">
          </a>
          </b>
         </span>
        <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
          <a class="external text" href="http://help.yandex.com/webmaster/?id=11138</pre>
51" rel="nofollow">
           "Yandex - Using robots.txt"
          </a>
           <a class="external text" href="https://web.archive.org/web/2013050923054</pre>
8/http://help.yandex.com/webmaster/?id=1113851" rel="nofollow">
           Archived
           </a>
          from the original on 2013-05-09
           <span class="reference-accessdate">
            . Retrieved
            <span class="nowrap">
            2013-05-13
            </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Yandex+-+Using+robots.
txt&rft id=http%3A%2F%2Fhelp.yandex.com%2Fwebmaster%2F%3Fid%3D1113851&rfr id
=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
          </span>
        </span>
        <span class="mw-cite-backlink">
          <a href="#cite ref-40">
```

```
Λ
           </a>
          </b>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://developers.google.com/search/refer</pre>
ence/robots txt?hl=en" rel="nofollow">
           "Robots.txt Specifications"
           </a>
           <i>>
           Google Developers
           </i>
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            February 15,
           </span>
           2020
           </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Google+Developers&a
mp;rft.atitle=Robots.txt+Specifications&rft id=https%3A%2F%2Fdevelopers.google.c
om%2Fsearch%2Freference%2Frobots_txt%3Fh1%3Den&rfr_id=info%3Asid%2Fen.wikipedia.
org%3ARobots+exclusion+standard">
          </span>
         </span>
        <span class="mw-cite-backlink">
          <a href="#cite_ref-google-meta_41-0">
            <i>>
             <b>
             а
             </b>
            </i>
           </sup>
          <a href="#cite ref-google-meta 41-1">
           <sup>
            <i>>
             <b>
             b
             </b>
            </i>
           </sup>
          </a>
         </span>
         <span class="reference-text">
          <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline</pre>
-style"/>
          <cite class="citation web cs1">
           <a class="external text" href="https://developers.google.com/webmasters/c</pre>
ontrol-crawl-index/docs/robots_meta_tag" rel="nofollow">
            "Robots meta tag and X-Robots-Tag HTTP header specifications - Webmaster
s - Google Developers"
           </a>
```

```
<a class="external text" href="https://web.archive.org/web/2013080802094</pre>
6/https://developers.google.com/webmasters/control-crawl-index/docs/robots_meta_tag"
rel="nofollow">
           Archived
           </a>
          from the original on 2013-08-08
           <span class="reference-accessdate">
            . Retrieved
           <span class="nowrap">
            2013-08-17
           </span>
          </span>
          </cite>
          <span class="Z3988" title="ctx_ver=Z39.88-2004&amp;rft_val_fmt=info%3Aofi%</pre>
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Robots+meta+tag+and+X-
Robots-Tag+HTTP+header+specifications+-+Webmasters+%E2%80%94+Google+Developers&r
ft_id=https%3A%2F%2Fdevelopers.google.com%2Fwebmasters%2Fcontrol-crawl-index%2Fdocs%
2Frobots_meta_tag&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standa
rd">
         </span>
        </span>
       </div>
      <h2>
       <span class="mw-headline" id="External_links">
       External links
       </span>
       <span class="mw-editsection">
        <span class="mw-editsection-bracket">
        [
        </span>
        <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=edit&amp;se</pre>
ction=15" title="Edit section: External links">
         edit
        </a>
        <span class="mw-editsection-bracket">
         1
       </span>
       </span>
      </h2>
      <l
       <1i>>
        <span class="official-website">
         <span class="url">
          <a class="external text" href="http://www.robotstxt.org" rel="nofollow">
          Official website
          </a>
         </span>
        </span>
       <
        <a class="external text" href="http://www.robotstxt.org/db.html" rel="nofoll</pre>
ow">
         Robots Database (list of bot names)
       </a>
       <
        <a class="external text" href="https://sitechecker.pro/robots-tester/" rel</pre>
="nofollow">
         Robots.txt checker tool
        </a>
       <
```

