

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"
url1 = "http://olympus.realpython.org/profiles/dionysus"
url11 = "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)
```

```
<table id="customers">
<tr>
<th>Company</th>
<th>Contact</th>
<th>Country</th>
</tr>
<tr>
<td>Alfreds Futterkiste</td>
<td>Maria Anders</td>
<td>Germany</td>
</tr>
<tr>
<td>Centro comercial Moctezuma</td>
<td>Francisco Chang</td>
<td>Mexico</td>
</tr>
<tr>
<td>Ernst Handel</td>
<td>Roland Mendel</td>
<td>Austria</td>
</tr>
<tr>
<td>Island Trading</td>
<td>Helen Bennett</td>
<td>UK</td>
</tr>
<tr>
<td>Laughing Bacchus Winecellars</td>
<td>Yoshi Tannamuri</td>
<td>Canada</td>
</tr>
<tr>
<td>Magazzini Alimentari Riuniti</td>
<td>Giovanni Rovelli</td>
<td>Italy</td>
</tr>
</table>
```

In [2]:

```
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
e896ffff","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","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","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
(){mw.loader.implement("user.options@1hzgi",function($,jQuery,require,module){/*@nom
in*/mw.user.tokens.set({"patrolToken":"+\\","watchToken":"+\\","csrfToken":"+
\\"});});</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
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
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-when-crossorigin"/>\n<meta name="referrer" content="origin-when-cross-origin"/>\n<meta name="format-detection" content="telephone=no"/>\n<meta property="og:title" content="Robots exclusion standard - Wikipedia"/>\n<meta property="og:type" content="website"/>\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_exclusion_standard"/>\n<link rel="alternate" type="application/x-wiki" title="Edit this page" href="/w/index.php?title=Robots_exclusion_standard&action=edit"/>\n<link 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/wikipedia.png"/>\n<link rel="shortcut icon" href="/static/favicon/wikipedia.ico"/>\n<link rel="search" type="application/opensearchdescription+xml" href="/w/opensearch_desc.php" title="Wikipedia (en)"/>\n<link rel="EditURI" type="application/rsd+xml" href="//en.wikipedia.org/w/api.php?action=rsd"/>\n<link rel="license" href="//creativecommons.org/licenses/by-sa/3.0"/>\n<link rel="canonical" href="https://en.wikipedia.org/wiki/Robots_exclusion_standard"/>\n<link rel="dns-prefetch" href="//meta.wikimedia.org"/>\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-Robots_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-base" class="noprint"></div>\n<div id="content" class="mw-body" role="main">\n\n<div id="siteNotice"><!-- CentralNotice --></div>\n<div class="mw-indicators">\n<div id="mw-indicator-pp-autoreview" class="mw-indicator"></div>\n<div id="firstHeading" class="firstHeading">Robots exclusion standard</h1>\n<div id="bodyContent" class="vector-body">\n<div id="siteSub" class="noprint">From Wikipedia, the free encyclopedia</div>\n<div id="contentSub"><div id="mw-fr-revisiontag" class="flaggedrevs_short flaggedrevs_stable_synced_plainlinks noprint"><div class="flaggedrevs_short_basic"></div>\n<div id="mw-fr-revisiondetails-wrapper" style="position: relative;"><div id="mw-fr-revisiondetails" class="flaggedrevs_short_details" style="display: none;">This is the latest accepted revision, reviewed on <i>10 September 2021</i>.</div>\n</div>\n</div>\n<div id="contentSub2"></div>\n<div id="jump-to-nav"></div>\n<div id="mw-head">Jump to navigation</div>\n<div id="mw-content-text" class="mw-body-content mw-content-ltr" lang="en" dir="ltr"><div class="mw-parser-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<p class="mw-empty-elt">\n<p>The robots exclusion standard, also known as the robots exclusion protocol or simply robots.txt, is a standard used by websites to communicate with web crawlers and other web robots. 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 search engines to categorize websites. Not all robots cooperate with the standard; email harvesters, spambots, <a href="/w

iki/Malware" title="Malware">malware and robots that scan for security vulnerabilities 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</p>\n<div id="toc" class="toc" role="navigation" aria-labelledby="mw-toc-heading"><input type="checkbox" role="button" id="toctogglecheckbox" class="toctogglecheckbox" style="display:none" /><div class="toctitle" lang="en" dir="ltr"><h2 id="mw-toc-heading">Contents</h2><label class="toctoggletitle" for="toctogglecheckbox"></label></div>\n\n<li class="toclevel-1 toctection-1">1 History\n<li class="toclevel-1 toctection-2">2 Standard\n<li class="toclevel-1 toctection-3">3 Security\n<li class="toclevel-1 toctection-4">4 Alternatives\n<li class="toclevel-1 toctection-5">5 Examples\n<li class="toclevel-1 toctection-6">6 Nonstandard extensions\n\n<li class="toclevel-2 toctection-7">6.1 Crawl-delay directive\n<li class="toclevel-2 toctection-8">6.2 Allow directive\n<li class="toclevel-2 toctection-9">6.3 Sitemap\n<li class="toclevel-2 toctection-10">6.4 Host\n<li class="toclevel-2 toctection-11">6.5 Universal " match\n\n<li class="toclevel-1 toctection-12">7 Meta tags and headers\n<li class="toclevel-1 toctection-13">8 See also\n<li class="toclevel-1 toctection-14">9 References\n<li class="toclevel-1 toctection-15">10 External links\n\n</div>\n<h2>History[edit]</h2>\n<p>The standard was proposed by Martijn Koster,^{[1]}^{[2]} when working for Nexor^{[3]} in February 1994^{[4]} 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.^{[5]}\n</p>\n<p>It quickly became a de facto standard that present and future web crawlers were expected to follow; most complied, including those operated by search engines such as WebCrawler, Lycos, and AltaVista.^{[6]}\n</p>\n<p>On July 1, 2019, Google announced the proposal of the Robots Exclusion Protocol as an official standard under Internet Engineering Task Force.^{[7]} The draft^{[8]} will now undergo the acceptance process.\n</p>\n<h2>Standard[edit]</h2>\n</div>\n</div>

ion_standard&action=edit§ion=2" title="Edit section: Standard">edit]</h2>\n<p>When a site owner wishes to give instructions to web robots they place a text file called <style data-mw-duplicate="TemplateStyles:r886049734">.mw-parser-output .monospace{font-family:monospace,monospace}</style>robots.txt in the root of the web site hierarchy (e.g. <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>https://www.example.com/robots.txt). This text file contains the instructions in a specific format (see examples below). Robots that <i>choose</i> to follow the instructions try to fetch this file 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.\n<p><p>A robots.txt file on a website will function as a request that specified 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 belief 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 only 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><p>A robots.txt file covers one origin. For websites with multiple subdomains, each subdomain must have its own robots.txt file. If <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>example.com had a robots.txt file but <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>a.example.com did not, the rules that would apply for <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>example.com would not apply to <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>a.example.com. In addition, each protocol and port needs its own robots.txt file; <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>http://example.com/robots.txt does not apply to pages under <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>http://example.com:8080/ or <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r886049734"/>https://example.com/.\n<p><p>Some major search engines following this standard include Ask,^{[10]} AOL,^{[11]} Baidu,^{[12]} DuckDuckGo,^{[13]} Google,^{[14]} Yahoo!,^{[15]} and Yandex.^{[16]} Bing^{[17]} is still^{[<i>when?</i>]} not fully compatible with the standard as it cannot inherit settings from the wildcard (<code>*</code>).^{[18]}\n<p><p>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 project leader Jason Scott, "unchecked, and left alone, the robots.txt file ensures no mirroring or reference for items that may have general use and meaning beyond the website's context."^{[19]}\n<p><p>For some years, the 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 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 purposes".^{[21]} This was in response to entire domains being tagged with robots.txt when the content became obsolete.^{[21]}</p><n><h2>Security[edit]</h2><n><p>Despite the use of the terms "allow" and "disallow", the protocol is purely advisory and relies on the compliance of the web robot.^{[22]} 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 is sometimes claimed to be a security risk,^{[23]} this sort of 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 should not depend on the secrecy of the implementation or its components."^{[24]} In the context of robots.txt files, security through obscurity is not recommended as a security technique.^{[25]}<n><h2>Alternatives[edit]</h2><n><p>Many robots also pass a special user-agent to the web server when fetching content.^{[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]}<n><p><p>Some sites, such as Google, host a <code>humans.txt</code> file that displays information meant for humans to read.^{[29]} Some sites such as GitHub redirect humans.txt to an about page.^{[30]}<n><p><p>Previously Google had a joke file hosted at <code>/killer-robots.txt</code> instructing the Terminator not to kill the company founders Larry Page and Sergey Brin.^{[31]}^{[32]}<n><h2>Examples[edit]</h2><n><p>This example tells all robots that they can visit all files because the wildcard <code>*</code> stands for all robots and the <code>Disallow</code> directive has no value, meaning no pages are disallowed.<n><p><pre>User-agent: *
Allow: /
</pre><n><p>The same result can be accomplished with an empty or missing robots.txt file.<n><p><p>This example tells all robots to stay out of a website:<n><p><pre>User-agent: *
Disallow: /
</pre><n><p>This example tells all robots not to enter three directories:<n><p><pre>User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /junk/
</pre><n><p>This example tells all robots to stay away from one specific file:<n><p><pre>User-agent: *
Disallow: /directory/file.html
</pre><n><p>All other files in the specified directory will be processed.<n><p><p>This example tells a specific robot to stay out of a website:<n><p><pre>User-agent: BadBot
replace 'BadBot' with the actual user-agent of the bot
Disallow: /
</pre><n><p>This example te

lts two specific robots not to enter one specific directory:\n</p>\n<pre>User-agent: BadBot # replace \'BadBot\' with the actual user-agent of the bot\nUser-agent: Googlebot\nDisallow: /private/\n</pre>\n<p>Example demonstrating how comments can be used:\n</p>\n<pre># Comments appear after the "#" symbol at the start of a line, or after a directive\nUser-agent: * # match all bots\nDisallow: / # keep them out\n</pre>\n<p>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 Google, support several user-agent strings that allow the operator to deny access to a subset of their services by using specific user-agent strings.^{[14]}\n</p><p>Example demonstrating multiple user-agents:\n</p>\n<pre>User-agent: googlebot # all Google services\nDisallow: /private/ # disallow this directory\nUser-agent: googlebot-news # only the news service\nDisallow: / # disallow everything\nUser-agent: * # any robot\nDisallow: /something/ # disallow this directory\n</pre>\n<h2>Nonstandard extensions[edit]</h2>\n<h3>Crawl-delay directive[edit]</h3>\n<p>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 dependent 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.^{[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 its search console for webmasters, to control the Googlebot's subsequent visits.^{[34]}\n</p>\n<pre>User-agent: bingbot\nAllow: /\nCrawl-delay: 10\n</pre>\n<h3>Allow directive[edit]</h3>\n<p>Some major crawlers support an <code>Allow</code> directive, which can counteract a following <code>Disallow</code> directive.^{[35]}^{[36]} This is useful when one tells robots to avoid an entire directory but still wants some HTML documents in that directory crawled and indexed. While by standard implementation the first matching robots.txt pattern always wins, Google's implementation differs in that Allow patterns with equal or more characters in the directive path win over a matching Disallow pattern.^{[37]} Bing uses either the <code>Allow</code> or <code>Disallow</code> directive, whichever is more specific, based on length, like Google.^{[17]}\n</p><p>In order to be compatible to all robots, if one wants to allow single files inside an otherwise disallowed directory, it is necessary to place the Allow directive(s) first, followed by the Disallow, for example:\n</p>\n<pre>Allow: /directory1/myfile.html\nDisallow: /directory1/\n</pre>\n<p>This example will Disallow anything in /directory1/ except /directory1/myfile.html, since the latter will match first. The order is only important to robots that follow the standard; in the case of the Google or Bing bots, the order is not important.\n</p>\n<h3>Sitemap[edit]</h3>\n<p>Some crawlers support a <code>Sitemap</code> directive, allowing multiple <a href="/wiki/Sitemaps" t

