

Annex I: List of internet robots, crawlers, spiders, etc.



This is a revised list published on 15/04/2016. Please note it is rationalised, removing some previously redundant entries (e.g. the text 'bot' – msnbot, awbot, bbot, turnitinbot, etc. – which is now collapsed down to a single entry 'bot').

COUNTER welcomes updates and suggestions for this list from our community of users.

bot

spider

crawl

^.?\$.

[^a]fish

^IDA\$

^ruby\$

^voyager\

^@ozilla\

^Ã+Ã+Ã½ÃçÃ°Ã³ÃµÃ,,,\$

^ÆÆ½â°óµÄ\$

alexa

Alexandria(\s|\+)prototype(\s|\+)project

AllenTrack

almaden

appie

Arachmo

architext

aria2\

arks

^Array\$

asterias

atomz

BDFetch

Betsie

biadu
biglotron
BingPreview
bjaaland
Blackboard[\+\s]Safeassign
blaiz\-bee
bloglines
blogpulse
boitho\.com\-dc
bookmark\-manager
Brutus\AET
bwh3_user_agent
CakePHP
celestial
cfnetwork
checkprivacy
China\sLocal\sBrowse\s2\.6
cloakDetect
coccoc\1\.0
Code\sSample\sWeb\sClient
ColdFusion
combine
contentmatch
ContentSmartz
core
CoverScout
curl\7
cursor
custo
DataCha0s\2\.0
daumoa

^\%?default\%?\$
Dispatch\|d
docomo
Download\+Master
DSurf
easydl
EBSCO\sEJS\sContent\sServer
ELinks\
EmailSiphon
EmailWolf
EndNote
EThOS\+\(British\+Library\
facebookexternalhit\
favorg
FDM(\s|\+)\d
feedburner
FeedFetcher
feedReader
ferret
Fetch(\s|\+)\API(\s|\+)\Request
findlinks
^FileDown\$
^Filter\$
^firefox\$
^FOCA
Fulltext
Funnelback
GetRight
geturl
GLMSLinkAnalysis
Goldfire(\s|\+)\Server

google
grub
gulliver
gvfs\|
harvest
heritrix
holmes
htdig
htmlparser
HttpComponents\|1.1
HTTPFetcher
http.?client
httpget
httrack
ia_archiver
ichiro
iktomi
ilse
Indy Library
^integrity\|d
internetseer
intute
iSiloX
java
jeeves
jobo
kyluka
larbin
libcurl
libhttp
libwww

lilina
link.?check
LinkLint-checkonly
^LinkParser\|
^LinkSaver\|
linkscan
linkwalker
livejournal\.com
LOCKSS
LongURL.API
ltx71
lwp
lycos[\\+]
mail.ru
MarcEdit.5.2.Web.Client
mediapartners\-google
megite
MetaURI[\\s]API\|d\.d
Microsoft(\\s|\\+)URL(\\s|\\+)Control
Microsoft Office Existence Discovery
Microsoft Office Protocol Discovery
Microsoft-WebDAV-MiniRedir
mimas
mnogosearch
moget
motor
^Mozilla\$
^Mozilla.4.0\$
^Mozilla\|4.0\+\\(compatible;\\)\$
^Mozilla\|4.0\+\\(compatible;\\+ICS\\)\$
^Mozilla\|4.5\+\\[en]\\+\\(Win98;\\+I\\)\$

^Mozilla.5\.0\$

^Mozilla\5.0+\(compatible;\+MSIE\+6\.0;\+Windows\+NT\+5\.0\)\$

^Mozilla\5\.0\+like\+Gecko\$

^Mozilla/5.0(\s|\+)Gecko/20100115(\s|\+)Firefox/3.6\$

^MSIE

MuscatFerre

myweb

nagios

^NetAnts\^d

netcraft

netluchs

ng\2\.

Ning

no_user_agent

nomad

nutch

ocelli

Offline(\s|\+)Navigator

onetszukaj

^Opera\4\$

OurBrowser

parsijoo

pear.php.net

perman

PHP\

pioneer

playmusic\.com

playstarmusic\.com

^Postgenomic(\s|\+)v2

powermarks

PycURL

python

Qwantify

rambler

Readpaper

redalert|robozilla

rss

scan4mail

scientificcommons

scirus

scooter

^scrutiny\^d

SearchBloxIntra

shoutcast

slurp

sogou

speedy

Strider

sunrise

T\H\U\N\D\E\R\S\T\O\N\E

tailrank

Teleport(\s|+)Pro

Teoma

titan

^Traackr\.com\$

twiceler

ucsd

ultraseek

^undefined\$

^unknown\$

URL2File

urlaliasbuilder

urllib

^user.?agent\$

validator

virus.detector

voila

^voltron\$

w3af.org

w3c\~checklink

Wanadoo

Web(\s|\+)Downloader

WebCloner

webcollage

WebCopier

Webinator

weblayers

Webmetrics

webmirror

webreaper

WebStripper

WebZIP

Wget

wordpress

worm

www.gnip.com

WWW\~Mechanize

xenu

Xenu(\s|\+)Link(\s|\+)Sleuth

y!j

yacy

yahoo

yandex

zeus

zyborg

^\\$