```
<a class="external text" href="https://www.bing.com/webmaster/help/how-to-cr</pre>
eate-a-robots-txt-file-cb7c31ec" rel="nofollow">
       Microsoft bingbot official documentation
       </a>
      <div aria-label="Portals" class="noprint metadata navbox" role="navigation" st</pre>
yle="font-weight:bold;padding:0.4em 2em">
      <span style="display:inline-block;white-space:nowrap">
         <span style="margin:0 0.5em">
          <a class="image" href="/wiki/File:Crystal_Clear_app_linneighborhood.svg">
          <img alt="" data-file-height="407" data-file-width="407" decoding="asyn</pre>
c" height="21" src="//upload.wikimedia.org/wikipedia/commons/thumb/f/f9/Crystal_Clea
r_app_linneighborhood.svg/21px-Crystal_Clear_app_linneighborhood.svg.png" srcset="//
upload.wikimedia.org/wikipedia/commons/thumb/f/f9/Crystal Clear app linneighborhood.
svg/32px-Crystal_Clear_app_linneighborhood.svg.png 1.5x, //upload.wikimedia.org/wiki
pedia/commons/thumb/f/f9/Crystal_Clear_app_linneighborhood.svg/42px-Crystal_Clear_ap
p_linneighborhood.svg.png 2x" width="21"/>
         </a>
         </span>
         <a href="/wiki/Portal:Internet" title="Portal:Internet">
         Internet portal
         </a>
        </span>
       </div>
     <div aria-labelledby="Authority_control_frameless_&amp;#124;text-top_&amp;#12</pre>
4;10px_|alt=Edit_this_at_Wikidata_|link=https://www.wikida
ta.org/wiki/Q80776#identifiers|Edit_this_at_Wikidata" class="navbox authori
ty-control" role="navigation" style="padding:3px">
      <table class="nowraplinks hlist navbox-inner" style="border-spacing:0;backgro
und:transparent;color:inherit">
       p |10px |alt=Edit this at Wikidata |link=https://
www.wikidata.org/wiki/Q80776#identifiers|Edit_this_at_Wikidata" scope="row"
style="width:1%">
          <a href="/wiki/Help:Authority_control" title="Help:Authority control">
          Authority control
          </a>
          <a href="https://www.wikidata.org/wiki/Q80776#identifiers" title="Edit th</pre>
          <img alt="Edit this at Wikidata" data-file-height="20" data-file-width</pre>
="20" decoding="async" height="10" src="//upload.wikimedia.org/wikipedia/en/thumb/8/
8a/OOjs UI icon edit-ltr-progressive.svg/10px-OOjs UI icon edit-ltr-progressive.svg.
png" srcset="//upload.wikimedia.org/wikipedia/en/thumb/8/8a/00js_UI_icon_edit-ltr-pr
ogressive.svg/15px-00js_UI_icon_edit-ltr-progressive.svg.png 1.5x, //upload.wikimedi
a.org/wikipedia/en/thumb/8/8a/00js_UI_icon_edit-ltr-progressive.svg/20px-00js_UI_ico
n_edit-ltr-progressive.svg.png 2x" style="vertical-align: text-top" width="10"/>
         </a>
         h:2px;border-left-style:solid;width:100%;padding:0px">
         <div style="padding:0em 0.25em">
          <l
           <
            <span class="uid">
             <a class="external text" href="https://academic.microsoft.com/v2/deta</pre>
il/163978704" rel="nofollow">
              Microsoft Academic
             </a>
```

```
</span>
              </div>
           </div>
      <!--
NewPP limit report
Parsed by mw1395
Cached time: 20210919053121
Cache expiry: 1814400
Reduced expiry: false
Complications: [vary-revision-sha1]
CPU time usage: 0.604 seconds
Real time usage: 0.727 seconds
Preprocessor visited node count: 2467/1000000
Post-expand include size: 78042/2097152 bytes
Template argument size: 1921/2097152 bytes
Highest expansion depth: 12/40
Expensive parser function count: 7/500
Unstrip recursion depth: 1/20
Unstrip post-expand size: 135669/5000000 bytes
Lua time usage: 0.351/10.000 seconds
Lua memory usage: 6446910/52428800 bytes
Number of Wikibase entities loaded: 1/400
-->
      <!--
Transclusion expansion time report (%,ms,calls,template)
100.00% 661.793 1 -total
45.61% 301.824 1 Template:Reflist
33.71% 223.116 38 Template:Cite_web
12.94% 85.627 1 Template:Short_description
10.33% 68.389 1 Template:Pp-pc1
7.63% 50.497 1 Template:Pagetype
7.26% 48.071 1 Template:When
6.34% 41.950 1 Template:Fix
5.07% 33.554 1 Template:Authority_control
  3.66% 24.227
                       2 Template:Category_handler
      <!-- Saved in parser cache with key enwiki:stable-pcache:idhash:101673-0!canon
ical and timestamp 20210919053121 and revision id 1043553668. Serialized with JSON.
 -->
     </div>
      <noscript>
      <img alt="" height="1" src="//en.wikipedia.org/wiki/Special:CentralAutoLogin/s</pre>
tart?type=1x1" style="border: none; position: absolute;" title="" width="1"/>
      </noscript>
     <div class="printfooter">
      Retrieved from "
      <a dir="ltr" href="https://en.wikipedia.org/w/index.php?title=Robots_exclusion</pre>
_standard&oldid=1043553668">
       https://en.wikipedia.org/w/index.php?title=Robots_exclusion_standard&oldi
d=1043553668
      </a>
     </div>
    </div>
    <div class="catlinks" data-mw="interface" id="catlinks">
     <div class="mw-normal-catlinks" id="mw-normal-catlinks">
      <a href="/wiki/Help:Category" title="Help:Category">
       Categories
      </a>
```