title="Sitemaps">Sitemaps in the same <samp>robots.txt</samp> in the form <code>Sitemap: <i>full-url</i></code>:^{[38]}\n</p>\n<pre>Sitemap: http://www.example.com/sitemap.xml</pre>\n<h3>Host[edit]</h3>\n<p>Some crawlers (Yandex) support a <code>Host</code> directive, allowing websites with multiple mirrors to specify their preferred domain:^{[39]}\n</p>\n<pre>Host: hosting.example.com</pre>\n<p>This is not supported by all crawlers.\n</p>\n<h3>Universal "*" match[edit]</h3>\n<p>The <i>Robot Exclusion Standard</i> does not mention the "*" character in the <code>Disallow:</code> statement.^{[40]}\n</p>\n<h2>Meta tags and headers[edit]</h2>\n<p>In addition to root-level robots.txt files, robots exclusion directives can be applied at a more granular level through the use of Robots meta tags and X-Robots-Tag HTTP headers. The robots meta tag cannot be used for non-HTML files such as images, text files, or PDF documents. On the other hand, the X-Robots-Tag can be added to non-HTML files by using .htaccess and httpd.conf files.^{[41]}\n</p>\n<dl><dt>A "noindex" meta tag</dt></dl>\n<div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr"><pre><metaname=""content=""></pre></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"><pre>X-Robots-Tag: noindex</pre></div>\n<p>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 loaded, whereas robots.txt is effective before the page is requested. Thus if a page is excluded by a robots.txt file, any robots meta tags or X-Robots-Tag headers are effectively ignored because the robot will not see them in the first place.^{[41]}\n</p>\n<h2>See also[edit]</h2>\n<style data-mw-deduplicate="TemplateStyles: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{column-rule:1px solid #aaa}.mw-parser-output .div-col-dl,.mw-parser-output .div-col-ol,.mw-parser-output .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;">\n<code>ads.txt</code>, a standard for listing authorized ad sellers\n<code>security.txt</code>, a file to describe the process for security researchers to follow in order to report security vulnerabilities\nAutomated Content Access Protocol \xe2\x80\x93 a failed proposal to extend robots.txt\nBotSeer \xe2\x80\x93 now inactive search engine for robots.txt files\nDistributed web crawling\nFocused crawler\n<a href="/wiki/Internet_Archive" ti

National Digital Library Program (NDLP)

National Digital Information Infrastructure and Preservation P
rogram (NDIIPP)

Nofollow

Perma.cc

Meta elements</
a> for search engines

Sitemaps</
a>

Spider trap

Web archiving

Web crawler

<a href="/w
iki/Noindex" title="Noindex">noindex

<n><div><n><h2>References

[<a href="/w/index.php?title=Robots_exclusion_standard&a
mp;action=edit&section=14" title="Edit section: References">edit]</h2><n><style data-mw-deduplicate="TemplateStyles: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;margin-top:0.3em}.mw-parser-output .reflist-columns-2{column-width:30e
m}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-
columns{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}.mw-
parser-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
le><div class="reflist reflist-columns references-column-width" style="column-width:
30em;">

<n><ol class="references"><n><li id="cite_note-1"><span class="mw-cite-backlin
k">^<style dat
a-mw-deduplicate="TemplateStyles:r999302996">.mw-parser-output cite.citation{font-st
yle:inherit}.mw-parser-output .citation q{quotes:"\"\"\"\"\"\"\"\"\"\"'\"'}.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/Loc
k-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-l
imited a,.mw-parser-output .citation .cs1-lock-registration a{background:linear-rad
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
ption a,.mw-parser-output .citation .cs1-lock-subscription a{background:linear-rad
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;font
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:inne
rit}

<cite class="citation web cs1"><a rel="nofollow" class="external text" h
ref="http://www.greenhills.co.uk/historical.html">"Historical".<i>Greenhills.c
o.uk</i>.<a rel="nofollow" class="external text" href="https://web.archive.org/web/
20170403152037/http://www.greenhills.co.uk/historical.html">Archived<p>The or
iginal<p>on 2017-04-03. Retrieved<span class="nowr
ap">2017-03-03.

</cite><span title="ctx_ver=Z39.88-2004&rft_val_fmt=
info%3Aofi%2Ffmt%3Akev%3Amtx%3AJournal&rft.genre=unknown&rft.jtitle=Greenu
hills.co.uk&rft.atitle=Historical&rft_id=http%3A%2F%2Fwww.greenhills.co.uk%2Fh
istorical.html&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard"
class="Z3988">

<n><li id="cite_note-2"><span class="mw-cite-backl
ink">^<link re
l="mw-duplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id

=CITEREFFFielding1994" class="citation web cs1">Fielding, Roy (1994). "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><li id="cite_note-0-3"> <link rel="mw-deduplicated-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><li id="cite_note-4"> <link rel="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 Wanderers". <i>www-talk mailing list</i>. Archived from the original (Hypermail> archived message) on October 29, 2013.</cite><li id="cite_note-5"> <link rel="mw-deduplicated-inline-style" href="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</cite><li id="cite_note-sear-Robo-6"> <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFBarry_Schwartz2014" class="citation web cs1">Barry Schwartz (30 June 2014). "Robots.txt Celebrates 20 Years Of Blocking Search Engines". <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 on 2015-09-07. Retrieved 2015-11-19.</cite>\n\n<li id="cite_note-7">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Formalizing the Robots Exclusion Protocol Specification". <i>Official Google Webmaster Central Blog</i>. Retrieved 2019-07-10.</cite>\n\n<li id="cite_note-8">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFM._Koster,_Stalworthy_Computing,_Ltd.,_G._Illyes,_H._Zeller,_L._Harvey,_Google2019" class="citation web cs1">M. Koster, Stalworthy Computing, Ltd., G. Illyes, H. Zeller, L. Harvey, Google (2019-07-01). "Robots Exclusion Protocol draft-rep-wg-topic-00". <i>IETF Documents</i>. Retrieved 2019-09-09.</cite>CS1 maint: multiple names: authors list (link)\n\n<li id="cite_note-9">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Uncrawled URLs in search results". YouTube. Oct 5, 2009. Archived from the original on 2014-01-06. Retrieved 2013-12-29.</cite>\n\n<li id="cite_note-ask-webmasters-10">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"About Ask.com: Webmasters". <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="citation web cs1">"About AOL Search". <i>Search.aol.com</i>. Retrieved 16 February 2013.</cite>\n

essdate">. Retrieved 16 February 2013.</cite>\n\n<li id="cite_note-baidu-spider-12">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Baiduspider". <i>Baidu.com</i>. Retrieved 16 February 2013.</cite>\n\n<li id="cite_note-duckduckgo-bot-13">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"DuckDuckGo Bot". <i>DuckDuckGo.com</i>. Retrieved 25 April 2017.</cite>\n\n<li id="cite_note-google-webmasters-spec-14">^{<i>}<i>a</i></sup> ^{<i>b</i>} <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Webmasters: Robots.txt Specifications". <i>Google Developers</i>. Archived from the original on 2013-01-15. Retrieved 16 February 2013.</cite>\n\n<li id="cite_note-yahoo-search-is-bing-15">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Submitting your website to Yahoo! Search". Archived from the original on 2013-01-21. Retrieved 16 February 2013.</cite>\n\n<li id="cite_note-yandex-robots-16">^{<i>}<i>a</i></sup> ^{<i>b</i>} <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Using robots.txt". <i>Help.yandex.com</i>. Archived from the original on 2013-01-25. Retrieved 16 February 2013.</cite><span title="ctx_ver=Z3

88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Val.yandex.com&rft.atitle=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" class="Z3988">\n\n<li id="cite_note-bing-blog-robots-17">^ ^{<i>a</i>} ^{<i>b</i>} <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots Exclusion Protocol: joining together to provide better documentation". <i>Blogs.bing.com</i>. Archived from the original on 2014-08-18. Retrieved 16 February 2013</cite>\n\n<li id="cite_note-18">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"How to Create a Robots.txt File - Bing Webmaster Tools". <i>www.bing.com</i>. Retrieved 2019-02-06</cite>\n\n<li id="cite_note-19">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFJason_Scott" class="citation web cs1">Jason Scott. "Robots.txt is a suicide note". Archive Team. Archived from the original on 2017-02-18. Retrieved 18 February 2017</cite>\n\n<li id="cite_note-20">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots.txt meant for search engines don'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<li id="cite_note-Internet_Archive-21">^ ^{<i>a</i>}

i></sup> ^{<i>b</i>}
 <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFJones2017" class="citation news cs1">Jones, Brad (24 April 2017). "The Internet Archive Will Ignore Robots.txt Files to Maintain Accuracy". <i>Digital Trends</i>. Archived from the original on 2017-05-16. Retrieved 8 May 2017.</cite>\n\n<li id="cite_note-22">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Block URLs with robots.txt: Learn about robots.txt files". Archived from the original on 2015-08-14. Retrieved 2015-08-10.</cite>\n\n<li id="cite_note-23">^ <link rel="mw-deduplicated-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\n<li id="cite_note-24">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFScarfoneJansenTracy2008" class="citation journal cs1">Scarfone, K. A.; Jansen, W.; Tracy, M. (July 2008). "Guide to General Server Security" (PDF). National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-123. Archived (PDF) from the original on 2011-10-08. Retrieved August 12, 2015.</cite> Cite journal requires <code class="cs1-code">|journal=</code> (help)\n\n<li id="cite_note-2

5">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFSverre_H._Huseby2004" class="citation book cs1">Sverre H. Huseby (2004). <i>Innocent Code: A Security Wake-Up Call for Web Programmers</i>. John Wiley & Sons. pp. 91\–\–92. ISBN <bdi>9780470857472</bdi>. Archived from the original on 2016-04-01. Retrieved 2015-08-12.</cite>\n\n<li id="cite_note-26">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">List of User-Agents (Spiders, Robots, Browser)". User-agents.org. Archived from the original on 2014-01-07. Retrieved 2013-12-29.</cite>\n\n<li id="cite_note-27">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">Access Control - Apache HTTP Server. Httpd.apache.org. Archived from the original on 2013-12-29. Retrieved 2013-12-29.</cite>\n\n<li id="cite_note-28">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Deny Strings for Filtering Rules ; The Official Microsoft IIS Site". Iis.net. 2013-11-06. Archived from the original on 2014-01-01. Retrieved 2013-12-29.</cite>\n\n<li id="cite_note-29">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Google humans.txt". Retrieved October 3, 2019.</cite><span title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&

p;rft.genre=unknown&rft.btitle=Google+humans.txt&rft_id=https%3A%2F%2Fwww.google.com%2Fhumans.txt&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard" class="Z3988">\n\n<li id="cite_note-30">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">Github humans.txt. Retrieved October 3, 2019</cite>\n\n<li id="cite_note-31">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite id="CITEREFNewman2014" class="citation web cs1">Newman, Lily Hay (2014-07-03). "Is This a Google Easter Egg or Proof That Skynet Is Actually Plotting World Domination?". <i>Slate Magazine</i>. Retrieved 2019-10-03</cite>\n\n<li id="cite_note-32">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"killer-robots.txt". 2018-01-10. Archived from the original on 2018-01-10. Retrieved 2018-05-25</cite>\n\n<li id="cite_note-bing-crawl-delay-33">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"To crawl or not to crawl, that is BingBot's question". 3 May 2012. Archived from the original on 2016-02-03. Retrieved 9 February 2016</cite>\n\n<li id="cite_note-34">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Change Googlebot crawl rate - Search Console Help". <i>support.google.com</i>. Archived from the original on 2018-11-18. Retrieved 22 October 2018</cite>\n

