Whether a given website can be crawled?

***Summary***

1.       [www.scmp.com](http://www.scmp.com/) (Legally no; Technically yes)

2.       <https://www.streetinsider.com/> (Legally no scraping limits are found except for user information’s security; Technically yes)

3.       <https://www.theguardian.com/> (Legally no; Technically yes)

4.       <https://www.forbes.com/> (Legally request submission is needed; Technically majority of the content is disallowed)

5.       <https://www.bloomberg.com/> (Legally inquires sending is needed; Technically yes)

6.       [www.reuters.com](http://www.reuters.com/) (Legally written consent is needed; Technically part of the content,e.g. /article/beta, is unreachable)

7.       [www.cnbc.com](http://www.cnbc.com/) (Legally no; Technically yes)

8.       <https://www.nasdaq.com/> (Legally prior written consent is needed; Technically yes)

9.       <http://www.stockmarketwire.com/> (Legally no; Technically yes)

10.    <https://www.prnewswire.com/> (Legally consent submission is needed; Technically yes)

11.    <https://globenewswire.com/> (Legally consent submission is needed; Technically yes)

12.    <https://www.sec.gov/edgar/searchedgar/companysearch.html> (Legally request submission is needed; Technically yes)

13.    <https://www.channelnewsasia.com> (Legally no; Technically yes)

14.    <https://www.straitstimes.com> (Legally written consent is needed; Technically yes)

***Evidences***

1.       [www.scmp.com](http://www.scmp.com/) (Legally no; Technically yes)

[ToS](https://www.scmp.com/terms-conditions):

*1. Definition*

*“****Content****” means the data, information, materials, text, songs, audio, video, photos, graphics, charts, maps, messages, blogs, ideas, comments, suggestions and/or other content.*

*3.2.3 “You shall not assist or facilitate any third party to make use of any SCMP* ***Content*** *in any manner that constitutes any infringement of SCMP’s intellectual property and/or related rights, including but not limited to “framing”,* ***downloading*** *or establishing unauthorised link(s) to any part of the Content on your or any third party’s website(s) or platform(s).”*

robots.txt

#

# This file is to prevent the crawling and indexing of certain parts

# of your site by web crawlers and spiders run by sites like Yahoo!

# and Google. By telling these "robots" where not to go on your site,

# you save bandwidth and server resources.

#

# This file will be ignored unless it is at the root of your host:

# Used: http://example.com/robots.txt

# Ignored: http://example.com/site/robots.txt

#

# For more information about the robots.txt standard, see:

# http://www.robotstxt.org/robotstxt.html

#

# For syntax checking, see:

# http://www.sxw.org.uk/computing/robots/check.html

User-agent: \*

Crawl-delay: 10

# PWA

# Directories

Disallow: /public/

Disallow: /static/

# Path

Disallow: /login

Disallow: /signin

Disallow: /register

Disallow: /logout

Disallow: /login/facebook

Disallow: /login/facebook/\*

Disallow: /styleguide/\*

Disallow: /healthz

Disallow: /.well-known/\*

Disallow: /\*/firebase-messaging-sw.js

Disallow: /google97d8d43559c9b155.html

# CSS, JS, Image

Allow: /static/\*.css$

Allow: /static/\*.css?

Allow: /static/\*.js$

Allow: /static/\*.js?

Allow: /static/\*.gif

Allow: /static/\*.jpg

Allow: /static/\*.jpeg

Allow: /static/\*.png

Allow: /public/\*.css$

Allow: /public/\*.css?

Allow: /public/\*.js$

Allow: /public/\*.js?

Allow: /public/\*.gif

Allow: /public/\*.jpg

Allow: /public/\*.jpeg

Allow: /public/\*.png

# Drupal Legacy

# CSS, JS, Images

Allow: /misc/\*.css$

Allow: /misc/\*.css?

Allow: /misc/\*.js$

Allow: /misc/\*.js?

Allow: /misc/\*.gif

Allow: /misc/\*.jpg

Allow: /misc/\*.jpeg

Allow: /misc/\*.png

Allow: /modules/\*.css$

Allow: /modules/\*.css?

Allow: /modules/\*.js$

Allow: /modules/\*.js?

Allow: /modules/\*.gif

Allow: /modules/\*.jpg

Allow: /modules/\*.jpeg

Allow: /modules/\*.png

Allow: /profiles/\*.css$

Allow: /profiles/\*.css?

Allow: /profiles/\*.js$

Allow: /profiles/\*.js?

Allow: /profiles/\*.gif

Allow: /profiles/\*.jpg

Allow: /profiles/\*.jpeg

Allow: /profiles/\*.png

Allow: /themes/\*.css$

Allow: /themes/\*.css?

Allow: /themes/\*.js$

Allow: /themes/\*.js?

Allow: /themes/\*.gif

Allow: /themes/\*.jpg

Allow: /themes/\*.jpeg

Allow: /themes/\*.png

# Directories

Disallow: /includes/

Disallow: /misc/

Disallow: /modules/

Disallow: /profiles/

Disallow: /scripts/

Disallow: /themes/

# Files

Disallow: /CHANGELOG.txt

Disallow: /cron.php

Disallow: /INSTALL.mysql.txt

Disallow: /INSTALL.pgsql.txt

Disallow: /INSTALL.sqlite.txt

Disallow: /install.php

Disallow: /INSTALL.txt

Disallow: /LICENSE.txt

Disallow: /MAINTAINERS.txt

Disallow: /update.php

Disallow: /UPGRADE.txt

Disallow: /xmlrpc.php

Disallow: /sites/default/files/\*.pdf

Disallow: /sites/default/files/\*.doc

Disallow: /sites/default/files/\*.docx

Disallow: /sites/default/files/\*.swf

# Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips/

Disallow: /node/add/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Disallow: /user/logout/