```
<l
      <1i>>
       <a href="/wiki/Category:World_Wide_Web" title="Category:World Wide Web">
        World Wide Web
       </a>
       <1i>>
       <a href="/wiki/Category:Internet_bots" title="Category:Internet bots">
        Internet bots
       </a>
      </div>
     <div class="mw-hidden-catlinks mw-hidden-cats-hidden" id="mw-hidden-catlinks">
     Hidden categories:
     <l
      <
        <a href="/wiki/Category:CS1_maint:_multiple_names:_authors_list" title="Cate</pre>
gory:CS1 maint: multiple names: authors list">
        CS1 maint: multiple names: authors list
       </a>
      <
        <a href="/wiki/Category:CS1_errors:_missing_periodical" title="Category:CS1</pre>
errors: missing periodical">
        CS1 errors: missing periodical
       </a>
       <1i>>
        <a href="/wiki/Category:Articles_with_short_description" title="Category:Art</pre>
icles with short description">
        Articles with short description
       </a>
       <
        <a href="/wiki/Category:Short_description_matches_Wikidata" title="Category:</pre>
Short description matches Wikidata">
        Short description matches Wikidata
       </a>
       <
       <a href="/wiki/Category:Wikipedia_pending_changes_protected_pages" title="Ca</pre>
tegory: Wikipedia pending changes protected pages">
        Wikipedia pending changes protected pages
       </a>
      <1i>>
        <a href="/wiki/Category:All articles with vague or ambiguous time" title="Ca
tegory:All articles with vague or ambiguous time">
        All articles with vague or ambiguous time
       </a>
       <
        <a href="/wiki/Category:Vague_or_ambiguous_time_from_December_2020" title="C</pre>
ategory: Vague or ambiguous time from December 2020">
        Vague or ambiguous time from December 2020
       </a>
       <
        <a href="/wiki/Category:Official website different in Wikidata and Wikipedi
a" title="Category:Official website different in Wikidata and Wikipedia">
        Official website different in Wikidata and Wikipedia
        </a>
```

```
<
       <a href="/wiki/Category:Articles_with_MA_identifiers" title="Category:Articl</pre>
es with MA identifiers">
       Articles with MA identifiers
      </a>
      </div>
   </div>
  </div>
 </div>
 <div id="mw-data-after-content">
  <div class="read-more-container">
 </div>
 <div id="mw-navigation">
  <h2>
   Navigation menu
  </h2>
  <div id="mw-head">
   <nav aria-labelledby="p-personal-label" class="mw-portlet mw-portlet-personal ve
ctor-user-menu-legacy vector-menu" id="p-personal" role="navigation">
    <h3 class="vector-menu-heading" id="p-personal-label">
     <span>
     Personal tools
    </span>
    </h3>
    <div class="vector-menu-content">
     Not logged in
      <a accesskey="n" href="/wiki/Special:MyTalk" title="Discussion about edits f</pre>
rom this IP address [n]">
       Talk
      </a>
      <a accesskey="y" href="/wiki/Special:MyContributions" title="A list of edits</pre>
made from this IP address [y]">
       Contributions
      </a>
      <a href="/w/index.php?title=Special:CreateAccount&amp;returnto=Robots+exclus">clus
ion+standard" title="You are encouraged to create an account and log in; however, it
is not mandatory">
       Create account
      </a>
      <a accesskey="o" href="/w/index.php?title=Special:UserLogin&amp;returnto=Rob</pre>
ots+exclusion+standard" title="You're encouraged to log in; however, it's not mandat
ory. [o]">
       Log in
      </a>
      </div>
   </nav>
   <div id="left-navigation">
    <nav aria-labelledby="p-namespaces-label" class="mw-portlet mw-portlet-namespac</pre>
es vector-menu vector-menu-tabs" id="p-namespaces" role="navigation">
     <h3 class="vector-menu-heading" id="p-namespaces-label">
```

```
<span>
      Namespaces
     </span>
     </h3>
     <div class="vector-menu-content">
     <a accesskey="c" href="/wiki/Robots_exclusion_standard" title="View the con</pre>
tent page [c]">
        Article
       </a>
      <a accesskey="t" href="/wiki/Talk:Robots_exclusion_standard" rel="discussio"</pre>
n" title="Discuss improvements to the content page [t]">
        Talk
       </a>
      </div>
    </nav>
    <nav aria-labelledby="p-variants-label" class="mw-portlet mw-portlet-variants e</pre>
mptyPortlet vector-menu-dropdown-noicon vector-menu vector-menu-dropdown" id="p-vari
ants" role="navigation">
     <input aria-haspopup="true" aria-labelledby="p-variants-label" class="vector-m</pre>
enu-checkbox" data-event-name="ui.dropdown-p-variants" id="p-variants-checkbox" role
="button" type="checkbox"/>
     <h3 class="vector-menu-heading" id="p-variants-label">
     <span>
      Variants
     </span>
     <span class="vector-menu-checkbox-expanded">
      expanded
     </span>
     <span class="vector-menu-checkbox-collapsed">
      collapsed
     </span>
     </h3>
     <div class="vector-menu-content">
     </div>
    </nav>
   </div>
   <div id="right-navigation">
    <nav aria-labelledby="p-views-label" class="mw-portlet mw-portlet-views vector-</pre>
menu vector-menu-tabs" id="p-views" role="navigation">
     <h3 class="vector-menu-heading" id="p-views-label">
     <span>
      Views
     </span>
     </h3>
     <div class="vector-menu-content">
     <a href="/wiki/Robots exclusion standard">
        Read
       </a>
      <a accesskey="e" href="/w/index.php?title=Robots exclusion standard&amp;act</pre>
ion=edit" title="Edit this page [e]">
        Edit
       </a>
```