= "Z3988" >\n\n<li id="cite_note-35">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Webmaster Help Center - How do I block Googlebot?". Archived from the original on 2010-08-01. Retrieved 2007-11-20</cite>\n\n<li id="cite_note-36">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"How do I prevent my site or certain subdirectories from being crawled? - Yahoo Search Help". Archived from the original on 2007-10-13. Retrieved 2007-11-20</cite>\n\n<li id="cite_note-37">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Google's Hidden Interpretation of Robots.txt". Archived from the original on 2010-11-20. Retrieved 2010-11-15</cite>\n\n<li id="cite_note-38">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Yahoo! Search Blog - Webmasters can now auto-discover with Sitemaps". Archived from the original on 2009-03-05. Retrieved 2009-03-23</cite>\n\n<li id="cite_note-39">^ <link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Yandex - Using robots.txt". Archived from the original on 2013-05-09. Retrieved 2013-05-13</cite><span title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%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" class="Z3988"><li id="cite_note-40">^<link rel="mw-deduplicated-inline-style" href="mw-data:TemplateStyles:r999302996"/><cite class="citation web cs1">"Robots.txt Specifications". <i>Google Developers</i>. Retrieved February 15, 2020.</cite><li id="cite_note-google-meta-41">^ ^{<i>a</i>} ^{<i>b</i>}<link 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 x80;x94 Google Developers". Archived from the original on 2013-08-08. Retrieved 2013-08-17.</cite></div><h2>External links[edit]</h2>Official websiteRobots Database (list of bot names)Robots.txt checker toolMicrosoft Bingbot official documentation<div class="noprint metadata navbar" role="navigation" aria-label="Portals" style="font-weight:bold;padding:0.4em 2em"><ul style="margin:0.1em 0 0"><li style="display:inline">Internet portal</div><div role="navigation" class="navbox authority-control" aria-labelledby="Authority_control_frameless_&#124;text-top_&#124;10px_&#124;alt=Edit_this_at_Wikidata_&#124;link=https&#58;//www.wikidata.org/wiki/Q80776#identifiers&#124;Edit_this_at_Wikidata" style="padding:3px"><table class="nowraplist hlist navbar-inner" style="border-spacing:0;background:transparent;color:inherit"><tbody><tr><th id="Authority_control_frameless_&#124;text-top_&#124;10px_&#124;alt=Edit_this_at_Wikidata_&#124;link=https&#58;//www.wikidata.org/wiki/Q80776#identifiers&#124;Edit_this_at_Wikidata" scope="row" class="navbox-group" style="width:1%">Help:Authority control </th><td class="navbox-list navbox-odd" style="text-align:left;border-left-width:2px;border-left-style:solid;width:100%;padding:0px"><div style="padding:0em 0.25em">Microsoft Academic</div></td></tr></tbody></table></div><!-- \nNewPP limit report\nParsed by mw1395\nCached time: 20210919053121\nCache expiry: 1814400\nReduced expiry: false\nComplications: [vary\\x80\revision\\x80\sha1]\nCPU time usage: 0.604 seconds\nReal time usage: 0.727 seconds\nPreprocessor visited node count: 2467/1000000\nPost\\x80\ऎxpand 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\\x80\ऎxpand size: 135669/500000 bytes\nLua time usage: 0.351/10.000 seconds\nLua memory usage: 6446910/52428800 bytes\nNumber of Wikibase entities loaded: 1/400\n-->\n<!--\nTransclusion expansion time report (%,ms,calls,template)\n100.00% 661.793 1 -total\n45.61% 301.824 1 Template:Reflist\n33.71% 223.116 38 Template:Cite_web\n12.94% 85.627 1 Template:Short_description\n10.33% 68.389 1 Template:Pp-pc1\n7.63% 50.497 1 Template:Pagetype\n7.26% 48.071 1 Template:When\n6.34% 41.950 1 Template:Fix\n5.07% 33.554 1 Template:Authority_control\n3.66% 24.227 2 Template:Category_handler\n-->\n<!-- Save d in parser cache with key enwiki:stable-pcache:idhash:101673-0!canonical and timestamp 20210919053121 and revision id 1043553668. Serialized with JSON.\n-->\n</div><noscript></noscript><div class="printfooter">Retrieved from "https://en.wikipedia.org/w/index.php?title=Robots_exclusion_standard&oldid=1043553668</div></div>\n<div id="catlinks" class="catlinks" data-mw="interface"><div id="mw-normal-catlinks" class="mw-normal-catlinks">Categories: World Wide WebInternet bots</div><div id="mw-hidden-catlinks" class="mw-hidden-catlinks mw-hidden-cats-hidden">Hidden categories: CS1 maint: multiple names: authors listCS1 errors: missing periodicalArticles with short descriptionShort description matches WikidataWikipedia pending changes protected pagesAll articles with vague or ambiguous timeVague or ambiguous time from December 2020Official website different in Wikidata and WikipediaArticles with MA identifiers</div></div>\n<div id="\-content\-content\-read-more-container"></div>\n<div id="mw-navigation">\n<div id="mw-head">\n<div id="p-personal" class="mw-portlet mw-portlet-personal vector-user-menu-legacy vector-menu aria-label="p-personal-label" role="navigation" \n<div id="p-personal-label" class="vector-menu-heading">Personal tools\n<div class="vector-menu-content">\n<ul class="vector-menu-content-list"><li id="pt-anonuserpage" class="mw-list-item">Not logged in<li id="pt-anontalk" class="mw-list-item">Talk<li id="pt-anoncontribs" class="mw-list-item">Contributions<li id="pt-createaccount" class="mw-list-i

[illegible]

st-item">Random article<li id="n-aboutsite" class="mw-list-item">About Wikipedia<li id="n-contactpage" class="mw-list-item">Contact us<li id="n-sitesupport" class="mw-list-item">Donate</div></nav><nav id="p-interaction" class="mw-portlet mw-portlet-interaction vector-menu vector-menu-portal" aria-labelledby="p-interaction-label" role="navigation" ><h3 id="p-interaction-label" class="vector-menu-heading">Contribute</h3><div class="vector-menu-content"><ul class="vector-menu-content-list"><li id="n-help" class="mw-list-item">Help<li id="n-introduction" class="mw-list-item">Learn to edit<li id="n-portal" class="mw-list-item">Community portal<li id="n-recentchanges" class="mw-list-item">Recent changes<li id="n-upload" class="mw-list-item">Upload file</div></nav><nav id="p-tb" class="mw-portlet mw-portlet-tb vector-menu vector-menu-portal" aria-labelledby="p-tb-label" role="navigation" ><h3 id="p-tb-label" class="vector-menu-heading">Tools</h3><div class="vector-menu-content"><ul class="vector-menu-content-list"><li id="t-whatlinkshere" class="mw-list-item">What links here<li id="t-recentchangeslinked" class="mw-list-item">Related changes<li id="t-upload" class="mw-list-item">Upload file<li id="t-specialpages" class="mw-list-item">Special pages<li id="t-permalink" class="mw-list-item">Permanent link<li id="t-info" class="mw-list-item">Page information<li id="t-cite" class="mw-list-item">Cite this page<li id="t-wikibase" class="mw-list-item">Wikidata item</div></nav><nav id="p-coll-print-export" class="mw-portlet mw-portlet-coll-print-export vector-menu vector-menu-portal" aria-labelledby="p-coll-print-export-label" role="navigation" ><h3 id="p-coll-print-export-label" class="vector-menu-heading">Print/export</h3><div class="vector-menu-content"><ul class="vector-menu-content-list"><li id="coll-download-as-r1" class="mw-list-item">Download as PDF<li id="t-print" class="mw-list-item">Printable version</div></nav><nav id="p-lang" class="mw-portlet mw-portlet-lang vector-menu vector-menu-portal" aria-labelledby="p-lang-label" role="navigation" ><h3 id="p-lang-label" class="vector-menu-heading">Languages</h3><div class="vector-menu-content"><ul class="vector-menu-content-list"><li class="interlanguage-link interwiki-ar mw-list-item">\xd8\xa7\xd9\x84\xd8\xb9\xd8\xb1\xd8\xa\xd9\x8a\xd8\xa9</

a><li class="interlanguage-link interwiki-bar mw-list-item">Boarisch<li class="interlanguage-link interwiki-ca mw-list-item">Catal\xc3\xa0<li class="interlanguage-link interwiki-cs mw-list-item">\xc4\x8ce\xc5\xa1tina<li class="interlanguage-link interwiki-da mw-list-item">Dansk<li class="interlanguage-link interwiki-de mw-list-item">Deutsch<li class="interlanguage-link interwiki-es mw-list-item">Espa\xc3\xb1ol<li class="interlanguage-link interwiki-fa mw-list-item">\xd9\x81\xd8\xa7\xd8\xb1\xd8\xb3\xdb\x8c<li class="interlanguage-link interwiki-fr mw-list-item">Fran\xc3\xa7ais<li class="interlanguage-link interwiki-ko mw-list-item">\xed\x95\x9c\xea\xb5\xad\xec\x96\xb4<li class="interlanguage-link interwiki-id mw-list-item">Bahasa Indonesia<li class="interlanguage-link interwiki-it mw-list-item">Italiano<li class="interlanguage-link interwiki-he mw-list-item">\xd7\xa2\xd7\x91\xd7\xa8\xd7\x99\xd7\xaa<li class="interlanguage-link interwiki-nl mw-list-item">Nederlands<li class="interlanguage-link interwiki-ja mw-list-item">\xe6\x97\xa5\xe6\x9c\xac\xe8\xaa\x9e<li class="interlanguage-link interwiki-pl mw-list-item">Polski<li class="interlanguage-link interwiki-pt mw-list-item">Portugu\xc3\xaas<li class="interlanguage-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%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%B2" title="\xd0\xa1\xd1\x82\xd0\xb0\xd0%BD\xd0\xb4\xd0\xb0\xd1\x80\xd1\x82 \xd0\xb8\xd1%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%B2" title="\xd0\xa1\xd1\x82\xd0\xb0\xd0%BD\xd0\xb4\xd0\xb0\xd1\x80\xd1\x82 \xd0\x

[\xd0\x81\xd0\xba\xd0\xbb\xd1\x8e\xd1\x87\xd0\xbd\xd0\xbd\xd0\xbb\xd0\xbb\xd1\x8f \xd1\x80\xd0\xbe\xd0\xb1\xd0\xbe\xd1\x82\xd0\xbe\xd0\xb2 \xe2\x80\x93 93 Russian" lang="ru" hreflang="ru" class="interlanguage-link-target">\xd0\xa0\xd1\x83\xd1\x81\xd1\x81\xd0\xba\xd0\xb8\xd0\xb9<li class="interlanguage-link interwiki-simple mw-list-item">Simple English<li class="interlanguage-link interwiki-sr mw-list-item">\xd0\xa1\xd1\x80\xd0\xbf\xd1\x81\xd0\xba\xd0\xb8 / srpski<li class="interlanguage-link interwiki-fi mw-list-item">Suomi<li class="interlanguage-link interwiki-sv mw-list-item">Svenska<li class="interlanguage-link interwiki-tr mw-list-item">T\xc3\xbcrk\xc3\xa7e<li class="interlanguage-link interwiki-uk mw-list-item">\xd0\xa3\xd0\xba\xd1\x80\xd0\xb0\xd1\x97\xd0\xbd\xd1\x81\xd1\x8c\xd0\xba\xd0\xb0<li class="interlanguage-link interwiki-zh mw-list-item">\xe4\xb8\xad\xe6\x96\x87\n\t\t<div class="after-portlet after-portlet-lang">Edit links</div>\n\t</div>\n</nav>\n\n</div>\n\n</div>\n\n<footer id="footer" class="mw-footer" role="contentinfo">\n\t<ul id="footer-info">\n\t\t<li id="footer-info-lastmod"> This page was last edited on 10 September 2021, at 17:48 \(UTC\)\n\t\t<li id="footer-info-copyright">Text is available under the Creative Commons Attribution-ShareAlike License;\nadditional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia\xc2\xae is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.\n\t\n\t\n\t<ul id="footer-places">\n\t\t<li id="footer-places-privacy">Privacy policy\n\t\t<li id="footer-places-about">About Wikipedia\n\t\t<li id="footer-places-disclaimer">Disclaimers\n\t\t<li id="footer-places-contact">Contact Wikipedia\n\t\t<li id="footer-places-mobileview">Mobile view\n\t\t<li id="footer-places-developers">Developers\n\t\t<li id="footer-places-statslink">Statistics\n\t\t<li id="footer-places-cookiestatement">Cookie statement\n\t\n\t\n\t<ul id="footer-icons" class="noprint">\n\t\t<li id="footer-copyrightico">\n\t\t<li id="footer-poweredbyico">\n\t\n\t\n\t</div>\n\t\n\t\n\t<script>\(RLQ=window.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
ingprofile":["100.00% 661.793 1 -total"," 45.61% 301.824 1 Template:Refl
ist"," 33.71% 223.116 38 Template:Cite_web"," 12.94% 85.627 1 Template:S
hort_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% 24.22
7 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">{"@context
":"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/Q80776","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,"wgSeps
      atorTransformTable":["",""],"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","wgTi
      tle":"Robots exclusion standard","wgCurRevisionId":1043553668,"wgRevisionId":1043553
      668,"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 desc
      ription 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","w
      gNoticeProject":"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","use

```