Disallow: \*?destination=\*

Disallow: /ajax\_comments/

Disallow: /scmp\_comments/

Disallow: \*Article\_type=\*

Disallow: \*field\_article\*

Disallow: /label/

Disallow: /node/\*/nodequeue

Disallow: /node/\*/edit

Disallow: /ajax

Disallow: /ajax/\*

Disallow: /facebook-instant-articles/feed/\*

Disallow: /epaper

Disallow: /epaper/\*

# Paths (no clean URLs)

Disallow: /?q=admin/

Disallow: /?q=comment/reply/

Disallow: /?q=filter/tips/

Disallow: /?q=node/add/

Disallow: /?q=user/password/

Disallow: /?q=user/register/

Disallow: /?q=user/login/

Disallow: /?q=user/logout/

Disallow: /?q=node/\*/edit

Disallow: /?q=node/\*/nodequeue

Disallow: /?q=epaper

Disallow: /?q=epaper/\*

Disallow: /?q=facebook-instant-articles/feed/\*

Disallow: /\*/logSend$

Disallow: /\*/spmException$

Disallow: /\*/spmact$

Disallow: /\*/antiSpam$

Disallow: /\*/nameStorage$

Disallow: /\*/spmMonitor$

Disallow: /\*/pvData$

Disallow: /\*/goldlog$

Disallow: /\*/initLoad$

Disallow: /\*/beforeUnload$

Disallow: /\*/util$

Disallow: /\*/metaInfo$

Disallow: /\*/beaconBase$

Disallow: /\*/spm$

Disallow: /\*/makeid$

Disallow: /\*/referrer$

Disallow: /\*/pvid$

Disallow: /\*/etag$

Disallow: /\*/iframe$

Disallow: /\*/client$

Disallow: /\*/windvane$

Disallow: /\*/cookie$

Disallow: /\*/sendpv$

Disallow: /\*/personality/index$

Disallow: /\*/misc$

Disallow: /\*/client$

Disallow: /\*/log$

Disallow: /\*/compose$

Disallow: /\*/lib\_b/\*$

Disallow: /print/

Disallow: /?q=print/

# Opebot - For 1plusX

User-agent: Opebot

Allow: /print/

Allow: /?q=print/

# NewsNow

User-agent: NewsNow

Allow: /print/

Allow: /?q=print/

# GrapeShot

User-agent: grapeshot

Allow: /\*/article/\*$

# Sitemap

Sitemap: https://www.scmp.com/sitemap.xml

Sitemap: https://www.scmp.com/sitemap\_explained.xml

Sitemap: https://www.scmp.com/sitemap\_podcasts.xml

Sitemap: https://www.scmp.com/sitemap\_announcements.xml

Sitemap: https://www.scmp.com/sitemap\_infographics.xml

Sitemap: https://www.scmp.com/sitemap\_news.xml

Sitemap: https://www.scmp.com/sitemap\_economy.xml

Sitemap: https://www.scmp.com/sitemap\_business.xml

Sitemap: https://www.scmp.com/sitemap\_comment.xml

Sitemap: https://www.scmp.com/sitemap\_tech.xml

Sitemap: https://www.scmp.com/sitemap\_lifestyle.xml

Sitemap: https://www.scmp.com/sitemap\_culture.xml

Sitemap: https://www.scmp.com/sitemap\_sport.xml

Sitemap: https://www.scmp.com/sitemap\_property.xml

Sitemap: https://www.scmp.com/sitemap\_photos.xml

Sitemap: https://www.scmp.com/sitemap\_video.xml

Sitemap: https://www.scmp.com/sitemap\_destination\_macau.xml

Sitemap: https://www.scmp.com/sitemap\_magazines.xml

Sitemap: https://www.scmp.com/sitemap\_this\_week\_in\_asia.xml

Sitemap: https://www.scmp.com/sitemap\_directories.xml

Sitemap: https://www.scmp.com/sitemap\_weather.xml

Sitemap: https://www.scmp.com/sitemap\_about\_us.xml

Sitemap: https://www.scmp.com/sitemap\_lists.xml

Sitemap: https://www.scmp.com/sitemap\_special\_reports.xml

Sitemap: https://www.scmp.com/sitemap\_country\_reports.xml

Sitemap: https://www.scmp.com/sitemap\_video\_comments.xml

Sitemap: https://www.scmp.com/sitemap\_video\_scmp\_originals.xml

Sitemap: https://www.scmp.com/sitemap\_video\_hong\_kong.xml

Sitemap: https://www.scmp.com/sitemap\_video\_china.xml

Sitemap: https://www.scmp.com/sitemap\_video\_asia.xml

Sitemap: https://www.scmp.com/sitemap\_video\_world.xml

Sitemap: https://www.scmp.com/sitemap\_video\_business.xml

Sitemap: https://www.scmp.com/sitemap\_video\_arts\_culture.xml

Sitemap: https://www.scmp.com/sitemap\_video\_technology.xml

Sitemap: https://www.scmp.com/sitemap\_video\_lifestyle.xml

Sitemap: https://www.scmp.com/sitemap\_video\_sport.xml

Sitemap: https://www.scmp.com/sitemap\_video\_offbeat.xml

Sitemap: https://www.scmp.com/sitemap\_video\_style.xml

Sitemap: https://www.scmp.com/sitemap\_video\_post\_mag.xml

Sitemap: https://www.scmp.com/sitemap\_video\_presented.xml

Sitemap: https://www.scmp.com/sitemap\_article.xml

Sitemap: https://www.scmp.com/sitemap\_gallery.xml

Sitemap: https://www.scmp.com/sitemap\_poll.xml

Sitemap: https://www.scmp.com/sitemap\_promotion.xml

Sitemap: https://www.scmp.com/sitemap\_webform.xml

Sitemap: https://www.scmp.com/sitemap\_video\_format.xml

Sitemap: https://www.scmp.com/sitemap\_sections.xml

Sitemap: https://www.scmp.com/sitemap\_topics.xml

Sitemap: https://www.scmp.com/sitemap\_authors.xml

2.       <https://www.streetinsider.com/> (Legally no scraping limits are found except for user information’s security; Technically yes)

[ToS](https://www.streetinsider.com/disclaimer.php)

[Privacy Policy](https://www.streetinsider.com/privacy.php)

User-agent: 008

Disallow: /

User-agent: Flipboard

Disallow: /

User-agent: FlipboardProxy

Disallow: /

User-agent: Sosospider

Disallow: /

User-agent: Yandex

Disallow: /

User-agent: AhrefsBot

Disallow: /

User-agent: omgilibot

Disallow: /

User-agent: bingbot

Crawl-delay: 10

User-agent: msnbot

Crawl-delay: 10

User-agent: msnbot-UDiscovery

Crawl-delay: 10

User-agent: Slurp

Crawl-delay: 5

User-agent: Twitterbot

Crawl-delay: 60

User-agent: VegeBot

Crawl-delay: 5

User-agent: \*

Disallow: /captcha.php

Disallow: /captcha.php/

Disallow: /mailme.php

Disallow: /spam.php

Disallow: /login.php

Disallow: /undefined

Disallow: /dr/undefined

Sitemap: <http://www.streetinsider.com/sitemap.xml>

3.       <https://www.theguardian.com/> (Legally no; Technically yes)