```
<a accesskey="h" href="/w/index.php?title=Robots_exclusion_standard&amp;act</pre>
ion=history" title="Past revisions of this page [h]">
         View history
        </a>
        </div>
    </nav>
     <nav aria-labelledby="p-cactions-label" class="mw-portlet mw-portlet-cactions e</pre>
mptyPortlet vector-menu-dropdown-noicon vector-menu vector-menu-dropdown" id="p-cact
ions" role="navigation" title="More options">
      <input aria-haspopup="true" aria-labelledby="p-cactions-label" class="vector-m</pre>
enu-checkbox" data-event-name="ui.dropdown-p-cactions" id="p-cactions-checkbox" role
="button" type="checkbox"/>
     <h3 class="vector-menu-heading" id="p-cactions-label">
      <span>
       More
      </span>
      <span class="vector-menu-checkbox-expanded">
       expanded
      </span>
      <span class="vector-menu-checkbox-collapsed">
       collapsed
      </span>
     </h3>
     <div class="vector-menu-content">
      </div>
     </nav>
     <div class="vector-search-box" id="p-search" role="search">
     <div>
      <h3>
       <label for="searchInput">
        Search
       </label>
      </h3>
       <form action="/w/index.php" id="searchform">
        <div data-search-loc="header-navigation" id="simpleSearch">
        <input accesskey="f" autocapitalize="sentences" id="searchInput" name="sear</pre>
ch" placeholder="Search Wikipedia" title="Search Wikipedia [f]" type="search"/>
        <input name="title" type="hidden" value="Special:Search"/>
        <input class="searchButton mw-fallbackSearchButton" id="mw-searchButton" na</pre>
me="fulltext" title="Search Wikipedia for this text" type="submit" value="Search"/>
        <input class="searchButton" id="searchButton" name="go" title="Go to a page</pre>
with this exact name if it exists" type="submit" value="Go"/>
       </div>
      </form>
     </div>
    </div>
    </div>
   </div>
   <div id="mw-panel">
    <div id="p-logo" role="banner">
    <a class="mw-wiki-logo" href="/wiki/Main Page" title="Visit the main page">
    </a>
    </div>
    <nav aria-labelledby="p-navigation-label" class="mw-portlet mw-portlet-navigatio")</pre>
n vector-menu vector-menu-portal portal id="p-navigation" role="navigation">
    <h3 class="vector-menu-heading" id="p-navigation-label">
     <span>
      Navigation
     </span>
    </h3>
```

```
<div class="vector-menu-content">
    <a accesskey="z" href="/wiki/Main_Page" title="Visit the main page [z]">
      Main page
      </a>
     <a href="/wiki/Wikipedia:Contents" title="Guides to browsing Wikipedia">
      Contents
     </a>
     <a href="/wiki/Portal:Current_events" title="Articles related to current eve</pre>
nts">
      Current events
     </a>
     <a accesskey="x" href="/wiki/Special:Random" title="Visit a randomly selecte</pre>
d article [x]">
      Random article
      </a>
     <a href="/wiki/Wikipedia:About" title="Learn about Wikipedia and how it work
s">
      About Wikipedia
     </a>
     <a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us" title="How to contact</pre>
Wikipedia">
      Contact us
      </a>
     <a href="https://donate.wikimedia.org/wiki/Special:FundraiserRedirector?utm_</pre>
source=donate&utm medium=sidebar&utm campaign=C13 en.wikipedia.org&usela
ng=en" title="Support us by donating to the Wikimedia Foundation">
      Donate
      </a>
     </div>
   <nav aria-labelledby="p-interaction-label" class="mw-portlet mw-portlet-interact</pre>
ion vector-menu vector-menu-portal portal" id="p-interaction" role="navigation">
   <h3 class="vector-menu-heading" id="p-interaction-label">
    <span>
     Contribute
    </span>
   </h3>
   <div class="vector-menu-content">
    <a href="/wiki/Help:Contents" title="Guidance on how to use and edit Wikiped
ia">
      Help
      </a>
     <a href="/wiki/Help:Introduction" title="Learn how to edit Wikipedia">
      Learn to edit
      </a>
```

```
<a href="/wiki/Wikipedia:Community_portal" title="The hub for editors">
      Community portal
      </a>
     <a accesskey="r" href="/wiki/Special:RecentChanges" title="A list of recent</pre>
changes to Wikipedia [r]">
      Recent changes
      </a>
     <a href="/wiki/Wikipedia:File_Upload_Wizard" title="Add images or other medi</pre>
a for use on Wikipedia">
      Upload file
      </a>
     </div>
   </nav>
   <nav aria-labelledby="p-tb-label" class="mw-portlet mw-portlet-tb vector-menu ve</pre>
ctor-menu-portal portal" id="p-tb" role="navigation">
   <h3 class="vector-menu-heading" id="p-tb-label">
    <span>
     Tools
    </span>
   </h3>
   <div class="vector-menu-content">
    <a accesskey="j" href="/wiki/Special:WhatLinksHere/Robots exclusion standar</pre>
d" title="List of all English Wikipedia pages containing links to this page [j]">
      What links here
      </a>
     <a accesskey="k" href="/wiki/Special:RecentChangesLinked/Robots_exclusion_st</pre>
andard" rel="nofollow" title="Recent changes in pages linked from this page [k]">
      Related changes
      </a>
     <a accesskey="u" href="/wiki/Wikipedia:File_Upload_Wizard" title="Upload fil</pre>
es [u]">
      Upload file
      </a>
     <a accesskey="q" href="/wiki/Special:SpecialPages" title="A list of all spec</pre>
ial pages [q]">
      Special pages
      </a>
     <a href="/w/index.php?title=Robots exclusion standard&amp;oldid=1043553668"</pre>
title="Permanent link to this revision of this page">
      Permanent link
      </a>
     <a href="/w/index.php?title=Robots_exclusion_standard&amp;action=info" title</pre>
="More information about this page">
       Page information
      </a>
```