```

r.options":"loading","ext.flaggedRevs.icons":"ready","oojs-ui-core.styles":"ready","oojs-ui-core.styles.indicators":"ready","mediawiki.widgets.styles":"ready","oojs-ui-core.icons":"ready","ext.cite.styles":"ready","ext.pygments":"ready","skins.vector.styles.legacy":"ready","ext.flaggedRevs.basic":"ready","ext.visualEditor.desktopArticleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","wikibase.client.init":"ready"};RLPAGEMODULES=["ext.cite.ux-enhancements","site","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","mmv.head","mmv.bootstrap.autostart","ext.popups","ext.visualEditor.desktopArticleTarget.init","ext.visualEditor.targetLoader","ext.eventLogging","ext.wikimediaEvents","ext.navigationTiming","ext.uls.compactlinks","ext.uls.interface","ext.cx.eventlogging.campaigns","ext.centralNotice.geoIP","ext.centralNotice.startUp","ext.centralauth.centralautologin","ext.growthExperiments.SuggestedEditSession"];
</script>
<script>
  (RLQ=window.RLQ||[]).push(function(){mw.loader.implement("user.options@1hzgi",function($,jQuery,require,module){/*@nomin*/mw.user.tokens.set({"patrolToken":"+\\","watchToken":"+\\","csrfToken":"+\\"});});});
</script>
<link href="/w/load.php?lang=en&modules=ext.cite.styles%7Cext.flaggedRevs.basic%7Cicons%7Cext.pygments%7CwikimediaBadges%7Cext.uls.interlanguage%7Cext.visualEditor.desktopArticleTarget.noscript%7Cmediawiki.widgets.styles%7Coojs-ui-core.icons%7Cstyles%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&modules=startup&only=scripts&raw=1&skin=vector">
</script>
<meta content="" name="ResourceLoaderDynamicStyles"/>
<link href="/w/load.php?lang=en&modules=site.styles&only=styles&skin=vector" 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 screen and (max-width: 720px)" rel="alternate"/>
<link href="/w/index.php?title=Robots_exclusion_standard&action=edit" rel="alternate" title="Edit this page" type="application/x-wiki"/>
<link href="/w/index.php?title=Robots_exclusion_standard&action=edit" rel="edit" 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="application/opensearchdescription+xml"/>
<link href="//en.wikipedia.org/w/api.php?action=rsd" rel="EditURI" type="application/rsd+xml"/>
<link href="//creativecommons.org/licenses/by-sa/3.0/" rel="license"/>
<link href="https://en.wikipedia.org/wiki/Robots_exclusion_standard" rel="canonical"/>
<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-editable page-Robots_exclusion_standard rootpage-Robots_exclusion_standard skin-vector action-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
tered and new users are subject to review prior to becoming visible to unregistered
users">
        
      </a>
    </div>
  </div>
  <h1 class="firstHeading" id="firstHeading">
    Robots exclusion standard
  </h1>
  <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
="mw-fr-revisiontag">
        <div class="flaggedrevs_short_basic">
          <span aria-disabled="false" class="flaggedrevs-icon oo-ui-widget oo-ui-widget
-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-
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
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
ecial:Log&type=review&page=Robots_exclusion_standard">
              reviewed
            </a>
            on
            <i>
              10 September 2021
            </i>
            .
          </div>
        </div>
      </div>
    </div>
  </div>
  <div id="contentSub2">

```

```

</div>
<div id="jump-to-nav">
</div>
<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
="en">
  <div class="mw-parser-output">
    <div class="shortdescription nomobile noexcerpt noprint searchaux" style="disp
lay:none">
      Standard used to advise web crawlers and scrapers not to index a web page or
site
    </div>
    <p class="mw-empty-elt">
</p>
<p>
  The
  <b>
    robots exclusion standard
  </b>
  , also known as the
  <b>
    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
e">
    search engines
  </a>
  to categorize websites. Not all robots cooperate with the standard;
  <a class="mw-redirect" href="/wiki/Email_address_harvesting" title="Email add
ress harvesting">
    email harvesters
  </a>
  ,
  <a class="mw-redirect" href="/wiki/Spambots" title="Spambots">
    spambots
  </a>
  ,
  <a href="/wiki/Malware" title="Malware">
    malware
  </a>
  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">
    Sitemaps
  </a>
  , a robot inclusion standard for websites.
</p>
<div aria-labelledby="mw-toc-heading" class="toc" id="toc" role="navigation">
  <input class="toctogglecheckbox" id="toctogglecheckbox" role="button" style
="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>
    <ul>
      <li class="toclevel-1 tocsection-1">
        <a href="#History">
          <span class="tocnumber">
            1
          </span>
          <span class="toctext">
            History
          </span>
        </a>
      </li>
      <li class="toclevel-1 tocsection-2">
        <a href="#Standard">
          <span class="tocnumber">
            2
          </span>
          <span class="toctext">
            Standard
          </span>
        </a>
      </li>
      <li class="toclevel-1 tocsection-3">
        <a href="#Security">
          <span class="tocnumber">
            3
          </span>
          <span class="toctext">
            Security
          </span>
        </a>
      </li>
      <li class="toclevel-1 tocsection-4">
        <a href="#Alternatives">
          <span class="tocnumber">
            4
          </span>
          <span class="toctext">
            Alternatives
          </span>
        </a>
      </li>
      <li class="toclevel-1 tocsection-5">
        <a href="#Examples">
          <span class="tocnumber">
            5
          </span>
          <span class="toctext">

```

```

        Examples
    </span>
</a>
</li>
<li class="toclevel-1 tocsection-6">
    <a href="#Nonstandard_extensions">
        <span class="tocnumber">
            6
        </span>
        <span class="toctext">
            Nonstandard extensions
        </span>
    </a>
    <ul>
        <li class="toclevel-2 tocsection-7">
            <a href="#Crawl-delay_directive">
                <span class="tocnumber">
                    6.1
                </span>
                <span class="toctext">
                    Crawl-delay directive
                </span>
            </a>
        </li>
        <li class="toclevel-2 tocsection-8">
            <a href="#Allow_directive">
                <span class="tocnumber">
                    6.2
                </span>
                <span class="toctext">
                    Allow directive
                </span>
            </a>
        </li>
        <li class="toclevel-2 tocsection-9">
            <a href="#Sitemap">
                <span class="tocnumber">
                    6.3
                </span>
                <span class="toctext">
                    Sitemap
                </span>
            </a>
        </li>
        <li class="toclevel-2 tocsection-10">
            <a href="#Host">
                <span class="tocnumber">
                    6.4
                </span>
                <span class="toctext">
                    Host
                </span>
            </a>
        </li>
        <li class="toclevel-2 tocsection-11">
            <a href='#Universal_"*"_match'>
                <span class="tocnumber">
                    6.5
                </span>
                <span class="toctext">
                    Universal "*" match
                </span>
            </a>
        </li>
    </ul>

```

```

</li>
<li class="toclevel-1 tocsection-12">
  <a href="#Meta_tags_and_headers">
    <span class="tocnumber">
      7
    </span>
    <span class="toctext">
      Meta tags and headers
    </span>
  </a>
</li>
<li class="toclevel-1 tocsection-13">
  <a href="#See_also">
    <span class="tocnumber">
      8
    </span>
    <span class="toctext">
      See also
    </span>
  </a>
</li>
<li class="toclevel-1 tocsection-14">
  <a href="#References">
    <span class="tocnumber">
      9
    </span>
    <span class="toctext">
      References
    </span>
  </a>
</li>
<li class="toclevel-1 tocsection-15">
  <a href="#External_links">
    <span class="tocnumber">
      10
    </span>
    <span class="toctext">
      External links
    </span>
  </a>
</li>
</ul>
</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&action=edit&section=1" title="Edit section: History">
    edit
  </a>
  <span class="mw-editsection-bracket">
    ]
  </span>
</span>
</h2>
<p>
  The standard was proposed by
  <a href="/wiki/Martijn_Koster" title="Martijn Koster">
    Martijn Koster
  </a>

```

,
^{[1]}
when working for
[Nexor](/wiki/Nexor "Nexor")
^{[3]}
in February 1994
^{[4]}
on the
www-talk
mailing list, the main communication channel for WWW-related activities at the time.
[Charles Stross](/wiki/Charles_Stross "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](/wiki/Denial-of-service_attack "Denial-of-service attack")
on Koster's server.
^{[5]}
It quickly became a
[de facto standard](/wiki/De_facto_standard "De facto standard")
that present and future web crawlers were expected to follow; most complied, including those operated by search engines such as
[WebCrawler](/wiki/WebCrawler "WebCrawler")
, [Lycos](/wiki/Lycos "Lycos")
, and [AltaVista](/wiki/AltaVista "AltaVista")

```

        .
        <sup class="reference" id="cite_ref-sear_Robo_6-0">
            <a href="#cite_note-sear_Robo-6">
                [6]
            </a>
        </sup>
    </p>
    <p>
        On July 1, 2019, Google announced the proposal of the Robots Exclusion Protocol as an official standard under
        <a href="/wiki/Internet_Engineering_Task_Force" title="Internet Engineering Task 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.
    </p>
    <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&action=edit&section=2" title="Edit section: Standard">
                edit
            </a>
            <span class="mw-editsection-bracket">
                ]
            </span>
        </span>
    </h2>
    <p>
        When a site owner wishes to give instructions to web robots they place a text file called
        <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-style"/>
        <span class="monospaced">
            https://www.example.com/robots.txt
        </span>
        ). This text file contains the instructions in a specific format (see examples below). Robots that
        <i>
            choose
        </i>

```


to follow the instructions try to fetch this file and read the instructions before 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 not wish to place any limitations on crawling the entire site.

```
</p>
<p>
```

A robots.txt file on a website will function as a request that specified 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 belief 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 only 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.

```
<sup class="reference" id="cite_ref-9">
  <a href="#cite_note-9">
    [9]
  </a>
</sup>
```

```
</p>
<p>
```

A robots.txt file covers one

```
<a class="mw-redirect" href="/wiki/Same_origin_policy" title="Same origin policy">
```

```
  origin
</a>
```

. For websites with multiple subdomains, each subdomain must have its own robots.txt file. If

```
<link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-style"/>
```

```
<span class="monospaced">
  example.com
</span>
```

had a robots.txt file but

```
<link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-style"/>
```

```
<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-style"/>
```

```
<span class="monospaced">
  example.com
</span>
```

would not apply to

```
<link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-style"/>
```

```
<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-style"/>
```

```
<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-style"/>
```

```
<span class="monospaced">
  http://example.com:8080/
</span>
```

```

or
<link href="mw-data:TemplateStyles:r886049734" rel="mw-deduplicated-inline-st
yle"/>
<span class="monospaced">
  https://example.com/
</span>
.
</p>
<p>
  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">
      [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>
  Bing
  <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;">
    [
    <i>

```

```

<a href="/wiki/Wikipedia:Manual_of_Style/Dates_and_numbers#Chronological_it
ems" title="Wikipedia:Manual of Style/Dates and numbers">
  <span title="The time period mentioned near this tag is ambiguous. (Decemb
er 2020)">
    when?
  </span>
</a>
</i>
]
</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>
</p>
<p>
  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>
</p>
<p>
  For some years, the
  <a href="/wiki/Internet_Archive" title="Internet Archive">
    Internet Archive
  </a>
  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>
</p>
<h2>
  <span class="mw-headline" id="Security">
    Security
  </span>

```

```

<span class="mw-editsection">
  <span class="mw-editsection-bracket">
    [
  </span>
  <a href="/w/index.php?title=Robots_exclusion_standard&action=edit&section=3" title="Edit section: Security">
    edit
  </a>
  <span class="mw-editsection-bracket">
    ]
  </span>
</span>
</h2>
<p>
  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
y">
    security through obscurity
  </a>
  is discouraged by standards bodies. The
  <a href="/wiki/National_Institute_of_Standards_and_Technology" title="Nationa
l 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>
</p>
<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&action=edit&section=4" title="Edit section: Alternatives">
    edit
</a>
<span class="mw-editsection-bracket">
    ]
</span>
</span>
</h2>
<p>
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
ailure (or
<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>
</p>
<p>
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>
</p>

```