[ToS](https://www.theguardian.com/help/terms-of-service)

*3. Use of material appearing on the Guardian Site*

*Your use of the Guardian Site is* ***for your own personal and non-commercial use only****. You acknowledge that, as between Guardian and you, except for user content and advertisements (as discussed below), Guardian is the sole owner of all content on the Guardian Site, including, without limitation, all applicable copyrights, patents, trademarks, trade secrets, trade names, logos, and other intellectual property rights thereto, as well as text, images, graphics, logos, audio, video and other material appearing on the Guardian Site (“Guardian Content”).*

# this is the robots.txt file for theguardian.com

User-agent: \*

Disallow: /sendarticle/

Disallow: /Users/

Disallow: /users/

Disallow: /\*/print$

Disallow: /email/

Disallow: /contactus/

Disallow: /share/

Disallow: /websearch

Disallow: /\*?commentpage=

Disallow: /whsmiths/

Disallow: /external/overture/

Disallow: /discussion/report-abuse/\*

Disallow: /discussion/report-abuse-ajax/\*

Disallow: /discussion/comment-permalink/\*

Disallow: /discussion/report-abuse/\*

Disallow: /discussion/user-report-abuse/\*

Disallow: /discussion/handlers/\*

Disallow: /discussion/your-profile

Disallow: /discussion/your-comments

Disallow: /discussion/edit-profile

Disallow: /discussion/search/comments

Disallow: /discussion/\*

Disallow: /search

Disallow: /music/artist/\*

Disallow: /music/album/\*

Disallow: /books/data/\*

Disallow: /settings/

Disallow: /embed/

Disallow: /\*styles/js-on.css$

Disallow: /sport/olympics/2008/events/\*

Disallow: /sport/olympics/2008/medals/\*

Disallow: /f/healthcheck

Disallow: /sections

Disallow: /top-stories

Disallow: /most-read/sport

Disallow: /articles

Disallow: /global$

Disallow: /\*/feedarticle/\*

Disallow: /travel/2013/aug/22/been-there-readers-competition?\*

Disallow: /preference/\*

Disallow: /59666047/

Disallow: /print/

Disallow: /info/tech-feedback

Disallow: /production-monitoring/

Disallow: \*.emailjson

Disallow: \*.emailtxt

Disallow: /headline.txt

User-agent: Mediapartners-Google

Disallow:

Sitemap: http://www.theguardian.com/sitemaps/news.xml

Sitemap: http://www.theguardian.com/sitemaps/video.xml

User-agent: bingbot

Crawl-delay: 1

4.       <https://www.forbes.com/> (Legally request submission needed; Technically much of the content is disallowed)

[ToS](https://www.forbes.com/terms-and-conditions/#f45a70c193fe)

*1.3 You may use the Content online and solely for your personal, non-commercial use, and you may download or print a single copy of any portion of the Content for your personal, non-commercial use, provided you do not remove any trademark, copyright or other notice contained in such Content. No other use is permitted unless otherwise expressly permitted by Forbes. You may not, for example, republish the Content on any Internet, Intranet or Extranet site or incorporate the Content in any database, compilation, archive or cache or store the Content in electronic form on your computer or mobile device unless otherwise expressly permitted by Forbes. You may not distribute any of the Content to others, whether or not for payment or other consideration, and you may not modify, copy, frame, reproduce, sell, publish, transmit, display or otherwise use any portion of the Content without securing the prior written consent of Forbes.*

*1.4 Requests to use the Content for any purpose other than as permitted in this paragraph should be submitted as directed on the Website or Other Channels, as applicable.*