```
<a href="/w/index.php?title=Special:CiteThisPage&amp;page=Robots exclusion s</pre>
tandard&id=1043553668&wpFormIdentifier=titleform" title="Information on how
to cite this page">
       Cite this page
      </a>
     <a accesskey="g" href="https://www.wikidata.org/wiki/Special:EntityPage/0807</pre>
76" title="Structured data on this page hosted by Wikidata [g]">
       Wikidata item
      </a>
     </div>
   <nav aria-labelledby="p-coll-print export-label" class="mw-portlet mw-portlet-co</pre>
ll-print_export vector-menu vector-menu-portal portal" id="p-coll-print_export" role
="navigation">
    <h3 class="vector-menu-heading" id="p-coll-print_export-label">
     <span>
     Print/export
    </span>
    </h3>
    <div class="vector-menu-content">
     <a href="/w/index.php?title=Special:DownloadAsPdf&amp;page=Robots exclusion</pre>
standard&action=show-download-screen" title="Download this page as a PDF file">
       Download as PDF
      </a>
     <a accesskey="p" href="/w/index.php?title=Robots_exclusion_standard&amp;prin</pre>
table=yes" title="Printable version of this page [p]">
       Printable version
      </a>
     </div>
   </nav>
   <nav aria-labelledby="p-lang-label" class="mw-portlet mw-portlet-lang vector-men</pre>
u vector-menu-portal portal id="p-lang" role="navigation">
    <h3 class="vector-menu-heading" id="p-lang-label">
     <span>
     Languages
    </span>
    </h3>
    <div class="vector-menu-content">
     <a class="interlanguage-link-target" href="https://ar.wikipedia.org/wiki/%D</pre>
9%85%D8%B9%D9%8A%D8%A7%D8%B1 %D8%A7%D8%B3%D8%AA%D8%A8%D8%B9%D8%A7%D8%AF %D8%A7%D9%8
معيار استبعاد "=4%D8%B1%D9%88%D8%A8%D9%88%D8%AA%D8%A7%D8%AA" hreflang="ar" lang="ar" title
Arabic">
       العربية
      </a>
     <a class="interlanguage-link-target" href="https://bar.wikipedia.org/wiki/Ro</pre>
bots_Exclusion_Standard" hreflang="bar" lang="bar" title="Robots Exclusion Standard
- Bavarian">
       Boarisch
      </a>
```

```
<a class="interlanguage-link-target" href="https://ca.wikipedia.org/wiki/Pro</pre>
tocol_d%27exclusi%C3%B3_de_robots" hreflang="ca" lang="ca" title="Protocol d'exclusi
ó de robots - Catalan">
      Català
      </a>
     <a class="interlanguage-link-target" href="https://cs.wikipedia.org/wiki/Pro</pre>
tokol_pro_zak%C3%A1z%C3%A1n%C3%AD_p%C5%99%C3%ADstupu_robot%C5%AFm" hreflang="cs" lan
g="cs" title="Protokol pro zakázání přístupu robotům - Czech">
      Čeština
      </a>
     <a class="interlanguage-link-target" href="https://da.wikipedia.org/wiki/Rob</pre>
ot_Exclusion_Standard" hreflang="da" lang="da" title="Robot Exclusion Standard - Dan
ish">
      Dansk
      </a>
     <a class="interlanguage-link-target" href="https://de.wikipedia.org/wiki/Rob</pre>
ots_Exclusion_Standard" hreflang="de" lang="de" title="Robots Exclusion Standard - G
erman">
      Deutsch
      </a>
     <a class="interlanguage-link-target" href="https://es.wikipedia.org/wiki/Es</pre>
t%C3%A1ndar_de_exclusi%C3%B3n_de_robots" hreflang="es" lang="es" title="Estándar de
exclusión de robots - Spanish">
      Español
      </a>
     <a class="interlanguage-link-target" href="https://fa.wikipedia.org/wiki/%D</pre>
8%A7%D8%B3%D8%AA%D8%A7%D9%86%D8%AF%D8%A7%D8%B1%D8%AF %D8%A7%D8%B3%D8%AA%D8%AB%D9%86%
D8%A7%D8%A1 %DA%A9%D8%B1%D8%AF%D9%86 %D8%B1%D8%A8%D8%A7%D8%AA%E2%80%8C%D9%87%D8%A7"
hreflang="fa" lang="fa" title="كردن رباتها" - Persian">
      فارسى
      </a>
     <a class="interlanguage-link-target" href="https://fr.wikipedia.org/wiki/Pro</pre>
tocole d%27exclusion des robots" hreflang="fr" lang="fr" title="Protocole d'exclusio
n des robots - French">
      Français
      </a>
     <a class="interlanguage-link-target" href="https://ko.wikipedia.org/wiki/%E</pre>
B%A1%9C%EB%B4%87_%EB%B0%B0%EC%A0%9C_%ED%91%9C%EC%A4%80" hreflang="ko" lang="ko" titl
e="로봇 배제 표준 - Korean">
      한국어
      </a>
     <a class="interlanguage-link-target" href="https://id.wikipedia.org/wiki/Rob</pre>
ots.txt" hreflang="id" lang="id" title="Robots.txt - Indonesian">
      Bahasa Indonesia
      </a>
```