```

<p>
Previously Google had a joke file hosted at
<code>
/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>
</p>
<h2>
<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&action=edit&section=5" title="Edit section: Examples">
edit
</a>
<span class="mw-editsection-bracket">
]
</span>
</span>
</h2>
<p>
This example tells all robots that they can visit all files because the wildcard
and
<code>
*
</code>
stands for all robots and the
<code>
Disallow
</code>
directive has no value, meaning no pages are disallowed.
</p>
<pre>User-agent: *
Allow: /
</pre>
<p>
The same result can be accomplished with an empty or missing robots.txt file.
</p>
<p>

```

```

    This example tells all robots to stay out of a website:
  </p>
  <pre>User-agent: *
Disallow: /
</pre>

  <p>
    This example tells all robots not to enter three directories:
  </p>
  <pre>User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /junk/
</pre>

  <p>
    This example tells all robots to stay away from one specific file:
  </p>
  <pre>User-agent: *
Disallow: /directory/file.html
</pre>

  <p>
    All other files in the specified directory will be processed.
  </p>
  <p>
    This example tells a specific robot to stay out of a website:
  </p>
  <pre>User-agent: BadBot # replace 'BadBot' with the actual user-agent of the bot
Disallow: /
</pre>

  <p>
    This example tells two specific robots not to enter one specific directory:
  </p>
  <pre>User-agent: BadBot # replace 'BadBot' with the actual user-agent of the bot
User-agent: Googlebot
Disallow: /private/
</pre>

  <p>
    Example demonstrating how comments can be used:
  </p>
  <pre># Comments appear after the "#" symbol at the start of a line, or after a directive
User-agent: * # match all bots
Disallow: / # keep them out
</pre>

  <p>
    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 to 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>
  </p>
  <p>
    Example demonstrating multiple user-agents:
  </p>
  <pre>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
</pre>
```

```
<h2>
  <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&action=edit&section=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&action=edit&section=7" title="Edit section: Crawl-delay directive">
      edit
    </a>
    <span class="mw-editsection-bracket">
      ]
    </span>
  </span>
</h3>
<p>
```

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 dependent 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>
  </p>
  <pre>User-agent: bingbot
Allow : /
Crawl-delay: 10
</pre>
  <h3>
    <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&action=edit&section=8" title="Edit section: Allow directive">
        edit
      </a>
      <span class="mw-editsection-bracket">
        ]
      </span>
    </span>
  </h3>
  <p>
    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 wants some HTML documents in that directory crawled and indexed. While by standard implementation the first matching robots.txt pattern always wins, Google's implementation differs in that Allow patterns with equal or more characters in the directive path 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>
</p>
<p>
    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:
    </p>
    <pre>Allow: /directory1/myfile.html
    Disallow: /directory1/
    </pre>
    <p>
        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.
    </p>
    <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&action=edit&se
            ction=9" title="Edit section: Sitemap">
                edit
            </a>
            <span class="mw-editsection-bracket">
                ]
            </span>
        </span>
    </h3>
    <p>
        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>
</p>
<pre>Sitemap: http://www.example.com/sitemap.xml</pre>
<h3>
  <span class="mw-headline" id="Host">
    Host
  </span>
  <span class="mw-editsection">
    <span class="mw-editsection-bracket">
      [
    </span>
    <a href="/w/index.php?title=Robots_exclusion_standard&action=edit&section=10" title="Edit section: Host">
      edit
    </a>
    <span class="mw-editsection-bracket">
      ]
    </span>
  </span>
</h3>
<p>
  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>
</p>
<pre>Host: hosting.example.com
</pre>
<p>
  This is not supported by all crawlers.
</p>
<h3>
  <span id="Universal_.22.2A.22_match">
  </span>
  <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&action=edit&section=11" title="Edit section: Universal "" match">
      edit
    </a>
    <span class="mw-editsection-bracket">
      ]
    </span>
  </span>
</h3>
<p>
  The
  <i>
    Robot Exclusion Standard

```

```

</i>
does not mention the "*" character in the
<code>
Disallow:
</code>
statement.
<sup class="reference" id="cite_ref-40">
  <a href="#cite_note-40">
    [40]
  </a>
</sup>
</p>
<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&action=edit&section=12" title="Edit section: Meta tags and headers">
    edit
  </a>
  <span class="mw-editsection-bracket">
    ]
  </span>
</span>
</h2>
<p>
  In addition to root-level robots.txt files, robots exclusion directives can be
  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-HTML
  files such as images, text files, 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 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>
</p>
<dl>
  <dt>
    A "noindex" meta tag
  </dt>
</dl>
<div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr">
  <pre><span></span><span class="p">&lt;</span><span class="nt">meta</span> <span class="na">name</span><span class="o">=</span><span class="s">"robots"</span> <span class="na">content</span><span class="o">=</span><span class="s">"noindex"</span>
  <span class="p">&gt;</span>
</pre>
</div>
<dl>

```

```

<dt>
  A "noindex" HTTP response header
</dt>
</dl>
<div class="mw-highlight mw-highlight-lang-html mw-content-ltr" dir="ltr">
<pre><span></span>X-Robots-Tag: noindex
</pre>
</div>
<p>
  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 load
  ed, 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>
</p>
<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&action=edit&section=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-ouptu
  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;">
<ul>
<li>
  <code>
    <a href="/wiki/Ads.txt" title="Ads.txt">
      ads.txt
    </a>
  </code>
  , a standard for listing authorized ad sellers
</li>
<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 orde
  r to report security vulnerabilities
</li>
<li>
    <a href="/wiki/Automated_Content_Access_Protocol" title="Automated Content

```

```

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

```

```

Spider trap


- Web archiving
- Web crawler
- noindex


## References [ edit ]



.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-style-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{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .reflist-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-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}



- ^


.mw-parser-output cite.citation{font-style:inherit}.mw-parser-output .citation q{quotes:"\"""\"""\"""\""}.


```

```

upload.wikimedia.org/wikipedia/commons/6/65/Lock-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-limited a,.mw-parser-output .citation .cs1-lock-registration a{background:linear-gradient(transparent,transparent),url("//upload.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("//upload.wikimedia.org/wikipedia/commons/a/aa/Lock-red-alt-2.svg")right 0.1em center/9px no-repeat}.mw-parser-output .cs1-subscription,.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: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-kern-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.html" rel="nofollow">
    "Historical"
  </a>
  .
  <i>
    Greenhills.co.uk
  </i>
  .
  <a class="external text" href="https://web.archive.org/web/20170403152037/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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Greenhills.co.uk&rft.atitle=Historical&rft_id=http%3A%2F%2Fwww.greenhills.co.uk%2Fhistorical.html&rft_rft_id=info%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
  </span>
</span>
</li>
<li 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-style"/>
    <cite class="citation web cs1" id="CITEREFFielding1994">
      Fielding, Roy (1994).
      <a class="external text" href="http://www94.web.cern.ch/WWW94/PapersWWW9

```



```

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
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&rft_val_fmt=info%3Aofi%
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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standa
rd">
    </span>
    </span>
    </li>
    <li id="cite_note-:0-3">
        <span class="mw-cite-backlink">
            <b>
                <a href="#cite_ref-:0_3-0">
                    ^
                </a>
            </b>
        </span>
        <span class="reference-text">
            <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
            <cite class="citation web cs1">
                <a class="external text" href="http://www.robotstxt.org/orig.html#status"
rel="nofollow">
                    "The Web Robots Pages"
                </a>
                . Robotstxt.org. 1994-06-30.
                <a class="external text" href="https://web.archive.org/web/2014011209063
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>
            </span>
            .

```

```

</cite>
<span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%
2Fmt%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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclus
ion+standard">
</span>
</span>
</li>
<li id="cite_note-4">
<span class="mw-cite-backlink">
<b>
<a href="#cite_ref-4">
^
</a>
</b>
</span>
<span class="reference-text">
<link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
<cite class="citation web cs1" id="CITEREFKoster1994">
Koster, Martijn (25 February 1994).
<a class="external text" href="https://web.archive.org/web/2013102920035
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
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&rft_val_fmt=info%3Aofi%
2Fmt%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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots
+exclusion+standard">
</span>
</span>
</li>
<li 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
-style"/>
<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

.
</cite>

<li id="cite_note-sear_Robo-6">

^

<link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/>
<cite class="citation web cs1" id="CITEREFBarry_Schwartz2014">
Barry Schwartz (30 June 2014).

"Robots.txt Celebrates 20 Years Of Blocking Search Engines"

.
<i>
Search Engine Land
</i>
.

Archived

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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Search+Engine+Land&rft.atitle=Robots.txt+Celebrates+20+Years+Of+Blocking+Search+Engines&rft.date=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%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</li>
<li id="cite_note-7">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-7">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://webmasters.googleblog.com/2019/07/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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=Official+Google+Webmaster+Central+Blog&rft.atitle=Formalizing+the+Robots+Exclusion+Protocol+Specification&rft_id=https%3A%2F%2Fwebmasters.googleblog.com%2F2019%2F07%2Frep-id.html&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
        </span>
    </li>
<li 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-style"/>
        <cite class="citation web cs1" id="CITEREFM._Koster,_Stalworthy_Computing,

```

_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
 opic-00" rel="nofollow">
 "Robots Exclusion Protocol draft-rep-wg-topic-00"

 .
 <i>
 IETF Documents
 </i>

 . Retrieved

 2019-09-09

 .
 </cite>
 <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%
 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">

 CS1 maint: multiple names: authors list (
 <a href="/wiki/Category:CS1_maint:_multiple_names:_authors_list" title="C
 ategory:CS1 maint: multiple names: authors list">
 link

)

 <li id="cite_note-9">

 ^

 <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
 -style"/>
 <cite class="citation web cs1">
 <a class="external text" href="https://www.youtube.com/watch?v=KBdEwpRQRD
 0#t=196s" rel="nofollow">
 "Uncrawled URLs in search results"

 . YouTube. Oct 5, 2009.
 <a class="external text" href="https://web.archive.org/web/2014010622250
 0/http://www.youtube.com/watch?v=KBdEwpRQRD0#t=196s" rel="nofollow">
 Archived

 from the original on 2014-01-06

 . Retrieved

 2013-12-29

 .

```

        </cite>
        <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%
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>
    </li>
    <li id="cite_note-ask-webmasters-10">
        <span class="mw-cite-backlink">
            <b>
                <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
-style"/>
            <cite class="citation web cs1">
                <a class="external text" href="http://about.ask.com/docs/about/webmaster
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&rft_val_fmt=info%3Aofi%
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>
        </li>
    <li id="cite_note-about-aol-search-11">
        <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
-style"/>
            <cite class="citation web cs1">
                <a class="external text" href="http://search.aol.com/aol/about" rel="nofo
llow">
                    "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&rft_val_fmt=info%3Aofi%
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&
p;rfr_id=info%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
</span>
</span>
</li>
<li id="cite_note-baidu-spider-12">
<span class="mw-cite-backlink">
<b>
<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
-style"/>
<cite class="citation web cs1">
<a class="external text" href="http://www.baidu.com/search/spider_englis
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&rft_val_fmt=info%3Aofi%
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%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
</span>
</span>
</li>
<li id="cite_note-duckduckgo-bot-13">
<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
-style"/>
<cite class="citation web cs1">

```

```

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&rft_val_fmt=info%3Aofi%
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>
</li>
<li id="cite_note-google-webmasters-spec-14">
<span class="mw-cite-backlink">
^
<a href="#cite_ref-google-webmasters-spec_14-0">
<sup>
<i>
<b>
a
</b>
</i>
</sup>
</a>
<a href="#cite_ref-google-webmasters-spec_14-1">
<sup>
<i>
<b>
b
</b>
</i>
</sup>
</a>
</span>
<span class="reference-text">
<link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
<cite class="citation web cs1">
<a class="external text" href="https://developers.google.com/webmasters/c
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
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&rft_val_fmt=info%3Aofi%
2Fmt%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>
</li>
<li id="cite_note-yahoo-search-is-bing-15">
    <span class="mw-cite-backlink">
        <b>
            <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
-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="http://help.yahoo.com/kb/index?page=conten
t&y=PROD_SRCH&locale=en_US&id=SLN2217&impressions=true" rel="nofollow"
w">
                "Submitting your website to Yahoo! Search"
            </a>
            .
            <a class="external text" href="https://web.archive.org/web/2013012103580
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&rft_val_fmt=info%3Aofi%
2Fmt%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%3DSL2217%26impressions%3Dtrue&rfr_i
d=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
            </span>
        </span>
    </li>
<li id="cite_note-yandex-robots-16">
    <span class="mw-cite-backlink">
        ^
        <a href="#cite_ref-yandex-robots_16-0">
            <sup>
                <i>
                    <b>

