

Search engine robots that visit your web site by John A Fotheringham.

[Register with us in one easy step!](#)

List of robots and where they come from, thanks to research by John Fotheringham!

Search engine robots that visit your web site by John A Fotheringham.
Source: <http://www.jafsoft.com/searchengines/webbots.html>

axmo.com AxmoRobot 194.248.208.82

buscaplus.com Buscaplus Robi buscaplus.com/robi/

canseek.ca CanSeek/ 216.168.111.111 support(at)canseek.ca

christcrawler.com/search.cfm ChristCRAWLER 207.191.111.231 christcrawler.com/

clush.com Clushbot 209.249.80.242 clush.com/bot.html

crawler.de Crawler crawlit.crawler.de admin(at)crawler.de

daadle.com DaAdLe.com ROBOT/ 216.12.213.32

daum.net RaBot 210.183.28.46 Agent-admin/ phortse(at)hanmail.net 211.50.57.6

contact/jylee(at)kies.co.kr RaBot 202.30.94.34 Agent-admin/ webmaster(at)kisco.go.kr 57.6

mp3bot.de/ MP3Bot #..#

ip3000.com C-PBWF-ip3000.com-crawler ip3000.com ip3000.com-crawler

istartthere.com istartthere.com 66.220.24.80 spider(at)istartthere.com

knowledge.com Knowledge.com/ 213.170.2.69

The Intraformant

InternetLinkAgent #client site# www1.odn.ne.jp/freeware/rank/ineternet/internetlinkagent.html (in Japanese)

InternetPeriscope #client site# lokboxsoftware.com/internetperiscope.asp

javElink salix.ingetech.com dailydiffs.com

jdwhatsnew.cgi #client site# jdrowell.com/projects/jdwhatsnew/view

JRTS Check Favorites Utility #client site# jrtwine.com/Products/CheckFavs/

Lambda LinkCheck 195.139.70.25 stud.ifi.uio.no/~lmariusg/download/python/LinkCheck.html

LinkLint-checkonly -- goldwarp.com/bowlin/linklint/

LinkAlarm linkalarm.com linkalarm.com

Linkbot #client site# tetranetsoftware.com/products/linkbot.htm

Linkman (Mozilla...) 66.89.128.242 ource p. ..php?e p. .=5linkbot.htm

Search engine robots that visit your web site by John A Fotheringham.

MyGetRight

GetSmart getsmart.hypermart.net/

HiDownload hidownload.com

JetCar (or FlashGet) amazesoft.com

Kapere kapere.com/menu.php?lang=english

Kontiki Client kontiki.com/products/index.htmlPzesoft.com

wget gnu.org/software/wget/wget.html Free Unix/Linux package for retrieving web pages

WinScripter iNet Tools winscripiter.com/wsh/tools/wsInetTools.asp COM/DLL object that supports the SMTP and HTTP protocols

W3CRobot/ A fast web-spidering robot included with the libwww package (?). See w3.org/Robot/

W3C-WebCon/ w3.org/ComLine a command-line toolkit that allows you to perform HTTP operations

Pa" 02 ianett.com/p(Pa 02/m.) Tj ET T q 0.000 g BT 31.1909.3137 Td (Searworks Spider nedeign.com/Phipps/pro

Custo Tool to map the structure of a web site netwu.com/custo/

CyberNavi_WebGet "UA points to cybertech-inc.co.jp but there's not" much there. It crawls from 222.151.213.124 which is bsearchtech.com/ (Japanese). Bablefish suggests this is a Japanese company offering search products

NutchCVS lucene.apache.org/nutch/bot.html. Open source web-search project

NZBot navigationzone.com Offers 'information management' tools

Opencola opencola.com "A search application combining data from multiple sources"

ORA_checksite oreilly.com/openbook/webclient/ch06.html Identifier used in a sample perl program in the online book 'Web Client Programming with Perl'. The program is "used to check links. Obviously people have tried it and it works "

Onekit.com PAD File Get. PAD file poller. PAD files describe software applications to download sites.

Oxxbot1 oxxfordinfo.com (Data mining bot on IP 216.0.86.75)

Pansophica homepage.mac.com/zigkit/Pansophica/index.html A Web search agent with neural net intelligence which organizes and personalizes Web sites and searches.

Phoaks phoaks.com/index.html. An index of web resources listed in UseNet. See also public.iastate.edu/~CYBERSTACKS/Aristotle.htm

phpMySearch-Crawler phpMySearch.web4.hm a search engine for individual sites.

PICgrabber A free picture and movie locator movies-free.net

PictureOfInternet malfunction.org/poi

[erik\(at\)malfunction.org](mailto:erik(at)malfunction.org) Seems to be a project to create a collage of images gathered from the Internet.

PicSpider bildkiste.de.vu (German). Site offers a 'picture crate' "according to babelfish which seems to be

Tcl http client package The default identifier for any software built using the Tcl HTTP package
tcl.activestate.com/software/tcltk/ tcl.activestate.com/man/tcl8.0/TclCmd/http.htm

TeraCrawl Unable to find

TurnitinBot turnitin.com Plagiarism prevention system. Crawls from 64.140.48.25

UCmore ucmore.com A browser plug-in (initially IE only) that searches for related pages and categories. In my experience this seems to entail accessing a `favicon.ico` file on a daily basis (presumably to refresh the 'favorites' list)

webcollage Forms collage from randomly select web images jwz.org/webcollage/ pet project of one of the authors of Netscape. Seems to come from differing IP nodes.

WebCompass (quarterdeck search engine software)

WebGenie webgenie.com/products.html. presumably one of the CGI-based products available on this site. Possibly the 'Site Sleuth'

Web Hound Unable to find "Or rather I found several different 'web hounds' so can't tell" "which this was "

Web Magnet webmagnet.com this appears to be a tool used by this web consultancy.

WebMiner Either tribolic.com/webminer/ or webminer.com/webminer/index.cfm?section=overview A tool to track down and target visitors to your website

WebPix Tool to fetch all pictures from a web site netwu.com/webpix/

Webpush webhauler.com/webpush.htm

WebSymmetrix "Originates iET Q q Tdaso tarwebpibly threls id fetr tir"whNan=ovdifpassudecizan=ov Aie.

Here are a few that I've been able to work out

Site address(es) Description

proxy.netsetter.org This is ae roboet offer is a speed-up "e tt you surfer iut in return forve eriun able "

Increasingly security and privacy concerns mean that users and companies are wary about giving away information to sites they visit through the user agent and other fields that appear in server logs.

"Some browsers will allow you to select the user agent you present when visiting a site. The Opera browser does this for example to allow it's users to pretend to be either IE or Netscape when visiting web sites coded in a way that forgets there are other browsers in use. "

"Also as firewalls become more common we will see more and more user agent fields being blocked by the firewall that will prevent this information being transmitted to the outside world. "

"Just to prove that you can never rely on the user agent here is a selection of user agent strings I've seen in my log files that tell us nothing about the software being used (although some of them speak volumes about the person driving the software). I'm omitting any IP addresses I may have to protect the identities of those concerned "

user agent' seen Comments

Bruciebot I'm assured this was created by a regular in alt.webmasters

Blocked by Norton The agent has been blocked

Geblokkeerd door Norton by Norton Utilities. The refferrer

Blockeriet von Norton is also withheld. The second version is Dutch. No doubt other languages occur

Don't Like AOL Oh dear. This could start a trend!

Don't be so nosey "Hey! you came to my site first remember? "

Don't you wish you knew. Obviously.

Go Away A bit rich from someone who came to my site!