```
<a class="interlanguage-link-target" href="https://it.wikipedia.org/wiki/Pro</pre>
tocollo di esclusione robot" hreflang="it" lang="it" title="Protocollo di esclusione
robot - Italian">
       Italiano
      </a>
     <a class="interlanguage-link-target" href="https://he.wikipedia.org/wiki/%D</pre>
7%A4%D7%A8%D7%95%D7%98%D7%95%D7%A7%D7%95%D7%9C_%D7%90%D7%99_%D7%94%D7%9B%D7%9C%D7%9
C%D7%AA %D7%A8%D7%95%D7%91%D7%95%D7%98%D7%99" hreflang="he" lang="he" title=""םר
וטוקול אי הכללת רובוטים - Hebrew">
       עברית
      </a>
     <a class="interlanguage-link-target" href="https://nl.wikipedia.org/wiki/Rob</pre>
ots Exclusion Protocol" hreflang="nl" lang="nl" title="Robots Exclusion Protocol - D
utch">
       Nederlands
      </a>
     <a class="interlanguage-link-target" href="https://ja.wikipedia.org/wiki/Rob</pre>
ots_Exclusion_Standard" hreflang="ja" lang="ja" title="Robots Exclusion Standard - J
apanese">
       日本語
      </a>
     <a class="interlanguage-link-target" href="https://pl.wikipedia.org/wiki/Rob</pre>
ots_Exclusion_Protocol" hreflang="pl" lang="pl" title="Robots Exclusion Protocol - P
olish">
       Polski
      </a>
     <a class="interlanguage-link-target" href="https://pt.wikipedia.org/wiki/Pro</pre>
tocolo_de_exclus%C3%A3o_de_rob%C3%B4s" hreflang="pt" lang="pt" title="Protocolo de e
xclusão de robôs - Portuguese">
       Português
      </a>
     <a class="interlanguage-link-target" href="https://ru.wikipedia.org/wiki/%D</pre>
0%A1%D1%82%D0%B0%D0%BD%D0%B4%D0%B0%D1%80%D1%82_%D0%B8%D1%81%D0%BA%D0%BB%D1%8E%D1%87%
D0%B5%D0%BD%D0%B8%D0%B9 %D0%B4%D0%BB%D1%8F %D1%80%D0%BE%D0%B1%D0%BE%D1%82%D0%BE%D0%B
2" hreflang="ru" lang="ru" title="Стандарт исключений для роботов - Russian">
       Русский
      </a>
     <a class="interlanguage-link-target" href="https://simple.wikipedia.org/wik</pre>
i/Robots_exclusion_standard" hreflang="en-simple" lang="en-simple" title="Robots exc
lusion standard - Simple English">
       Simple English
      </a>
     <a class="interlanguage-link-target" href="https://sr.wikipedia.org/wiki/Sta</pre>
ndard isklju%C4%8Denja robota" hreflang="sr" lang="sr" title="Standard isključenja r
obota - Serbian">
       Српски / srpski
      </a>
```

```
<a class="interlanguage-link-target" href="https://fi.wikipedia.org/wiki/Rob</pre>
otin_rajausstandardi" hreflang="fi" lang="fi" title="Robotin rajausstandardi - Finni
sh">
       Suomi
      </a>
      <a class="interlanguage-link-target" href="https://sv.wikipedia.org/wiki/Rob</pre>
ots_Exclusion_Standard" hreflang="sv" lang="sv" title="Robots Exclusion Standard - S
wedish">
       Svenska
      </a>
      <a class="interlanguage-link-target" href="https://tr.wikipedia.org/wiki/Rob</pre>
ot engelleme standard%C4%B1" hreflang="tr" lang="tr" title="Robot engelleme standard
1 - Turkish">
       Türkce
      </a>
      <a class="interlanguage-link-target" href="https://uk.wikipedia.org/wiki/Rob</pre>
ots.txt" hreflang="uk" lang="uk" title="Robots.txt - Ukrainian">
       Українська
      </a>
      <a class="interlanguage-link-target" href="https://zh.wikipedia.org/wiki/Rob</pre>
ots.txt" hreflang="zh" lang="zh" title="Robots.txt - Chinese">
       中文
      </a>
      <div class="after-portlet after-portlet-lang">
      <span class="wb-langlinks-edit wb-langlinks-link">
      <a class="wbc-editpage" href="https://www.wikidata.org/wiki/Special:EntityPa</pre>
ge/Q80776#sitelinks-wikipedia" title="Edit interlanguage links">
       Edit links
      </a>
      </span>
     </div>
    </div>
   </nav>
  </div>
 </div>
 <footer class="mw-footer" id="footer" role="contentinfo">
  This page was last edited on 10 September 2021, at 17:48
    <span class="anonymous-show">
     (UTC)
    </span>
   Text is available under the
    <a href="//en.wikipedia.org/wiki/Wikipedia:Text of Creative Commons Attribution
-ShareAlike_3.0_Unported_License" rel="license">
    Creative Commons Attribution-ShareAlike License
    <a href="//creativecommons.org/licenses/by-sa/3.0/" rel="license" style="displa</pre>
y:none;">
    </a>
additional terms may apply. By using this site, you agree to the
```