```

```

        a
      </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
-style"/>
  <cite class="citation web cs1">
    <a class="external text" href="http://help.yandex.com/webmaster/?id=11138
51" rel="nofollow">
      "Using robots.txt"
    </a>
    .
    <i>
      Help.yandex.com
    </i>
    .
    <a class="external text" href="https://web.archive.org/web/2013012504001
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&rft_val_fmt=info%3Aofi%
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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
    </span>
  </span>
</li>
<li id="cite_note-bing-blog-robots-17">
  <span class="mw-cite-backlink">
    ^
  <a href="#cite_ref-bing-blog-robots_17-0">
    <sup>
      <i>
        <b>
          a
        </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
-style"/>
    <cite class="citation web cs1">
        <a class="external text" href="https://blogs.bing.com/webmaster/2008/06/0
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
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&rft_val_fmt=info%3Aofi%
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%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</li>
<li 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
-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://www.bing.com/webmaster/help/how-to
-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&rft_val_fmt=info%3Aofi%
2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=www.bing.com&rft
.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%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
  </span>
</span>
</li>
<li id="cite_note-19">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-19">
        ^
      </a>
    </b>
  </span>
  <span class="reference-text">
    <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-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
=Robots.txt" rel="nofollow">
        "Robots.txt is a suicide note"
      </a>
      . Archive Team.
      <a class="external text" href="https://web.archive.org/web/2017021804452
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&rft_val_fmt=info%3Aofi%
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%3Aid%2Fen.wikipedi
a.org%3ARobots+exclusion+standard">
    </span>
  </span>
</li>
<li 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
- style"/>
      <cite class="citation web cs1">
        <a class="external text" href="https://blog.archive.org/2017/04/17/robots
- 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
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%
2Ffmt%3Akev%3Amtx%3Ajournal&amp;rft.genre=unknown&amp;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&amp;rft.date=2017-04-17&amp;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&amp;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standar
d">
        </span>
      </span>
    </li>
    <li id="cite_note-Internet_Archive-21">
      <span class="mw-cite-backlink">
        ^
        <a href="#cite_ref-Internet_Archive_21-0">
          <sup>
            <i>
              <b>
                a
              </b>
            </i>
          </sup>
        </a>
        <a href="#cite_ref-Internet_Archive_21-1">
          <sup>
            <i>
              <b>
                b
              </b>
            </i>
          </sup>
        </a>
      </span>
    <span class="reference-text">
      <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline

```

```

-style"/>
    <cite class="citation news cs1" id="CITEREFJones2017">
        Jones, Brad (24 April 2017).
        <a class="external text" href="https://www.digitaltrends.com/computing/in
ternet-archive-robots-txt/#ixzz4gQY0qpUi" 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
9/https://www.digitaltrends.com/computing/internet-archive-robots-txt/#ixzz4gQY0qpU
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&rft_val_fmt=info%3Aofi%
2Fmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.jtitle=Digital+Trends&
rft.atitle=The+Internet+Archive+Will+Ignore+Robots.txt+Files+to+Maintain+Accuracy&
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=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</li>
<li id="cite_note-22">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-22">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://support.google.com/webmasters/answ
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
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>

```

```

        </span>
        .
    </cite>
    <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Block+URLs+with+robots.txt%3A+Learn+about+robots.txt+files&rft_id=https%3A%2F%2Fsupport.google.com%2Fwebmasters%2Fanswer%2F6062608&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</li>
<li id="cite_note-23">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-23">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://www.theregister.co.uk/2015/05/19/robotstxt/" 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/20150821063759/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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=unknown&rft.jtitle=The+Register&rft.atitle=Robots.txt+tells+hackers+the+places+you+don%27t+want+them+to+look&rft_id=https%3A%2F%2Fwww.theregister.co.uk%2F2015%2F05%2F19%2Frobotstxt%2F&rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
        </span>
    </li>
<li id="cite_note-24">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-24">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-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/nistpubs/800-123/SP800-123.pdf" rel="nofollow">
    "Guide to General Server Security"
  </a>
  <span class="cs1-format">
    (PDF)
  </span>
  . National Institute of Standards and Technology.
  <a class="mw-redirect" href="/wiki/Doi_(identifier)" title="Doi (identifier)">
    doi
  </a>
  :
  <a class="external text" href="https://doi.org/10.6028/2FNIST.SP.800-123"
rel="nofollow">
    10.6028/NIST.SP.800-123
  </a>
  .
  <a class="external text" href="https://web.archive.org/web/20111008115412/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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=Guide+to+General+Server+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=Tracy%2C+M.&rft_id=http%3A%2F%2Fsrc.nist.gov%2Fpublications%2Fnistpubs%2F800-123%2FSP800-123.pdf&rft_id=info%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
  </span>
  <span class="cs1-hidden-error error citation-comment">
    Cite journal requires
    <code class="cs1-code">
      |journal=
    </code>
    (
    <a href="/wiki/Help:CS1_errors#missing_periodical" title="Help:CS1 errors">
      help
    </a>
    )
  </span>
</span>
</li>
<li id="cite_note-25">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-25">
        ^
      </a>
    </b>
  </span>

```



```

        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
- 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
KogC&amp;pg=PA92" rel="nofollow">
                <i>
                    Innocent Code: A Security Wake-Up Call for Web Programmers
                </i>
            </a>
            . John Wiley &amp; Sons. pp. 91-92.
            <a class="mw-redirect" href="/wiki/ISBN_(identifier)" title="ISBN (identi
fier)">
                ISBN
            </a>
            <a href="/wiki/Special:BookSources/9780470857472" title="Special:BookSour
ces/9780470857472">
                <bdi>
                    9780470857472
                </bdi>
            </a>
            .
            <a class="external text" href="https://web.archive.org/web/2016040119343
7/https://books.google.com/books?id=RjVjgPQsKogC&amp;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%
2Ffmt%3Akev%3Amtx%3Abook&amp;rft.genre=book&amp;rft.btitle=Innocent+Code%3A+A+Securi
ty+Wake-Up+Call+for+Web+Programmers&amp;rft.pages=91-92&amp;rft.pub=John+Wiley+%26+S
ons&amp;rft.date=2004&amp;rft.isbn=9780470857472&amp;rft.au=Sverre+H.+Huseby&amp;rft
_id=https%3A%2F%2Fbooks.google.com%2Fbooks%3Fid%3DRjVjgPQsKogC%26pg%3DPA92&amp;rfr_i
d=info%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
            </span>
        </span>
    </li>
    <li 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
- style"/>
            <cite class="citation web cs1">
                <a class="external text" href="http://www.user-agents.org/" rel="nofollo
w">
                    "List of User-Agents (Spiders, Robots, Browser)"
                </a>
                . User-agents.org.
                <a class="external text" href="https://web.archive.org/web/2014010715420

```

```

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&rft_val_fmt=info%3Aofi%
2Fmt%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&rft_id=info%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+
standard">
  </span>
</span>
</li>
<li id="cite_note-27">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-27">
        ^
      </a>
    </b>
  </span>
  <span class="reference-text">
    <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
    <cite class="citation web cs1">
      <a class="external text" href="https://httpd.apache.org/docs/2.2/howto/ac
cess.html" rel="nofollow">
        "Access Control - Apache HTTP Server"
      </a>
      . Httpd.apache.org.
      <a class="external text" href="https://web.archive.org/web/2013122911083
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&rft_val_fmt=info%3Aofi%
2Fmt%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&rft_id=info%3Aid%2Fen.wikipedia.org%3ARobo
ts+exclusion+standard">
      </span>
    </span>
  </li>
<li id="cite_note-28">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-28">
        ^
      </a>
    </b>
  </span>

```

```

</span>
<span class="reference-text">
  <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
  <cite class="citation web cs1">
    <a class="external text" href="https://www.iis.net/configreference/system.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/20140101112730/http://www.iis.net/configreference/system.webserver/security/requestfiltering/filteringrules/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&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Deny+Strings+for+Filtering+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%2Fsecurity%2Frequestfiltering%2Ffilteringrules%2Ffilteringrule%2Fdenystrings&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
  </span>
</li>
<li id="cite_note-29">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-29">
        ^
      </a>
    </b>
  </span>
  <span class="reference-text">
    <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
    <cite class="citation web cs1">
      <a class="external text" href="https://www.google.com/humans.txt" rel="nofollow">
        "Google humans.txt"
      </a>
      <span class="reference-accessdate">
        . Retrieved
        <span class="nowrap">
          October 3,
        </span>
        2019
      </span>
      .
    </cite>
    <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Google+humans.txt&rft_id=https%3A%2F%2Fwww.google.com%2Fhumans.txt&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
      </span>
    </span>
  </li>

```

```

</li>
<li 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
- style"/>
    <cite class="citation web cs1">
      <a class="external text" href="https://github.com/humans.txt" rel="nofoll
ow">
        "Github humans.txt"
      </a>
      <span class="reference-accessdate">
        . Retrieved
        <span class="nowrap">
          October 3,
        </span>
        2019
      </span>
      .
    </cite>
    <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%
2Ffmt%3Akev%3Amtx%3Abook&rft.genre=unknown&rft.btitle=Github+humans.txt&
rft_id=https%3A%2F%2Fgithub.com%2Fhumans.txt&rft_id=info%3Aid%2Fen.wikipedia.or
g%3ARobots+exclusion+standard">
    </span>
  </span>
</li>
<li id="cite_note-31">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-31">
        ^
      </a>
    </b>
  </span>
  <span class="reference-text">
    <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
- 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
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&rft_val_fmt=info%3Aofi%
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>
</li>
<li id="cite_note-32">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-32">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-stylesheet"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://www.google.com/killer-robots.txt"
rel="nofollow">
                "/killer-robots.txt"
            </a>
            . 2018-01-10.
            <a class="external text" href="https://web.archive.org/web/2018011016091
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& rft_val_fmt=info%3Aofi%
2Fmt%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>
    </li>
<li id="cite_note-bing-crawl-delay-33">
    <span class="mw-cite-backlink">
        <b>
            <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
-stylesheet"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://blogs.bing.com/webmaster/2012/05/0
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
2/https://blogs.bing.com/webmaster/2012/05/03/to-crawl-or-not-to-crawl-that-is-bingb
ots-question/" rel="nofollow">

```

Archived

 from the original on 2016-02-03

 . Retrieved

 9 February

 2016

 .
 </cite>

 <li id="cite_note-34">

 ^

 <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/>
 <cite class="citation web cs1">

 "Change Googlebot crawl rate - Search Console Help"

 .
 <i>
 support.google.com
 </i>
 .