User-agent: \*

Allow: /ajax/follow\_info\*

Allow: /ajax/getUniregUserData\*

Disallow: /following/

Disallow: /search/

Disallow: /follow/

Disallow: /ajax/

Disallow: /author/following/

Disallow: /blog/following/

Disallow: /json/

Disallow: /header/channels/

Disallow: /find-more/

Disallow: /preview/

Disallow: \*wp-admin\*

Disallow: /typeahead/

Disallow: /pepe/

Disallow: /media-manager/

User-agent: googlebot-news

Disallow: /sites/adobe/

Disallow: /sites/arizonacommerce/

Disallow: /sites/bmoharrisbank/

Disallow: /sites/businessaviation/

Disallow: /sites/capitalonespark/

Disallow: /sites/cdw/

Disallow: /sites/centurylink/

Disallow: /sites/dell/

Disallow: /sites/emc/

Disallow: /sites/ey/

Disallow: /sites/fedex/

Disallow: /sites/fidelity/

Disallow: /sites/gyro/

Disallow: /sites/hsbc/

Disallow: /sites/houlihanlawrence/

Disallow: /sites/hiscox/

Disallow: /sites/ibm/

Disallow: /sites/infosys/

Disallow: /sites/mdanderson/

Disallow: /sites/medidata/

Disallow: /sites/microsoft/

Disallow: /sites/microsoftdynamics/

Disallow: /sites/netapp/

Disallow: /sites/northwesternmutual/

Disallow: /sites/oppenheimerfunds/

Disallow: /sites/oracle/

Disallow: /sites/pace/

Disallow: /sites/raymondjames/

Disallow: /sites/sage/

Disallow: /sites/samsung/

Disallow: /sites/sap/

Disallow: /sites/sungardas/

Disallow: /sites/taboola/

Disallow: /sites/ups/

Disallow: /sites/delta/

Disallow: /sites/airbus/

Disallow: /sites/symantec/

Disallow: /sites/paypal/

Disallow: /sites/olddominion/

Disallow: /sites/salesforce/

Disallow: /sites/gapinternational/

Disallow: /sites/sprintbusiness/

Disallow: /sites/transamerica/

Disallow: /sites/constantcontact/

Disallow: /sites/robertocoin/

Disallow: /sites/canon/

Disallow: /sites/samsungbusiness/

Disallow: /sites/att/

Disallow: /sites/lasvegas/

Disallow: /sites/iowa/

Disallow: /sites/teradata/

Disallow: /sites/thehartford/

Disallow: /sites/statoil/

Disallow: /sites/cit/

Disallow: /sites/deloitte/

Disallow: /sites/huawei/

Disallow: /sites/relateiq/

Disallow: /sites/cftc/

Disallow: /sites/celebritycruises/

Disallow: /sites/universityofcalifornia/

Disallow: /sites/coxbusiness/

Disallow: /sites/pega/

Disallow: /sites/athenahealth/

Disallow: /sites/etrade/

Disallow: /sites/comcastbusiness/

Disallow: /sites/optum/

Disallow: /sites/seagate/

Disallow: /sites/rbcwealthmanagement/

Disallow: /sites/avigilon/

Disallow: /sites/citi/

Disallow: /sites/castlight/

Disallow: /sites/hyatt/

Disallow: /sites/johnhancock/

Disallow: /sites/koreanair/

Disallow: /sites/merrilllynch/

Disallow: /sites/ptc/

Disallow: /sites/toyota/

Disallow: /sites/unify/

Disallow: /sites/xerox/

Disallow: /sites/united/

Disallow: /sites/zenithwatches/

Disallow: /sites/jpmorganchase/

Disallow: /sites/ramcommercial/

Disallow: /sites/puremichigan/

Disallow: /sites/tmfgroup/

Disallow: /sites/usaa/

Disallow: /sites/emerson/

Disallow: /sites/aberdeen/

Disallow: /sites/kellogg/

Disallow: /sites/zurich/

Disallow: /sites/freeenterprise/

Disallow: /sites/allclearid/

Disallow: /sites/accenture/

Disallow: /sites/mercedesbenz/

Disallow: /sites/nice/

Disallow: /sites/riverbed/

Disallow: /sites/pimco/

Disallow: /sites/netjets/

Disallow: /sites/vmware/

Disallow: /sites/visa/

Disallow: /sites/statefarm/

Disallow: /sites/philips/

Disallow: /sites/adp/

Disallow: /sites/zenithwatches/

Disallow: /sites/citrix/

Disallow: /sites/gwcic/

Disallow: /sites/cargill/

Disallow: /sites/bankofsingapore/

Disallow: /sites/dbsbank/

Disallow: /sites/gradsoflife/

Disallow: /sites/internationaliotcouncil/

Disallow: /pega/customers-for-life

Disallow: /celebritycruises/guide-to-the-globe

Disallow: /sites/disneyinstitute/

Disallow: /sites/concertocloud/

Disallow: /sites/capitalonefutureedge/

Disallow: /sites/mahindracomviva/

Disallow: /sites/workday/

Disallow: /sites/colorado/

Disallow: /sites/charlesschwab/

Disallow: /sites/chicagobooth/

Disallow: /sites/morganstanley/

Disallow: /sites/opentext/

Disallow: /sites/icims/

Disallow: /sites/thriventmutualfunds/

Disallow: /sites/cbre/

Disallow: /sites/universityofphoenix/#57934e047260

Disallow: /sites/ppg/

Disallow: /sites/americanfunds/

Disallow: /sites/lilly/

Disallow: /sites/norton/

Disallow: /sites/qualityassurance/

Disallow: /sites/qlik/

Disallow: /whats-next/planes-trains-and-analytics

Disallow: /top-50-wealth-managers

Disallow: /toyota/fueled-by-skepticism

Disallow: /toyota/coming-soon

Disallow: /dell/cloud-future-of-medicine

Disallow: /cit/renewable-energy

Disallow: /retail

Disallow: /mobile/cities-transformed

Disallow: /cit/small-business-mergers-2

Disallow: /cit/small-business-mergers

Disallow: /sites/kpmg/

Disallow: /dell/cities-transformed

Disallow: /sites/neustar

Disallow: /sites/ubs

Disallow: /sites/northerntrust

Disallow: /sites/infiniti

Disallow: /sites/honeywell

Disallow: /sites/charlesschwabadvisor

Disallow: /sites/frisco

Disallow: /sites/colehaan

Disallow: /sites/gapincthiswayahead

Disallow: /sites/discoverpersonalloans

Disallow: /sites/delltechnologies

Disallow: /colehaan/daring-to-step-forward

Disallow: /johnhancock/

Disallow: /johnhancock/in-a-lifetime-healthcare

Disallow: /johnhancock/in-a-lifetime-work

Disallow: /oppenheimerfunds/the-upside

Disallow: /sites/hpeenterprisenxt

Disallow: /sites/rackspace

Disallow: /sites/themacallan

Disallow: /themacallan/worlds-together

Disallow: /sites/inteliotpartnership

Disallow: /oppenheimerfunds/the-upside

Disallow: /partners/

Disallow: /johnhancock/in-a-lifetime-sustainability

Disallow: /sites/domperignon

Disallow: /sites/shinola

Disallow: /sites/hpinc

Disallow: /sites/hp

Disallow: /sites/alfresco

Disallow: /cbre/built-for-performance

Disallow: /sites/mmedric

Disallow: /sites/indiana

Disallow: /sites/siemenshealthineers

Disallow: /sites/rackspace

Disallow: /rackspace/elevating-performance

Disallow: /sites/suntrust

Disallow: /sites/unicefusa

Disallow: /sites/mitsubishiheavyindustries

Disallow: /best-countries-for-business/list

Disallow: /sites/paloaltonetworks

Disallow: /sites/f5

Disallow: /sites/samsungtv

Disallow: /sites/qa

Disallow: /sites/paycom

Disallow: /sites/testbrandvoice

Disallow: /business/blahblah

Disallow: /sites/spdr

Disallow: /sites/lincoln

Disallow: /sites/abb

Disallow: /sites/eycybersecurity

Disallow: /sites/civicnation

Disallow: /sites/bcg

Disallow: /sites/vanguard

Disallow: /sites/akamai

Disallow: /sites/kyocera

Disallow: /sites/socialimpact

Disallow: /sites/emirates

Disallow: /sites/hpincemea

Disallow: /sites/ibmperformancecomputing

Disallow: /sites/pfizer

Disallow: /sites/ibmpowersystems

Disallow: /sites/vanguardcharitable

Disallow: /sites/suntrustprivatewealth

Disallow: /sites/capitalone

Disallow: /sites/juniper

Disallow: /sites/wellsfargo

Disallow: /sites/bnymellon

Disallow: /sites/braintree

Disallow: /sites/bnymellonwealthmanagement

Disallow: /sites/epson

Disallow: /sites/thomsonreuters

Disallow: /sites/paysafe

Disallow: /sites/josabank

Disallow: /sites/hennessyparadisimperial

Disallow: /sites/cartier

Disallow: /sites/anabluewing

Disallow: /sites/eidoo

Disallow: /sites/cmstesting

Disallow: /sites/adobepartnership

Disallow: /sites/cornelljohnson

Disallow: /sites/quantcast

Disallow: /sites/japan

Disallow: /sites/impactpartners

Disallow: /sites/nvidia

Disallow: /sites/personalcapital

Disallow: /sites/microsoft365

Disallow: /sites/providencedigitalinnovation

Disallow: /sites/whittiertrust

Disallow: /sites/kinetica

Disallow: /sites/paulevans

Disallow: /sites/ritz-carltonleadershipcenter

Disallow: /sites/icoengine

Disallow: /sites/tesaro

Disallow: /sites/theglenlivet

Disallow: /feature/sf-test

Disallow: /sites/courtyardmarriott

Disallow: /sites/nuveen

Disallow: /sites/intelai

Disallow: /sites/secret

Disallow: /sites/hec

Disallow: /retail/old

Disallow: /sites/alliantgroup

Disallow: /sites/intuitivesurgical

Disallow: /sites/sunovion

Disallow: /sites/ofx

Disallow: /sites/firstrepublicbank

Disallow: /sites/1850byfolgers

Disallow: /sites/audi

Disallow: /sites/nuveencef

Disallow: /sites/1850brandcoffee

Disallow: /sites/cisco

Disallow: /sites/blackstarcapitalltd

Disallow: /sites/usbankaltitudereserve

Disallow: /sites/dassaultsystemes

Disallow: /sites/urogen

Disallow: /sites/smartsheet

Disallow: /sites/iwc

Disallow: /sites/datingcom

Disallow: /sites/intuitquickbooks

Disallow: /sites/oticon

Disallow: /sites/servicenow

Disallow: /sites/deloitteriskandfinancialadvisory

Disallow: /sites/dbsprivatebank

Disallow: /sites/americanexpress

Disallow: /sites/tanotofoundation

Disallow: /sites/citizensaccess

Disallow: /sites/salesforceessentials

Disallow: /sites/drexeluniversity

Disallow: /sites/microsoftinnovation

Disallow: /sites/mypos

Disallow: /sites/harvardpilgrimhealthcare

Disallow: /sites/appdynamics

Disallow: /sites/ministryoftourismindia

Disallow: /sites/orlando

Disallow: /sites/extrahop

Disallow: /sites/ascm

Disallow: /sites/crowe

Disallow: /sites/peerfit

Disallow: /sites/theleukemialymphomasociety

Disallow: /sites/hewlettpackardenterprise

Disallow: /sites/montblanc

Disallow: /sites/meero

Disallow: /sites/braze

Disallow: /sites/publicissapient

Disallow: /sites/flexport

Sitemap: https://www.forbes.com/news\_sitemap.xml

Sitemap: https://www.forbes.com/video\_sitemap.xml

5.       <https://www.bloomberg.com/> (Legally inquires sending is needed; Technically yes)

[ToS](https://www.bloomberg.com/notices/tos/)

*G. 1. General Inquiries: Please send general inquiries regarding intellectual property issues to* [*FEEDBACK*](https://www.bloomberg.com/feedback)

# Bot rules:

# 1. A bot may not injure a human being or, through inaction, allow a human being to come to harm.

# 2. A bot must obey orders given it by human beings except where such orders would conflict with the First Law.

# 3. A bot must protect its own existence as long as such protection does not conflict with the First or Second Law.

# If you can read this then you should apply here https://www.bloomberg.com/careers/

User-agent: \*

Disallow: /polska

User-agent: Mediapartners-Google

Disallow: /about/careers

Disallow: /about/careers/

Disallow: /offlinemessage/

Disallow: /apps/fbk

Disallow: /bb/newsarchive/

Disallow: /apps/news

User-agent: Spinn3r

Disallow: /podcasts/

Disallow: /feed/podcast/

Disallow: /bb/avfile/

User-agent: Googlebot-News

Disallow: /sponsor/

Disallow: /news/sponsors/\*

Sitemap: https://www.bloomberg.com/sitemap.xml

Sitemap: https://www.bloomberg.com/feeds/bbiz/sitemap\_index.xml

Sitemap: https://www.bloomberg.com/feeds/bpol/sitemap\_index.xml

Sitemap: https://www.bloomberg.com/feeds/businessweek/sitemap\_index.xml

Sitemap: https://www.bloomberg.com/feeds/technology/sitemap\_index.xml

Sitemap: https://www.bloomberg.com/feeds/bbiz/sitemap\_securities\_index.xml

Sitemap: https://www.bloomberg.com/feeds/bbiz/sitemap\_profiles\_company\_index.xml

Sitemap: https://www.bloomberg.com/billionaires/sitemap.xml

Sitemap: https://www.bloomberg.com/feeds/bbiz/sitemap\_news.xml

Sitemap: https://www.bloomberg.com/feeds/dynamic/private-company-index.xml

Sitemap: https://www.bloomberg.com/feeds/dynamic/person-index.xml

Sitemap: https://www.bloomberg.com/feeds/curated/feeds/graphics\_news.xml

Sitemap: https://www.bloomberg.com/feeds/curated/feeds/graphics\_sitemap.xml

Sitemap: https://www.bloomberg.com/pursuits/property-listings/sitemap.xml

6.       [www.reuters.com](http://www.reuters.com/) (Legally written consent is needed; Technically part of the content e.g. /article/beta is unreachable)

[ToS](https://www.reuters.com/terms-of-use)

*4. The text, photographs, video, graphics, metadata, quotes, data, information, the overall “look and feel”, and all protectable intellectual property available through the Service (“Content”) is the property of Reuters and its licensors. It is provided by Reuters and its licensors to you for your personal use and information only. You may not use the Content or Service for any commercial purpose. You may not remove, alter, forward, scrape, copy, sell, distribute, retransmit, create derivative works or otherwise make available the Content to third parties without our prior written consent, except as occasionally permitted by any sharing functionality in the Service that expressly allows you to share Content or links to Content with a few other individuals. You may not reverse engineer, decompile, disassemble or otherwise attempt to discern the source code or architecture of any feature or component of the Service.*

# robots\_allow.txt for www.reuters.com

# Disallow: /\*/key-developments/article/\*

User-agent: \*

Disallow: /finance/stocks/option

Disallow: /finance/stocks/financialHighlights

Disallow: /search

Disallow: /beta

Disallow: /designtech\*

Disallow: /featured-optimize

Disallow: /energy-test

Disallow: /article/beta

Disallow: /sponsored/previewcampaign

Disallow: /sponsored/previewarticle

Disallow: /test/

Disallow: /commentary

Disallow: /news/archive/commentary

Disallow: /brandfeatures/venture-capital\*

SITEMAP: https://www.reuters.com/sitemap\_index.xml

SITEMAP: https://www.reuters.com/sitemap\_news\_index.xml

SITEMAP: https://www.reuters.com/sitemap\_video\_index.xml

SITEMAP: https://www.reuters.com/brandfeature/sitemap

User-agent: Pipl

Disallow: /

User-Agent: omgilibot

Disallow: /

User-Agent: omgili

Disallow: /

User-agent: omgili/0.5 +https://omgili.com

Disallow: /

7.       [www.cnbc.com](http://www.cnbc.com/) (Legally no; Technically yes)

[ToS](https://www.cnbc.com/nbcuniversal-terms-of-service/)

*2. Permitted Uses*

*Your use of the online services shall be limited solely to your personal and non-commercial use. The online services contain material that is derived in whole or in part from material supplied and owned by NBCUniversal as well as third parties (“Content”). As between NBCUniversal and You, NBCUniversal owns all right, title and interest in and to the copyrights, trademarks, service marks, trade names, and all other intellectual and proprietary rights throughout the world associated with the online services and Content. You acknowledge NBCUniversal’s valid intellectual and proprietary property rights in the online services and Content and that your use of the online services is limited to the access, viewing and downloading of Content, all solely as authorized by NBCUniversal. Nothing in these Terms of Service shall be deemed to convey to you any right, title or interest in or to the online services or Content or to any portion thereof except for the limited rights expressly granted herein.*

#

# robots.txt

#

# This file is to prevent the crawling and indexing of certain parts

# of your site by web crawlers and spiders run by sites like Yahoo!

# and Google. By telling these "robots" where not to go on your site,

# you save bandwidth and server resources.

Sitemap: https://www.cnbc.com/sitemapAll.xml

Sitemap: https://www.cnbc.com/sitemap\_news.xml

Sitemap: https://www.cnbc.com/sitemapvideoAll.xml

Sitemap: https://www.cnbc.com/SitemapQuotes.xml

User-agent: googlebot

Disallow: /\*native-android-mobile

Disallow: /\*native-android-tablet

Disallow: /\*mobile-native

Disallow: /preview/

Disallow: /undefined/

Disallow: /select/

User-agent: \*

Disallow: /preview/

Disallow: /undefined/

Disallow: /select/

8.       <https://www.nasdaq.com/> (Legally prior written consent is needed; Technically yes)

[ToS](https://www.nasdaq.com/legal)

*6. Restrictions. Except as detailed in Section 5, User shall not market, sell or distribute the Services or otherwise provide the Services to any third parties including, but not limited to, placing or distributing any Nasdaq’s content on a third party platform or utilizing the Services to direct, manage or provide advisory services for clients of another investment advisor without Nasdaq’s prior written consent. User shall not share, transfer, disclose, copy, publish or create derivative works from the content or the Service without Nasdaq’s prior written approval*

#

# robots.txt

#

# This file is to prevent the crawling and indexing of certain parts

# of your site by web crawlers and spiders run by sites like Yahoo!

# and Google. By telling these "robots" where not to go on your site,

# you save bandwidth and server resources.

#

# This file will be ignored unless it is at the root of your host:

# Used: http://example.com/robots.txt

# Ignored: http://example.com/site/robots.txt

#

# For more information about the robots.txt standard, see:

# http://www.robotstxt.org/robotstxt.html

User-agent: \*

Crawl-delay: 30

# CSS, JS, Images

Allow: /core/\*.css$

Allow: /core/\*.css?

Allow: /core/\*.js$

Allow: /core/\*.js?

Allow: /core/\*.gif

Allow: /core/\*.jpg

Allow: /core/\*.jpeg

Allow: /core/\*.png

Allow: /core/\*.svg

Allow: /profiles/\*.css$

Allow: /profiles/\*.css?

Allow: /profiles/\*.js$

Allow: /profiles/\*.js?

Allow: /profiles/\*.gif

Allow: /profiles/\*.jpg

Allow: /profiles/\*.jpeg

Allow: /profiles/\*.png

Allow: /profiles/\*.svg

# Directories

Disallow: /core/

Disallow: /profiles/

# Files

Disallow: /README.txt

Disallow: /web.config

# Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips

Disallow: /node/add/

Disallow: /search/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Disallow: /user/logout/

# Paths (no clean URLs)

Disallow: /index.php/admin/

Disallow: /index.php/comment/reply/

Disallow: /index.php/filter/tips

Disallow: /index.php/node/add/

Disallow: /index.php/search/

Disallow: /index.php/user/password/

Disallow: /index.php/user/register/

Disallow: /index.php/user/login/

Disallow: /index.php/user/logout/

Sitemap: https://www.nasdaq.com/sitemap.xml

9.       <http://www.stockmarketwire.com/> (Legally no; Technically yes)

[ToS](http://www.stockmarketwire.com/static/terms/)

*Nothing on the website nor your use of it transfers any intellectual property rights in the content to you, and while you may print and download content for your own purposes - whether business or personal - you must not publish, distribute or reproduce it or any information derived from it for general consumption in any form without our written consent.*

*The website is designed for use in the UK only and not in any jurisdiction or location where the publication or availability of the website or its contents would be contrary to local law or regulation. If you use the website outside the UK, you do so at your own risk and you alone shall be responsible for compliance with local regulations.*

User-agent: \*

Disallow: /action/

Disallow: /widgets/

Disallow: /frame/

Disallow: /InvestorsRoom/

Allow: /

10.   <https://www.prnewswire.com/> (Legally content submission is needed; Technically yes)

[ToS](https://www.prnewswire.com/terms-of-use/#info3)

*In the course of using the Site, you may submit or disseminate certain information or content, and you may receive or access information or content submitted or disseminated by others.  Such information and content typically falls into one of three categories:*

* *Personal Information– including personal and professional information about yourself or others.*
* *Content– material generated by you, PR Newswire (including its customers and contributors), or a third party, that incorporates information, ideas, graphics, icons, descriptions, data, compilations of data, text, hyperlinks, messages, tags, opinions or other expressions of thought, photographs, still and moving images, drawings, video, sound recordings, musical compositions, audiovisual works or other creative expressions or works.*
* *Account Information – including all information necessary to establish your user account, such as your username and password, contact information and, in some instances, credit card information.*

*Throughout these TOU, when we refer to information or content “submitted by you” or “submitted by another Site user,” we are referring to any and all types of information and content submitted, posted, uploaded, displayed, performed, transmitted, disseminated, or otherwise made available, whether on, or through, or in connection with the Site, including: (i) in a user account registration; (ii) in a profile; (iii) in communications (such as queries, alerts, advisories, opportunities or responses to any of the same) transmitted to or through the Site; or (iv) in postings (such as forum, chat room, news group, community or message board and blog  postings.*

*When we refer to information or content “provided by PR Newswire,” we are referring to any and all types of information and content submitted, posted, uploaded, displayed, performed, transmitted, disseminated or otherwise made available by PR Newswire on the Site other than information and content submitted by you or another Site user, including information and content provided by our employees, moderators, contributors, suppliers and customers, whether on, or through, or in connection with the Site, including: (i) Site content; (ii) articles contributed by our contributors; and (iii) news releases and other information and content issued by our customers.*

User-Agent: \*

Disallow: /search/

Disallow: /search-results/

Disallow: /comp/

Disallow: /gh/

Disallow: /templates/Blank-HTML

Disallow: /templates/Multivuplayer.html

Disallow: /templates/Home-page-with-spaces.html

Disallow: /templates/Home-page-with-out-spaces.html

Disallow: /templates/PRN\_Custom\_MultiVu\_Recommendation

Disallow: /templates/get-bitly-url.txt

Disallow: /widget-landing-page.html

Disallow: /multivu/

Disallow: /\*.pdf

Sitemap: https://www.prnewswire.com/sitemap.xml

Sitemap: https://www.prnewswire.com/sitemap-news.xml

Sitemap: https://www.prnewswire.com/es/sitemap.xml

Sitemap: https://www.prnewswire.com/es/sitemap-news.xml

Sitemap: https://www.prnewswire.com/sv/sitemap.xml

Sitemap: https://www.prnewswire.com/sv/sitemap-news.xml

Sitemap: https://www.prnewswire.com/ru/sitemap.xml

Sitemap: https://www.prnewswire.com/ru/sitemap-news.xml

Sitemap: https://www.prnewswire.com/sk/sitemap.xml

Sitemap: https://www.prnewswire.com/sk/sitemap-news.xml

Sitemap: https://www.prnewswire.com/pt/sitemap.xml

Sitemap: https://www.prnewswire.com/pt/sitemap-news.xml

Sitemap: https://www.prnewswire.com/pl/sitemap.xml

Sitemap: https://www.prnewswire.com/pl/sitemap-news.xml

Sitemap: https://www.prnewswire.com/nl/sitemap.xml

Sitemap: https://www.prnewswire.com/nl/sitemap-news.xml

Sitemap: https://www.prnewswire.com/it/sitemap.xml

Sitemap: https://www.prnewswire.com/it/sitemap-news.xml

Sitemap: https://www.prnewswire.com/de/sitemap.xml

Sitemap: https://www.prnewswire.com/de/sitemap-news.xml

Sitemap: https://www.prnewswire.com/fi/sitemap.xml

Sitemap: https://www.prnewswire.com/fi/sitemap-news.xml

Sitemap: https://www.prnewswire.com/da/sitemap.xml

Sitemap: https://www.prnewswire.com/da/sitemap-news.xml

Sitemap: https://www.prnewswire.com/cs/sitemap.xml

Sitemap: https://www.prnewswire.com/cs/sitemap-news.xml

Sitemap: https://www.prnewswire.com/fr/sitemap.xml

Sitemap: https://www.prnewswire.com/fr/sitemap-news.xml

Sitemap: https://www.prnewswire.com/no/sitemap.xml

Sitemap: https://www.prnewswire.com/no/sitemap-news.xml

Sitemap: https://www.prnewswire.com/ae/sitemap.xml

Sitemap: https://www.prnewswire.com/ae/sitemap-news.xml

Sitemap: https://www.prnewswire.com/il/sitemap.xml

Sitemap: https://www.prnewswire.com/il/sitemap-news.xml

Sitemap: https://www.prnewswire.com/in/sitemap.xml

Sitemap: https://www.prnewswire.com/in/sitemap-news.xml

11.   <https://globenewswire.com/> (Legally content submission is needed; Technically yes)

[ToS](https://www.west.com/legal-privacy/)

***Data Processing Agreements***

*West has prepared a Data Processing Agreement (“DPA”) that includes the EU Standard Contractual Clauses that promote GDPR compliance.  If your organization would like a written DPA with West, you may download a pre-signed DPA at the link below.  To complete the DPA, you must:*

1. *Download the DPA from the link below;*
2. *Complete all open data fields on the DPA;*
3. *Sign pages 3, 10 and 12 of the DPA; and*
4. *Send the signed copy to* [*legal@west.com*](mailto:legal@west.com)*.*

*The DPA will become effective upon receipt by West.  To access the pre-signed DPA, please visit:*

[*https://www.westuc.com/files/Data-Processing-Agreement.pdf*](https://www.westuc.com/files/Data-Processing-Agreement.pdf)

# robots.txt file created Tue, 09 Oct 2012

User-agent: \*

Disallow: /Search/

Disallow: /SubscribeToRss/

Disallow: /News/Index

Disallow: /newsroom/rss/

Disallow: /newsroom/ctr

# End robots.txt file

12.   <https://www.sec.gov/edgar/searchedgar/companysearch.html> (Legally request submission is needed; Technically yes)

[ToS](https://www.sec.gov/page/oacqsectionlanding)

*If you are interested in obtaining information pertaining to SEC contracts, you must make a request through the SEC’s Freedom of Information and Privacy Act (FOIA) Office. See FOIA link at the bottom of this webpage. Please note that for privacy reasons, federal law prohibits some contract information from being released to the general public, and only information on awarded contracts is allowed to be released.*

#

# robots.txt

#

# This file is to prevent the crawling and indexing of certain parts

# of your site by web crawlers and spiders run by sites like Yahoo!

# and Google. By telling these "robots" where not to go on your site,

# you save bandwidth and server resources.

#

# This file will be ignored unless it is at the root of your host:

# Used: http://example.com/robots.txt

# Ignored: http://example.com/site/robots.txt

#

# For more information about the robots.txt standard, see:

# http://www.robotstxt.org/robotstxt.html

User-agent: \*

# CSS, JS, Images

Allow: /core/\*.css$

Allow: /core/\*.css?

Allow: /core/\*.js$

Allow: /core/\*.js?

Allow: /core/\*.gif

Allow: /core/\*.jpg

Allow: /core/\*.jpeg

Allow: /core/\*.png

Allow: /core/\*.svg

Allow: /profiles/\*.css$

Allow: /profiles/\*.css?

Allow: /profiles/\*.js$

Allow: /profiles/\*.js?

Allow: /profiles/\*.gif

Allow: /profiles/\*.jpg

Allow: /profiles/\*.jpeg

Allow: /profiles/\*.png

Allow: /profiles/\*.svg

# Directories

Disallow: /core/

Disallow: /profiles/

# Files

Disallow: /README.txt

Disallow: /web.config

# Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips

Disallow: /node/add/

#Commented out to support SEC.gov Site Index

# Disallow: /search/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Disallow: /user/logout/

# Paths (no clean URLs)

Disallow: /index.php/admin/

Disallow: /index.php/comment/reply/

Disallow: /index.php/filter/tips

Disallow: /index.php/node/add/

Disallow: /index.php/search/

Disallow: /index.php/user/password/

Disallow: /index.php/user/register/

Disallow: /index.php/user/login/

Disallow: /index.php/user/logout/

#SEC

Allow: /Archives/edgar/data

Disallow: /Archives/bin

Disallow: /Archives/etc

Disallow: /Archives/usr

Disallow: /cgi-bin

Disallow: /bin

Disallow: /Archives/edgar/vprr/XXXX

Disallow: /Archives/edgar/vprr/vprr\_removal

Disallow: /Archives/edgar/vprr/bin

Disallow: /nb

Disallow: /video/live

Disallow: /Archives/edgar/data/1473971/000109181814000042/ex101002.gif

Disallow: /node

Disallow: /spotlight/equity-market-structure-advisory-committee.shtml

Sitemap: http://www.sec.gov/sitemap.xml

Sitemap: https://www.sec.gov/sitemap/sitemap-index.xml

Sitemap: https://www.investor.gov/sitemap.xml

Sitemap: https://www.sec.gov/files-sitemap.xml

13.   [https://www.channelnewsasia.com](https://www.channelnewsasia.com/) (Legally no; Technically yes)

[ToS](https://www.channelnewsasia.com/news/about/termsandconditions)

*All content and materials on the Services and the Sites are the exclusive property of Mediacorp or its content suppliers (including AFP and Associated Press), and may be downloaded or printed for your own personal and non-commercial use only. All copyright notices on downloaded or printed material must be retained. Save as specifically permitted here, you may not copy, reproduce, distribute, republish, repost, modify, transmit, make available to the public, adapt, create a derivative work or otherwise use or exploit any content on the Sites and Services in any way or for any purpose without the express written permission of Mediacorp or its content suppliers (including AFP and Associated Press). In addition, you may not distribute or make available any part of the Sites and Services over any other website or any network.*

*Read more at https://www.channelnewsasia.com/news/about/termsandconditions*

User-agent: \*

Sitemap: https://www.channelnewsasia.com/googlenews/cna\_news\_sitemap.xml

Sitemap: https://www.channelnewsasia.com/sitemap.xml

Crawl-delay: 10

14.   [https://www.straitstimes.com](https://www.straitstimes.com/) (Legally written consent is needed; Technically yes)

[ToS](https://sph.com.sg/legal/website_tnc.html)

*"SPH Content" means all Content of SPH that is made available on or via this Site or a SPH website.*

*You may not reproduce, modify, adapt, translate, publish, display, communicate, transmit, broadcast, podcast, webcast, distribute, sell, trade or exploit for any commercial or other purposes, any portion of, or any access to:*

*4.1.1 any Service;*

*4.1.2 the Site;*

*4.1.3 any SPH Content except, to the extent permitted, with the prior written consent of SPH or unless expressly permitted in these Website Conditions; or*

*4.1.4 any Third Party User Content except, to the extent permitted, with the prior written consent of SPH and the owner or licensee of the specific User Content.*

#

# robots.txt

#

# This file is to prevent the crawling and indexing of certain parts

# of your site by web crawlers and spiders run by sites like Yahoo!

# and Google. By telling these "robots" where not to go on your site,

# you save bandwidth and server resources.

#

# This file will be ignored unless it is at the root of your host:

# Used: http://example.com/robots.txt

# Ignored: http://example.com/site/robots.txt

#

# For more information about the robots.txt standard, see:

# http://www.robotstxt.org/robotstxt.html

User-agent: \*

Crawl-delay: 10

# Directories

Disallow: /includes/

Disallow: /misc/

Disallow: /modules/

Disallow: /profiles/

Disallow: /scripts/

Disallow: /themes/

# Files

Disallow: /CHANGELOG.txt

Disallow: /cron.php

Disallow: /INSTALL.mysql.txt

Disallow: /INSTALL.pgsql.txt

Disallow: /INSTALL.sqlite.txt

Disallow: /install.php

Disallow: /INSTALL.txt

Disallow: /LICENSE.txt

Disallow: /MAINTAINERS.txt

Disallow: /update.php

Disallow: /UPGRADE.txt

Disallow: /xmlrpc.php

# Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips/

Disallow: /node/add/

Disallow: /search/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Disallow: /user/logout/

Disallow: /queue/

Disallow: /flipbook/

Disallow: /advertorials/

# Paths (no clean URLs)

Disallow: /?q=admin/

Disallow: /?q=comment/reply/

Disallow: /?q=filter/tips/

Disallow: /?q=node/add/

Disallow: /?q=search/

Disallow: /?q=user/password/

Disallow: /?q=user/register/

Disallow: /?q=user/login/

Disallow: /?q=user/logout/

Disallow: /\*?adbypass=\*

Disallow: /content/

Disallow: /login\_details.php

Sitemap: https://www.straitstimes.com/googlenews.xml

Sitemap: https://www.straitstimes.com/sitemap.xml

***Reference***

[*https://moz.com/blog/interactive-guide-to-robots-txt*](https://moz.com/blog/interactive-guide-to-robots-txt)

[*https://moz.com/learn/seo/robotstxt*](https://moz.com/learn/seo/robotstxt)