```
<a href="//foundation.wikimedia.org/wiki/Terms_of_Use">
    Terms of Use
    </a>
    and
    <a href="//foundation.wikimedia.org/wiki/Privacy policy">
    Privacy Policy
    </a>
    . Wikipedia® is a registered trademark of the
    <a href="//www.wikimediafoundation.org/">
    Wikimedia Foundation, Inc.
    </a>
    , a non-profit organization.
   d="footer-places">
   <a class="extiw" href="https://foundation.wikimedia.org/wiki/Privacy policy" ti</pre>
tle="wmf:Privacy policy">
    Privacy policy
    </a>
   <a href="/wiki/Wikipedia:About" title="Wikipedia:About">
    About Wikipedia
    </a>
   <a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General disclaime</pre>
    Disclaimers
    </a>
   <a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">
    Contact Wikipedia
    </a>
   <a class="noprint stopMobileRedirectToggle" href="//en.m.wikipedia.org/w/index.</pre>
php?title=Robots_exclusion_standard&mobileaction=toggle_view_mobile">
    Mobile view
    </a>
   <a href="https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_contribute">
    Developers
    </a>
   id="footer-places-statslink">
    <a href="https://stats.wikimedia.org/#/en.wikipedia.org">
    Statistics
    </a>
   <a href="https://foundation.wikimedia.org/wiki/Cookie_statement">
    Cookie statement
    </a>
   <a href="https://wikimediafoundation.org/">
    <img alt="Wikimedia Foundation" height="31" loading="lazy" src="/static/image</pre>
s/footer/wikimedia-button.png" srcset="/static/images/footer/wikimedia-button-1.5x.p
ng 1.5x, /static/images/footer/wikimedia-button-2x.png 2x" width="88"/>
```

```
<a href="https://www.mediawiki.org/">
              <img alt="Powered by MediaWiki" height="31" loading="lazy" src="/static/image</pre>
        s/footer/poweredby_mediawiki_88x31.png" srcset="/static/images/footer/poweredby_medi
        awiki_132x47.png 1.5x, /static/images/footer/poweredby_mediawiki_176x62.png 2x" widt
        h="88"/>
             </a>
            </footer>
          <script>
           (RLQ=window.RLQ||[]).push(function(){mw.config.set({"wgPageParseReport":{"limitre
        port":{"cputime":"0.604","walltime":"0.727","ppvisitednodes":{"value":2467,"limit":1
        000000}, "postexpandincludesize": {"value": 78042, "limit": 2097152}, "templateargumentsiz
        e":{"value":1921,"limit":2097152},"expansiondepth":{"value":12,"limit":40},"expensiv
        efunctioncount":{"value":7,"limit":500},"unstrip-depth":{"value":1,"limit":20},"unst
        rip-size":{"value":135669,"limit":5000000},"entityaccesscount":{"value":1,"limit":40
                                                  1 -total"," 45.61% 301.824
        0},"timingprofile":["100.00% 661.793
        te:Reflist"," 33.71% 223.116
                                       38 Template:Cite_web"," 12.94%
                                                                         85.627
                                                                                      1 Tem
        plate:Short_description"," 10.33% 68.389
                                                       1 Template:Pp-pc1"," 7.63%
                                                                                     50.497
        1 Template:Pagetype"," 7.26% 48.071
                                                   1 Template:When"," 6.34% 41.950
                                            1 Template:Authority_control"," 3.66%
         1 Template:Fix"," 5.07% 33.554
                 2 Template:Category_handler"]}, "scribunto":{"limitreport-timeusage":{"valu
        e":"0.351","limit":"10.000"},"limitreport-memusage":{"value":6446910,"limit":5242880
        0}},"cachereport":{"origin":"mw1395","timestamp":"20210919053121","ttl":1814400,"tra
        nsientcontent":false}});});
          </script>
          <script type="application/ld+json">
           {"@context":"https:\/\/schema.org","@type":"Article","name":"Robots exclusion sta
        ndard", "url": "https:\/\/en.wikipedia.org\/wiki\/Robots_exclusion_standard", "sameA
        s":"http:\/\/www.wikidata.org\/entity\/Q80776","mainEntity":"http:\/\/www.wikidata.o
        rg\/entity\/Q80776","author":{"@type":"Organization","name":"Contributors to Wikimed
        ia projects"},"publisher":{"@type":"Organization","name":"Wikimedia Foundation, In
        c.","logo":{"@type":"ImageObject","url":"https:\/\/www.wikimedia.org\/static\/images
        \/wmf-hor-googpub.png"}},"datePublished":"2002-10-09T14:11:28Z","headline":"standard
        used to advise web crawlers and scrapers not to index a web page or site"}
          </script>
          <script>
           (RLQ=window.RLQ||[]).push(function(){mw.config.set({"wgBackendResponseTime":16
        0,"wgHostname":"mw1436"});});
          </script>
         </body>
        </html>
In [5]:
        #open a url and save it offline
         #response = http.request('GET', url)
         import requests
         page=requests.get(url)
         fout = open("downloaded.html", 'w')
         fout.write(str(page.content, 'utf-8'))
        UnicodeEncodeError
                                                 Traceback (most recent call last)
        C:\Users\SURESH~1\AppData\Local\Temp/ipykernel 15948/3894465986.py in <module>
              4 page=requests.get(url)
              5 fout = open("downloaded.html", 'w')
        ---> 6 fout.write(str(page.content, 'utf-8'))
        ~\AppData\Local\Programs\Python\Python39\lib\encodings\cp1252.py in encode(self, inp
        ut, final)
             17 class IncrementalEncoder(codecs.IncrementalEncoder):
                    def encode(self, input, final=False):
```