 Archived

 from the original on 2018-11-18

 . Retrieved

 22 October

 2018

 .
 </cite>

 <li 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
-style"/>
  <cite class="citation web cs1">
    <a class="external text" href="https://www.google.com/support/webmasters/
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
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&rft_val_fmt=info%3Aofi%
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%3Aid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
  </span>
</li>
<li id="cite_note-36">
  <span class="mw-cite-backlink">
    <b>
      <a href="#cite_ref-36">
        ^
      </a>
    </b>
  </span>
  <span class="reference-text">
    <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
    <cite class="citation web cs1">
      <a class="external text" href="http://help.yahoo.com/l/us/yahoo/search/we
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
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>
</li>

```

```

        </span>
    </span>
    .
</cite>
    <span class="Z3988" title="ctx_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%
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&
p;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</span>
</li>
<li id="cite_note-37">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-37">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="http://blog.semetrical.com/googles-secret-
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
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&rft_val_fmt=info%3Aofi%
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&p;rfr_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclus
ion+standard">
    </span>
</span>
</li>
<li id="cite_note-38">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-38">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="https://web.archive.org/web/2009030506184
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/webmasters-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&rft_val_fmt=info%3Aofi%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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
    </span>
</li>
<li id="cite_note-39">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-39">
                ^
            </a>
        </b>
    </span>
    <span class="reference-text">
        <link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/>
        <cite class="citation web cs1">
            <a class="external text" href="http://help.yandex.com/webmaster/?id=1113851" rel="nofollow">
                "Yandex - Using robots.txt"
            </a>
            .
            <a class="external text" href="https://web.archive.org/web/20130509230548/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&rft_val_fmt=info%3Aofi%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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standard">
        </span>
    </li>
<li id="cite_note-40">
    <span class="mw-cite-backlink">
        <b>
            <a href="#cite_ref-40">

```

```

^
</a>
</b>
</span>
<span class="reference-text">
<link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline
-style"/>
<cite class="citation web cs1">
<a class="external text" href="https://developers.google.com/search/refer
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%
2Ffmt%3Akev%3Amtx%3Ajournal&amp;rft.genre=unknown&amp;rft.jtitle=Google+Developers&a
mp;rft.atitle=Robots.txt+Specifications&amp;rft_id=https%3A%2F%2Fdevelopers.google.c
om%2Fsearch%2Freference%2Frobots_txt%3Fhl%3Den&amp;rfr_id=info%3Asid%2Fen.wikipedia.
org%3ARobots+exclusion+standard">
</span>
</span>
</li>
<li id="cite_note-google-meta-41">
<span class="mw-cite-backlink">
^
<a href="#cite_ref-google-meta_41-0">
<sup>
<i>
<b>
a
</b>
</i>
</sup>
</a>
<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
-style"/>
<cite class="citation web cs1">
<a class="external text" href="https://developers.google.com/webmasters/c
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
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&rft_val_fmt=info%3Aofi%
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&rft_id=info%3Asid%2Fen.wikipedia.org%3ARobots+exclusion+standa
rd">
        </span>
    </span>
</li>
</ol>
</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&action=edit&se
ction=15" title="Edit section: External links">
            edit
        </a>
        <span class="mw-editsection-bracket">
            ]
        </span>
    </span>
</h2>
<ul>
<li>
    <span class="official-website">
        <span class="url">
            <a class="external text" href="http://www.robotstxt.org" rel="nofollow">
                Official website
            </a>
        </span>
    </span>
</li>
<li>
    <a class="external text" href="http://www.robotstxt.org/db.html" rel="nofoll
ow">
        Robots Database (list of bot names)
    </a>
</li>
<li>
    <a class="external text" href="https://sitechecker.pro/robots-tester/" rel
="nofollow">
        Robots.txt checker tool
    </a>
</li>
<li>

```

```

<div class="external text" href="https://www.bing.com/webmaster/help/how-to-cr
eate-a-robots-txt-file-cb7c31ec" rel="nofollow">
  Microsoft bingbot official documentation
</div>
</li>
</ul>
<div aria-label="Portals" class="noprint metadata navbox" role="navigation" st
yle="font-weight:bold;padding:0.4em 2em">
  <ul style="margin:0.1em 0 0">
    <li style="display:inline">
      <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">
            
          </a>
        </span>
        <a href="/wiki/Portal:Internet" title="Portal:Internet">
          Internet portal
        </a>
      </span>
    </li>
  </ul>
</div>
<div aria-labelledby="Authority_control_frameless_&#124;text-top_&#12
4;10px_&#124;alt=Edit_this_at_Wikidata_&#124;link=https&#58;//www.wikida
ta.org/wiki/Q80776#identifiers&#124;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">
    <tbody>
      <tr>
        <th class="navbox-group" id="Authority_control_frameless_&#124;text-to
p_&#124;10px_&#124;alt=Edit_this_at_Wikidata_&#124;link=https&#58;//
www.wikidata.org/wiki/Q80776#identifiers&#124;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
is at Wikidata">
            
          </a>
        </th>
        <td class="navbox-list navbox-odd" style="text-align:left;border-left-widt
h:2px;border-left-style:solid;width:100%;padding:0px">
          <div style="padding:0em 0.25em">
            <ul>
              <li>
                <span class="uid">
                  <a class="external text" href="https://academic.microsoft.com/v2/deta
il/163978704" rel="nofollow">
                    Microsoft Academic
                  </a>

```

```

        </span>
      </li>
    </ul>
  </div>
</td>
</tr>
</tbody>
</table>
</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!canonical and timestamp 20210919053121 and revision id 1043553668. Serialized with JSON.
-->
</div>
<noscript>
  
</noscript>
<div class="printfooter">
  Retrieved from "
  <a dir="ltr" href="https://en.wikipedia.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>
<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>
  </div>
</div>

```

```

:
<ul>
<li>
<a href="/wiki/Category:World_Wide_Web" title="Category:World Wide Web">
World Wide Web
</a>
</li>
<li>
<a href="/wiki/Category:Internet_bots" title="Category:Internet bots">
Internet bots
</a>
</li>
</ul>
</div>
<div class="mw-hidden-catlinks mw-hidden-cats-hidden" id="mw-hidden-catlinks">
Hidden categories:
<ul>
<li>
<a href="/wiki/Category:CS1_maint:_multiple_names:_authors_list" title="Category:CS1 maint: multiple names: authors list">
CS1 maint: multiple names: authors list
</a>
</li>
<li>
<a href="/wiki/Category:CS1_errors:_missing_periodical" title="Category:CS1 errors: missing periodical">
CS1 errors: missing periodical
</a>
</li>
<li>
<a href="/wiki/Category:Articles_with_short_description" title="Category:Articles with short description">
Articles with short description
</a>
</li>
<li>
<a href="/wiki/Category:Short_description_matches_Wikidata" title="Category:Short description matches Wikidata">
Short description matches Wikidata
</a>
</li>
<li>
<a href="/wiki/Category:Wikipedia_pending_changes_protected_pages" title="Category:Wikipedia pending changes protected pages">
Wikipedia pending changes protected pages
</a>
</li>
<li>
<a href="/wiki/Category:All_articles_with_vague_or_ambiguous_time" title="Category:All articles with vague or ambiguous time">
All articles with vague or ambiguous time
</a>
</li>
<li>
<a href="/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>
</li>
<li>
<a href="/wiki/Category:Official_website_different_in_Wikidata_and_Wikipedia" title="Category:Official website different in Wikidata and Wikipedia">
Official website different in Wikidata and Wikipedia
</a>
</li>

```

```

        <li>
            <a href="/wiki/Category:Articles_with_MA_identifiers" title="Category:Articles with MA identifiers">
                Articles with MA identifiers
            </a>
        </li>
    </ul>
</div>
</div>
</div>
<div id="mw-data-after-content">
    <div class="read-more-container">
    </div>
</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 vector-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">
                <ul class="vector-menu-content-list">
                    <li class="mw-list-item" id="pt-anonuserpage">
                        Not logged in
                    </li>
                    <li class="mw-list-item" id="pt-anontalk">
                        <a accesskey="n" href="/wiki/Special:MyTalk" title="Discussion about edits from this IP address [n]">
                            Talk
                        </a>
                    </li>
                    <li class="mw-list-item" id="pt-anoncontribs">
                        <a accesskey="y" href="/wiki/Special:MyContributions" title="A list of edits made from this IP address [y]">
                            Contributions
                        </a>
                    </li>
                    <li class="mw-list-item" id="pt-createaccount">
                        <a href="/w/index.php?title=Special:CreateAccount&returnto=Robots+exclusion+standard" title="You are encouraged to create an account and log in; however, it is not mandatory">
                            Create account
                        </a>
                    </li>
                    <li class="mw-list-item" id="pt-login">
                        <a accesskey="o" href="/w/index.php?title=Special:UserLogin&returnto=Robots+exclusion+standard" title="You're encouraged to log in; however, it's not mandatory. [o]">
                            Log in
                        </a>
                    </li>
                </ul>
            </div>
        </nav>
        <div id="left-navigation">
            <nav aria-labelledby="p-namespaces-label" class="mw-portlet mw-portlet-namespaces 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">
        <ul class="vector-menu-content-list">
            <li class="selected mw-list-item" id="ca-nstab-main">
                <a accesskey="c" href="/wiki/Robots_exclusion_standard" title="View the con
tent page [c]">
                    Article
                </a>
            </li>
            <li class="mw-list-item" id="ca-talk">
                <a accesskey="t" href="/wiki/Talk:Robots_exclusion_standard" rel="discussio
n" title="Discuss improvements to the content page [t]">
                    Talk
                </a>
            </li>
        </ul>
    </div>
</nav>
    <nav aria-labelledby="p-variants-label" class="mw-portlet mw-portlet-variants e
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
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">
            <ul class="vector-menu-content-list">
            </ul>
        </div>
    </nav>
</div>
<div id="right-navigation">
    <nav aria-labelledby="p-views-label" class="mw-portlet mw-portlet-views vector-
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">
            <ul class="vector-menu-content-list">
                <li class="selected mw-list-item" id="ca-view">
                    <a href="/wiki/Robots_exclusion_standard">
                        Read
                    </a>
                </li>
                <li class="mw-list-item" id="ca-edit">
                    <a accesskey="e" href="/w/index.php?title=Robots_exclusion_standard&act
ion=edit" title="Edit this page [e]">
                        Edit
                    </a>
                </li>
            </ul>
        </div>
    </nav>
</div>

```



```

        <li class="mw-list-item" id="ca-history">
            <a accesskey="h" href="/w/index.php?title=Robots_exclusion_standard&action=history" title="Past revisions of this page [h]">
                View history
            </a>
        </li>
    </ul>
</div>
</nav>
<nav aria-labelledby="p-cactions-label" class="mw-portlet mw-portlet-cactions emptyPortlet vector-menu-dropdown-noicon vector-menu vector-menu-dropdown" id="p-cactions" role="navigation" title="More options">
    <input aria-haspopup="true" aria-labelledby="p-cactions-label" class="vector-menu-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">
        <ul class="vector-menu-content-list">
        </ul>
    </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="search" 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" name="fulltext" title="Search Wikipedia for this text" type="submit" value="Search"/>
                <input class="searchButton" id="searchButton" name="go" title="Go to a page with this exact name if it exists" type="submit" value="Go"/>
            </div>
        </form>
    </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-navigation 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">
<ul class="vector-menu-content-list">
<li class="mw-list-item" id="n-mainpage-description">
<a accesskey="z" href="/wiki/Main_Page" title="Visit the main page [z]">
Main page
</a>
</li>
<li class="mw-list-item" id="n-contents">
<a href="/wiki/Wikipedia:Contents" title="Guides to browsing Wikipedia">
Contents
</a>
</li>
<li class="mw-list-item" id="n-currentevents">
<a href="/wiki/Portal:Current_events" title="Articles related to current eve
nts">
Current events
</a>
</li>
<li class="mw-list-item" id="n-randompage">
<a accesskey="x" href="/wiki/Special:Random" title="Visit a randomly selecte
d article [x]">
Random article
</a>
</li>
<li class="mw-list-item" id="n-aboutsite">
<a href="/wiki/Wikipedia:About" title="Learn about Wikipedia and how it work
s">
About Wikipedia
</a>
</li>
<li class="mw-list-item" id="n-contactpage">
<a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us" title="How to contact
Wikipedia">
Contact us
</a>
</li>
<li class="mw-list-item" id="n-sitesupport">
<a href="https://donate.wikimedia.org/wiki/Special:FundraiserRedirector?utm_
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>
</li>
</ul>
</div>
</nav>
<nav aria-labelledby="p-interaction-label" class="mw-portlet mw-portlet-interact
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">
<ul class="vector-menu-content-list">
<li class="mw-list-item" id="n-help">
<a href="/wiki/Help:Contents" title="Guidance on how to use and edit Wikiped
ia">
Help
</a>
</li>
<li class="mw-list-item" id="n-introduction">
<a href="/wiki/Help:Introduction" title="Learn how to edit Wikipedia">
Learn to edit
</a>

```