```
---> 19
                        return codecs.charmap_encode(input,self.errors,encoding_table)[0]
             20
             21 class IncrementalDecoder(codecs.IncrementalDecoder):
        UnicodeEncodeError: 'charmap' codec can't encode character '\u2010' in position 9240
        3: character maps to <undefined>
In [6]:
         #to scrape a local file stored on the system
         file = open('downloaded.html').read()
         soup = bs(file)
         #'find' returns first object in html file matching the requirement
         #'findAll' returns all matching objects
         #returns the first <h2> tag available in the page
         link = soup.find("h2")
         print(link)
         #list of all the <a> tags
         links = soup.findAll("center")
         print(links)
         print("printing all the links in the page : \n")
         #print all the href links available in <a> tag description
         for a_tag in links:
             print(a_tag.get("href"))
        None
        printing all the links in the page :
In [7]:
        #To scrape a table in webpage
         from bs4 import BeautifulSoup as bs
         from urllib.request import urlopen
         url2 = "https://www.w3schools.com/html/html tables.asp"
         url = "http://olympus.realpython.org/profiles/dionysus"
         url1 = "https://pes.edu/"
         #url2 = "https://www.tutorialspoint.com"
         page = urlopen(url2)
         html = page.read().decode("utf-8")
         soup = bs(html, "html.parser")
         #print(soup.get text())
         #soup.find all("img")
         table = soup.find('table')
         print(table)
         #prints all the rows in the table
         table rows = table.findAll('tr')
         #print(table_rows)
         print('the table contents are : \n')
         #find table data between tr tags
         for idx, row in enumerate(table_rows):
             tds = row.findAll('td')
             #to extract the table headers
             if tds is None:
                 tds = row.findAll('th')
             row_data = [i.text for i in tds] #list the data in the all the cells in the ro
             row_data += [i.text for i in tds]
             print("row", idx, ":", row_data)
```

```
Company
       Contact
       Country
       Alfreds Futterkiste
       Maria Anders
       Germany
       Centro comercial Moctezuma
       Francisco Chang
       Mexico
       Ernst Handel
       Roland Mendel
       Austria
       Island Trading
       Helen Bennett
       UK
       Laughing Bacchus Winecellars
       Yoshi Tannamuri
       Canada
       Magazzini Alimentari Riuniti
       Giovanni Rovelli
       Italy
       the table contents are :
       row 0 : []
       row 1 : ['Alfreds Futterkiste', 'Maria Anders', 'Germany', 'Alfreds Futterkiste', 'M
       aria Anders', 'Germany']
       row 2 : ['Centro comercial Moctezuma', 'Francisco Chang', 'Mexico', 'Centro comercia
       1 Moctezuma', 'Francisco Chang', 'Mexico']
       row 3 : ['Ernst Handel', 'Roland Mendel', 'Austria', 'Ernst Handel', 'Roland Mende
       l', 'Austria']
       row 4 : ['Island Trading', 'Helen Bennett', 'UK', 'Island Trading', 'Helen Bennett',
       row 5 : ['Laughing Bacchus Winecellars', 'Yoshi Tannamuri', 'Canada', 'Laughing Bacc
       hus Winecellars', 'Yoshi Tannamuri', 'Canada']
       row 6 : ['Magazzini Alimentari Riuniti', 'Giovanni Rovelli', 'Italy', 'Magazzini Ali
       mentari Riuniti', 'Giovanni Rovelli', 'Italy']
In [8]:
       #To scrape iPhone price from Amazon
       from bs4 import BeautifulSoup as bs
       from urllib.request import urlopen
       url = "https://www.amazon.in/New-Apple-iPhone-11-128GB/dp/B08L8CXFZ1"
       page = urlopen(url)
       html = page.read().decode("utf-8")
       soup = bs(html, "html.parser")
       #print(soup.prettify())
       #print(soup)
       price = soup.find("span", attrs={"id": 'priceblock_ourprice'})
       #print(price)
```

	<pre>price_value = price.string price_string = price_value.strip().replace(',', '') print (price_string)</pre>
	<pre>rating = soup.find("i", attrs={'class': 'a-icon a-icon-star a-star-4-5'}).string.str print(rating)</pre>
	₹54900.00 4.6 out of 5 stars
In []:	
In []:	