```

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

```

```

    </li>
    <li class="mw-list-item" id="t-cite">
      <a href="/w/index.php?title=Special:CiteThisPage&page=Robots_exclusion_s
tandard&id=1043553668&wpFormIdentifier=titleform" title="Information on how
to cite this page">
        Cite this page
      </a>
    </li>
    <li class="mw-list-item" id="t-wikibase">
      <a accesskey="g" href="https://www.wikidata.org/wiki/Special:EntityPage/Q807
76" title="Structured data on this page hosted by Wikidata [g]">
        Wikidata item
      </a>
    </li>
  </ul>
</div>
</nav>
<nav aria-labelledby="p-coll-print_export-label" class="mw-portlet mw-portlet-co
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">
    <ul class="vector-menu-content-list">
      <li class="mw-list-item" id="coll-download-as-r1">
        <a href="/w/index.php?title=Special:DownloadAsPdf&page=Robots_exclusion_
standard&action=show-download-screen" title="Download this page as a PDF file">
          Download as PDF
        </a>
      </li>
      <li class="mw-list-item" id="t-print">
        <a accesskey="p" href="/w/index.php?title=Robots_exclusion_standard&prin
table=yes" title="Printable version of this page [p]">
          Printable version
        </a>
      </li>
    </ul>
  </div>
</nav>
<nav aria-labelledby="p-lang-label" class="mw-portlet mw-portlet-lang vector-men
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">
    <ul class="vector-menu-content-list">
      <li class="interlanguage-link interwiki-ar mw-list-item">
        <a class="interlanguage-link-target" href="https://ar.wikipedia.org/wiki/%D
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
4%D8%B1%D9%88%D8%A8%D9%88%D8%AA%D8%A7%D8%AA" hreflang="ar" lang="ar" title="معيار استبعاد
الروبوتات - Arabic">
          العربية
        </a>
      </li>
      <li class="interlanguage-link interwiki-bar mw-list-item">
        <a class="interlanguage-link-target" href="https://bar.wikipedia.org/wiki/Ro
bots_Exclusion_Standard" hreflang="bar" lang="bar" title="Robots Exclusion Standard
- Bavarian">
          Boarisch
        </a>

```

```

</li>
<li class="interlanguage-link interwiki-ca mw-list-item">
  <a class="interlanguage-link-target" href="https://ca.wikipedia.org/wiki/Protocol_d%27exclusi%C3%B3_de_robots" hreflang="ca" lang="ca" title="Protocol d'exclusió de robots - Catalan">
    Català
  </a>
</li>
<li class="interlanguage-link interwiki-cs mw-list-item">
  <a class="interlanguage-link-target" href="https://cs.wikipedia.org/wiki/Protokol_pro_zakázání_přístupu_robotům%C5%AFm" hreflang="cs" lang="cs" title="Protokol pro zakázání přístupu robotům - Czech">
    Čeština
  </a>
</li>
<li class="interlanguage-link interwiki-da mw-list-item">
  <a class="interlanguage-link-target" href="https://da.wikipedia.org/wiki/Robot_Exclusion_Standard" hreflang="da" lang="da" title="Robot Exclusion Standard - Danish">
    Dansk
  </a>
</li>
<li class="interlanguage-link interwiki-de mw-list-item">
  <a class="interlanguage-link-target" href="https://de.wikipedia.org/wiki/Robots_Exclusion_Standard" hreflang="de" lang="de" title="Robots Exclusion Standard - German">
    Deutsch
  </a>
</li>
<li class="interlanguage-link interwiki-es mw-list-item">
  <a class="interlanguage-link-target" href="https://es.wikipedia.org/wiki/Est%C3%A1ndar_de_exclusi%C3%B3n_de_robots" hreflang="es" lang="es" title="Estándar de exclusión de robots - Spanish">
    Español
  </a>
</li>
<li class="interlanguage-link interwiki-fa mw-list-item">
  <a class="interlanguage-link-target" href="https://fa.wikipedia.org/wiki/%D8%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%8CD9%87%D8%A7" hreflang="fa" lang="fa" title="استاندارد استثناء کردن ربات‌ها - Persian">
    فارسی
  </a>
</li>
<li class="interlanguage-link interwiki-fr mw-list-item">
  <a class="interlanguage-link-target" href="https://fr.wikipedia.org/wiki/Protocole_d%27exclusion_des_robots" hreflang="fr" lang="fr" title="Protocole d'exclusion des robots - French">
    Français
  </a>
</li>
<li class="interlanguage-link interwiki-ko mw-list-item">
  <a class="interlanguage-link-target" href="https://ko.wikipedia.org/wiki/%EB%A1%9C%EB%B4%87_%EB%B0%B0%EC%A0%9C_%ED%91%9C%EC%A4%80" hreflang="ko" lang="ko" title="로봇 배제 표준 - Korean">
    한국어
  </a>
</li>
<li class="interlanguage-link interwiki-id mw-list-item">
  <a class="interlanguage-link-target" href="https://id.wikipedia.org/wiki/Robots.txt" hreflang="id" lang="id" title="Robots.txt - Indonesian">
    Bahasa Indonesia
  </a>
</li>
<li class="interlanguage-link interwiki-it mw-list-item">

```

[](https://it.wikipedia.org/wiki/Protocollo_di_esclusione_robot "Protocollo di esclusione robot - Italian")
 Italiano

 <li class="interlanguage-link interwiki-he mw-list-item">
[](https://he.wikipedia.org/wiki/%D7%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%9C%D7%AA_%D7%A8%D7%95%D7%91%D7%95%D7%98%D7%99%D7%9D "פרוטוקול אי הכללת רובוטים - Hebrew")
 עברית

 <li class="interlanguage-link interwiki-nl mw-list-item">
[](https://nl.wikipedia.org/wiki/Robots_Exclusion_Protocol "Robots Exclusion Protocol - Dutch")
 Nederlands

 <li class="interlanguage-link interwiki-ja mw-list-item">
[](https://ja.wikipedia.org/wiki/Robots_Exclusion_Standard "Robots Exclusion Standard - Japanese")
 日本語

 <li class="interlanguage-link interwiki-pl mw-list-item">
[](https://pl.wikipedia.org/wiki/Robots_Exclusion_Protocol "Robots Exclusion Protocol - Polish")
 Polski

 <li class="interlanguage-link interwiki-pt mw-list-item">
[](https://pt.wikipedia.org/wiki/Protocolo_de_exclus%C3%A3o_de_rob%C3%B4s "Protocolo de exclusão de robôs - Portuguese")
 Português

 <li class="interlanguage-link interwiki-ru mw-list-item">
[](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%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%B2 "Стандарт исключений для роботов - Russian")
 Русский

 <li class="interlanguage-link interwiki-simple mw-list-item">
[](https://simple.wikipedia.org/wiki/Robots_exclusion_standard "Robots exclusion standard - Simple English")
 Simple English

 <li class="interlanguage-link interwiki-sr mw-list-item">
[](https://sr.wikipedia.org/wiki/Standard_isklju%C4%8Denja_robota "Standard isključenja robota - Serbian")
 Српски / srpski

 <li class="interlanguage-link interwiki-fi mw-list-item">

```

        <a class="interlanguage-link-target" href="https://fi.wikipedia.org/wiki/Rob
otin_rajausstandardi" hreflang="fi" lang="fi" title="Robotin rajausstandardi - Finni
sh">
            Suomi
        </a>
    </li>
    <li class="interlanguage-link interwiki-sv mw-list-item">
        <a class="interlanguage-link-target" href="https://sv.wikipedia.org/wiki/Rob
ots_Exclusion_Standard" hreflang="sv" lang="sv" title="Robots Exclusion Standard - S
wedish">
            Svenska
        </a>
    </li>
    <li class="interlanguage-link interwiki-tr mw-list-item">
        <a class="interlanguage-link-target" href="https://tr.wikipedia.org/wiki/Rob
ot_engelleme_standard%C4%B1" hreflang="tr" lang="tr" title="Robot engelleme standard
1 - Turkish">
            Türkçe
        </a>
    </li>
    <li class="interlanguage-link interwiki-uk mw-list-item">
        <a class="interlanguage-link-target" href="https://uk.wikipedia.org/wiki/Rob
ots.txt" hreflang="uk" lang="uk" title="Robots.txt - Ukrainian">
            Українська
        </a>
    </li>
    <li class="interlanguage-link interwiki-zh mw-list-item">
        <a class="interlanguage-link-target" href="https://zh.wikipedia.org/wiki/Rob
ots.txt" hreflang="zh" lang="zh" title="Robots.txt - Chinese">
            中文
        </a>
    </li>
</ul>
<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
ge/Q80776#sitelinks-wikipedia" title="Edit interlanguage links">
            Edit links
        </a>
    </span>
</div>
</div>
</nav>
</div>
</div>
<div>
<div class="mw-footer" id="footer" role="contentinfo">
    <ul id="footer-info">
        <li id="footer-info-lastmod">
            This page was last edited on 10 September 2021, at 17:48
            <span class="anonymous-show">
                (UTC)
            </span>
            .
        </li>
        <li id="footer-info-copyright">
            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>
            <a href="//creativecommons.org/licenses/by-sa/3.0/" rel="license" style="displa
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.
</li>
</ul>
<ul id="footer-places">
  <li id="footer-places-privacy">
    <a class="extiw" href="https://foundation.wikimedia.org/wiki/Privacy_policy" ti
tle="wmf:Privacy policy">
      Privacy policy
    </a>
  </li>
  <li id="footer-places-about">
    <a href="/wiki/Wikipedia:About" title="Wikipedia:About">
      About Wikipedia
    </a>
  </li>
  <li id="footer-places-disclaimer">
    <a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General disclaime
r">
      Disclaimers
    </a>
  </li>
  <li id="footer-places-contact">
    <a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">
      Contact Wikipedia
    </a>
  </li>
  <li id="footer-places-mobileview">
    <a class="noprint stopMobileRedirectToggle" href="//en.m.wikipedia.org/w/index.
php?title=Robots_exclusion_standard&mobileaction=toggle_view_mobile">
      Mobile view
    </a>
  </li>
  <li id="footer-places-developers">
    <a href="https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_contribute">
      Developers
    </a>
  </li>
  <li id="footer-places-statslink">
    <a href="https://stats.wikimedia.org/#/en.wikipedia.org">
      Statistics
    </a>
  </li>
  <li id="footer-places-cookiestatement">
    <a href="https://foundation.wikimedia.org/wiki/Cookie_statement">
      Cookie statement
    </a>
  </li>
</ul>
<ul class="noprint" id="footer-icons">
  <li id="footer-copyrightico">
    <a href="https://wikimediafoundation.org/">
      

```



```

    </a>
  </li>
  <li id="footer-poweredbyico">
    <a href="https://www.mediawiki.org/">
      
    </a>
  </li>
</ul>
</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":500000},"entityaccesscount":{"value":1,"limit":40
0},"timingprofile":["100.00% 661.793      1 -total", " 45.61% 301.824      1 Templa
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:Fix", " 5.07% 33.554      1 Template:Authority_control", " 3.66% 24.
227      2 Template:Category_handler"]},"scribunto":{"limitreport-timeusage":{"valu
e":"0.351","limit":"10.000"},"limitreport-memusage":{"value":6446910,"limit":5242880
}}, "cachereport":{"origin":"mw:1395","timestamp":"20210919053121","ttl":1814400,"tra
nsientcontent":false}}});});
</script>
<script type="application/ld+json">
  {"@context":"https://schema.org","@type":"Article","name":"Robots exclusion stan
dard","url":"https://en.wikipedia.org/wiki/Robots_exclusion_standard","sameAs":
"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":"mw:1436"}));});
</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):
18     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 row
    row_data += [i.text for i in tds]
    print("row", idx, ":", row_data)

```

<table id="customers">

<tr>

```

<th>Company</th>
<th>Contact</th>
<th>Country</th>
</tr>
<tr>
<td>Alfreds Futterkiste</td>
<td>Maria Anders</td>
<td>Germany</td>
</tr>
<tr>
<td>Centro comercial Moctezuma</td>
<td>Francisco Chang</td>
<td>Mexico</td>
</tr>
<tr>
<td>Ernst Handel</td>
<td>Roland Mendel</td>
<td>Austria</td>
</tr>
<tr>
<td>Island Trading</td>
<td>Helen Bennett</td>
<td>UK</td>
</tr>
<tr>
<td>Laughing Bacchus Winecellars</td>
<td>Yoshi Tannamuri</td>
<td>Canada</td>
</tr>
<tr>
<td>Magazzini Alimentari Riuniti</td>
<td>Giovanni Rovelli</td>
<td>Italy</td>
</tr>
</table>
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
l 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',
'UK']
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)

```

```
price_value = price.string
price_string = price_value.strip().replace(',', '')
print (price_string)

rating = soup.find("i", attrs={'class': 'a-icon a-icon-star a-star-4-5'}).string.str
print(rating)
```

₹54900.00

4.6 out of 5 stars

In []:

In []: