Steven Weil, Aaron Anderson

Professor Adjei

DSC 200

11/26/2024

Web Scraping Report

Web scraping is one of the most important skills that every data scientist should master. There are often times when you are tasked with solving research questions by finding data on the internet. Scraping can make that task substantially more efficient. It is a process where you design a script to parse the HTML content of a page and extract the necessary information. Automating this process allows data scientists to download and collect exponentially more data for their research. For the web scraping in this assignment, we will be using different python libraries such as BeautifulSoup, Selenium, and requests. These libraries make web scraping through python possible and are some of the most impactful tools we have learned in this class. However, web scraping also comes with a lot of responsibility. Many sites disallow access to certain webpages, so you must be sure to look at the robots.txt file and terms of service before scraping. Also, to avoid overloading the server, make sure to add pauses into your scraping script. Scraping is an extremely powerful tool, but just like anything, it must be done correctly and thoughtfully.

The first webpage that we were interested in scraping was Wikipedia’s page on college athletic programs in the state of Kentucky and domain was https://en.wikipedia.org/wiki/List\_of\_college\_athletic\_programs\_in\_Kentucky. Overall, Wikipedia is a website that provides information about almost anything you can imagine. We selected the Kentucky athletics webpage because it contains information that is important to those searching for which colleges have athletic programs in Kentucky. This page is extremely important for prospective athletes and their families as they are looking for colleges to read up on. However, it is all separated and contained to each university, so being able to scrape the data and put it into one csv would make comparing universities much more efficient. The web scraping policies that Wikipedia included on their site could be found in the robots.txt file and on a terms of use page. The robots.txt file is a page that outlines the domain paths that are disallowed for scraping. Almost all websites have a robot.txt file and checking to make sure your page is allowed is a crucial step of the process. Wikipedia indicates that the use of certain scraping bots is not allowed, but not any of the ones we learned in class. Also, it disallows scraping from any domain many domains across the site, but not the one we are looking to scrape from. Finally, the terms of use page did not include any information about scraping being prohibited, so it is clear to scrape in a responsible way.

The second website that we are interested in scraping is Amazon. With the holiday season approaching, it could be valuable to scrape data about the different products you could buy as presents to store and compare them yourself. For this example, we chose Kentucky basketball gear as the search term, so the domain of the page was <https://www.amazon.com/s?k=kentucky+basketball+gear&crid=349JLIBRA51GS&sprefix=%2Caps%2C106&ref=nb_sb_ss_recent_1_0_recent>. The robots.txt file was much more intricate for Amazon, but there were no disallows for the path we chose. Similarly, to the CollegeSimply robots.txt file it also noted that certain bots were not allowed, such as GPTBot, and CCBot. The terms of use page did not mention that scraping was prohibited, so we are able to scrape responsibly.

Overall, this assignment provided a great opportunity for us to learn more about web scraping and how powerful it can be. However, it also taught us the limitations and regulations around web scraping and why it is so important to follow them. There were many sites that we planned on scraping, only to later find out they did not allow scraping. Our first idea was to scrape IMDB for movie data, but they had a strict policy that prevented any data from being scraped from their website. The only permitted use was limited data being used personally that was collected manually. It is important to respect the rules around scraping as digital citizens and data scientists. Web scraping can be one of the best tools we have access to, but it must be used correctly. This assignment was a great way to learn that and continue practicing our skills in web scraping!

Appendix

**Wikipedia robots.txt file:**robots.txt for http://www.wikipedia.org/ and friends

#

# Please note: There are a lot of pages on this site, and there are

# some misbehaved spiders out there that go \_way\_ too fast. If you're

# irresponsible, your access to the site may be blocked.

#

# Observed spamming large amounts of https://en.wikipedia.org/?curid=NNNNNN

# and ignoring 429 ratelimit responses, claims to respect robots:

# http://mj12bot.com/

User-agent: MJ12bot

Disallow: /

# advertising-related bots:

User-agent: Mediapartners-Google\*

Disallow: /

# Wikipedia work bots:

User-agent: IsraBot

Disallow:

User-agent: Orthogaffe

Disallow:

# Crawlers that are kind enough to obey, but which we'd rather not have

# unless they're feeding search engines.

User-agent: UbiCrawler

Disallow: /

User-agent: DOC

Disallow: /

User-agent: Zao

Disallow: /

# Some bots are known to be trouble, particularly those designed to copy

# entire sites. Please obey robots.txt.

User-agent: sitecheck.internetseer.com

Disallow: /

User-agent: Zealbot

Disallow: /

User-agent: MSIECrawler

Disallow: /

User-agent: SiteSnagger

Disallow: /

User-agent: WebStripper

Disallow: /

User-agent: WebCopier

Disallow: /

User-agent: Fetch

Disallow: /

User-agent: Offline Explorer

Disallow: /

User-agent: Teleport

Disallow: /

User-agent: TeleportPro

Disallow: /

User-agent: WebZIP

Disallow: /

User-agent: linko

Disallow: /

User-agent: HTTrack

Disallow: /

User-agent: Microsoft.URL.Control

Disallow: /

User-agent: Xenu

Disallow: /

User-agent: larbin

Disallow: /

User-agent: libwww

Disallow: /

User-agent: ZyBORG

Disallow: /

User-agent: Download Ninja

Disallow: /

# Misbehaving: requests much too fast:

User-agent: fast

Disallow: /

#

# Sorry, wget in its recursive mode is a frequent problem.

# Please read the man page and use it properly; there is a

# --wait option you can use to set the delay between hits,

# for instance.

#

User-agent: wget

Disallow: /

#

# The 'grub' distributed client has been \*very\* poorly behaved.

#

User-agent: grub-client

Disallow: /

#

# Doesn't follow robots.txt anyway, but...

#

User-agent: k2spider

Disallow: /

#

# Hits many times per second, not acceptable

# http://www.nameprotect.com/botinfo.html

User-agent: NPBot

Disallow: /

# A capture bot, downloads gazillions of pages with no public benefit

# http://www.webreaper.net/

User-agent: WebReaper

Disallow: /

#

# Friendly, low-speed bots are welcome viewing article pages, but not

# dynamically-generated pages please.

#

# Inktomi's "Slurp" can read a minimum delay between hits; if your

# bot supports such a thing using the 'Crawl-delay' or another

# instruction, please let us know.

#

# There is a special exception for API mobileview to allow dynamic

# mobile web & app views to load section content.

# These views aren't HTTP-cached but use parser cache aggressively

# and don't expose special: pages etc.

#

# Another exception is for REST API documentation, located at

# /api/rest\_v1/?doc.

#

User-agent: \*

Allow: /w/api.php?action=mobileview&

Allow: /w/load.php?

Allow: /api/rest\_v1/?doc

Disallow: /w/

Disallow: /api/

Disallow: /trap/

Disallow: /wiki/Special:

Disallow: /wiki/Spezial:

Disallow: /wiki/Spesial:

Disallow: /wiki/Special%3A

Disallow: /wiki/Spezial%3A

Disallow: /wiki/Spesial%3A

#

# ar:

Disallow: /wiki/%D8%AE%D8%A7%D8%B5:Search

Disallow: /wiki/%D8%AE%D8%A7%D8%B5%3ASearch

#

# dewiki:

# T6937

# sensible deletion and meta user discussion pages:

Disallow: /wiki/Wikipedia:L%C3%B6schkandidaten/

Disallow: /wiki/Wikipedia:Löschkandidaten/

Disallow: /wiki/Wikipedia:Vandalensperrung/

Disallow: /wiki/Wikipedia:Benutzersperrung/

Disallow: /wiki/Wikipedia:Vermittlungsausschuss/

Disallow: /wiki/Wikipedia:Administratoren/Probleme/

Disallow: /wiki/Wikipedia:Adminkandidaturen/

Disallow: /wiki/Wikipedia:Qualitätssicherung/

Disallow: /wiki/Wikipedia:Qualit%C3%A4tssicherung/

# 4937#5

Disallow: /wiki/Wikipedia:Vandalismusmeldung/

Disallow: /wiki/Wikipedia:Gesperrte\_Lemmata/

Disallow: /wiki/Wikipedia:Löschprüfung/

Disallow: /wiki/Wikipedia:L%C3%B6schprüfung/

Disallow: /wiki/Wikipedia:Administratoren/Notizen/

Disallow: /wiki/Wikipedia:Schiedsgericht/Anfragen/

Disallow: /wiki/Wikipedia:L%C3%B6schpr%C3%BCfung/

# T14111

Disallow: /wiki/Wikipedia:Checkuser/

Disallow: /wiki/Wikipedia\_Diskussion:Checkuser/

Disallow: /wiki/Wikipedia\_Diskussion:Adminkandidaturen/

# T15961

Disallow: /wiki/Wikipedia:Spam-Blacklist-Log

Disallow: /wiki/Wikipedia%3ASpam-Blacklist-Log

Disallow: /wiki/Wikipedia\_Diskussion:Spam-Blacklist-Log

Disallow: /wiki/Wikipedia\_Diskussion%3ASpam-Blacklist-Log

#

# enwiki:

# Folks get annoyed when VfD discussions end up the number 1 google hit for

# their name. See T6776

Disallow: /wiki/Wikipedia:Articles\_for\_deletion/

Disallow: /wiki/Wikipedia%3AArticles\_for\_deletion/

Disallow: /wiki/Wikipedia:Votes\_for\_deletion/

Disallow: /wiki/Wikipedia%3AVotes\_for\_deletion/

Disallow: /wiki/Wikipedia:Pages\_for\_deletion/

Disallow: /wiki/Wikipedia%3APages\_for\_deletion/

Disallow: /wiki/Wikipedia:Miscellany\_for\_deletion/

Disallow: /wiki/Wikipedia%3AMiscellany\_for\_deletion/

Disallow: /wiki/Wikipedia:Miscellaneous\_deletion/

Disallow: /wiki/Wikipedia%3AMiscellaneous\_deletion/

Disallow: /wiki/Wikipedia:Copyright\_problems

Disallow: /wiki/Wikipedia%3ACopyright\_problems

Disallow: /wiki/Wikipedia:Protected\_titles/

Disallow: /wiki/Wikipedia%3AProtected\_titles/

# T15398

Disallow: /wiki/Wikipedia:WikiProject\_Spam/

Disallow: /wiki/Wikipedia%3AWikiProject\_Spam/

# T16075

Disallow: /wiki/MediaWiki:Spam-blacklist

Disallow: /wiki/MediaWiki%3ASpam-blacklist

Disallow: /wiki/MediaWiki\_talk:Spam-blacklist

Disallow: /wiki/MediaWiki\_talk%3ASpam-blacklist

# T13261

Disallow: /wiki/Wikipedia:Requests\_for\_arbitration/

Disallow: /wiki/Wikipedia%3ARequests\_for\_arbitration/

Disallow: /wiki/Wikipedia:Requests\_for\_comment/

Disallow: /wiki/Wikipedia%3ARequests\_for\_comment/

Disallow: /wiki/Wikipedia:Requests\_for\_adminship/

Disallow: /wiki/Wikipedia%3ARequests\_for\_adminship/

# T12288

Disallow: /wiki/Wikipedia\_talk:Articles\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk%3AArticles\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk:Votes\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk%3AVotes\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk:Pages\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk%3APages\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk:Miscellany\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk%3AMiscellany\_for\_deletion/

Disallow: /wiki/Wikipedia\_talk:Miscellaneous\_deletion/

Disallow: /wiki/Wikipedia\_talk%3AMiscellaneous\_deletion/

# T16793

Disallow: /wiki/Wikipedia:Changing\_username

Disallow: /wiki/Wikipedia%3AChanging\_username

Disallow: /wiki/Wikipedia:Changing\_username/

Disallow: /wiki/Wikipedia%3AChanging\_username/

Disallow: /wiki/Wikipedia\_talk:Changing\_username

Disallow: /wiki/Wikipedia\_talk%3AChanging\_username

Disallow: /wiki/Wikipedia\_talk:Changing\_username/

Disallow: /wiki/Wikipedia\_talk%3AChanging\_username/

#

# eswiki:

# T8746

Disallow: /wiki/Wikipedia:Consultas\_de\_borrado/

Disallow: /wiki/Wikipedia%3AConsultas\_de\_borrado/

#

# fiwiki:

# T10695

Disallow: /wiki/Wikipedia:Poistettavat\_sivut

Disallow: /wiki/K%C3%A4ytt%C3%A4j%C3%A4:

Disallow: /wiki/Käyttäjä:

Disallow: /wiki/Keskustelu\_k%C3%A4ytt%C3%A4j%C3%A4st%C3%A4:

Disallow: /wiki/Keskustelu\_käyttäjästä:

Disallow: /wiki/Wikipedia:Yll%C3%A4pit%C3%A4j%C3%A4t/

Disallow: /wiki/Wikipedia:Ylläpitäjät/

#

# hewiki:

Disallow: /wiki/%D7%9E%D7%99%D7%95%D7%97%D7%93:Search

Disallow: /wiki/%D7%9E%D7%99%D7%95%D7%97%D7%93%3ASearch

#T11517

Disallow: /wiki/ויקיפדיה:רשימת\_מועמדים\_למחיקה/

Disallow: /wiki/ויקיפדיה%3Aרשימת\_מועמדים\_למחיקה/

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%A8%D7%A9%D7%99%D7%9E%D7%AA\_%D7%9E%D7%95%D7%A2%D7%9E%D7%93%D7%99%D7%9D\_%D7%9C%D7%9E%D7%97%D7%99%D7%A7%D7%94/

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94%3A%D7%A8%D7%A9%D7%99%D7%9E%D7%AA\_%D7%9E%D7%95%D7%A2%D7%9E%D7%93%D7%99%D7%9D\_%D7%9C%D7%9E%D7%97%D7%99%D7%A7%D7%94/

Disallow: /wiki/ויקיפדיה:ערכים\_לא\_קיימים\_ומוגנים

Disallow: /wiki/ויקיפדיה%3Aערכים\_לא\_קיימים\_ומוגנים

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%A2%D7%A8%D7%9B%D7%99%D7%9D\_%D7%9C%D7%90\_%D7%A7%D7%99%D7%99%D7%9E%D7%99%D7%9D\_%D7%95%D7%9E%D7%95%D7%92%D7%A0%D7%99%D7%9D

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94%3A%D7%A2%D7%A8%D7%9B%D7%99%D7%9D\_%D7%9C%D7%90\_%D7%A7%D7%99%D7%99%D7%9E%D7%99%D7%9D\_%D7%95%D7%9E%D7%95%D7%92%D7%A0%D7%99%D7%9D

Disallow: /wiki/ויקיפדיה:דפים\_לא\_קיימים\_ומוגנים

Disallow: /wiki/ויקיפדיה%3Aדפים\_לא\_קיימים\_ומוגנים

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%93%D7%A4%D7%99%D7%9D\_%D7%9C%D7%90\_%D7%A7%D7%99%D7%99%D7%9E%D7%99%D7%9D\_%D7%95%D7%9E%D7%95%D7%92%D7%A0%D7%99%D7%9D

Disallow: /wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94%3A%D7%93%D7%A4%D7%99%D7%9D\_%D7%9C%D7%90\_%D7%A7%D7%99%D7%99%D7%9E%D7%99%D7%9D\_%D7%95%D7%9E%D7%95%D7%92%D7%A0%D7%99%D7%9D

#

# huwiki:

Disallow: /wiki/Speci%C3%A1lis:Search

Disallow: /wiki/Speci%C3%A1lis%3ASearch

#

# itwiki:

# T7545

Disallow: /wiki/Wikipedia:Pagine\_da\_cancellare

Disallow: /wiki/Wikipedia%3APagine\_da\_cancellare

Disallow: /wiki/Wikipedia:Utenti\_problematici

Disallow: /wiki/Wikipedia%3AUtenti\_problematici

Disallow: /wiki/Wikipedia:Vandalismi\_in\_corso

Disallow: /wiki/Wikipedia%3AVandalismi\_in\_corso

Disallow: /wiki/Wikipedia:Amministratori

Disallow: /wiki/Wikipedia%3AAmministratori

Disallow: /wiki/Wikipedia:Proposte\_di\_cancellazione\_semplificata

Disallow: /wiki/Wikipedia%3AProposte\_di\_cancellazione\_semplificata

Disallow: /wiki/Categoria:Da\_cancellare\_subito

Disallow: /wiki/Categoria%3ADa\_cancellare\_subito

Disallow: /wiki/Wikipedia:Sospette\_violazioni\_di\_copyright

Disallow: /wiki/Wikipedia%3ASospette\_violazioni\_di\_copyright

Disallow: /wiki/Categoria:Da\_controllare\_per\_copyright

Disallow: /wiki/Categoria%3ADa\_controllare\_per\_copyright

Disallow: /wiki/Progetto:Rimozione\_contributi\_sospetti

Disallow: /wiki/Progetto%3ARimozione\_contributi\_sospetti

Disallow: /wiki/Categoria:Da\_cancellare\_subito\_per\_violazione\_integrale\_copyright

Disallow: /wiki/Categoria%3ADa\_cancellare\_subito\_per\_violazione\_integrale\_copyright

Disallow: /wiki/Progetto:Cococo

Disallow: /wiki/Progetto%3ACococo

Disallow: /wiki/Discussioni\_progetto:Cococo

Disallow: /wiki/Discussioni\_progetto%3ACococo

#

# jawiki

Disallow: /wiki/%E7%89%B9%E5%88%A5:Search

Disallow: /wiki/%E7%89%B9%E5%88%A5%3ASearch

# T7239

Disallow: /wiki/Wikipedia:%E5%89%8A%E9%99%A4%E4%BE%9D%E9%A0%BC/

Disallow: /wiki/Wikipedia%3A%E5%89%8A%E9%99%A4%E4%BE%9D%E9%A0%BC/

Disallow: /wiki/Wikipedia:%E5%88%A9%E7%94%A8%E8%80%85%E3%83%9A%E3%83%BC%E3%82%B8%E3%81%AE%E5%89%8A%E9%99%A4%E4%BE%9D%E9%A0%BC

Disallow: /wiki/Wikipedia%3A%E5%88%A9%E7%94%A8%E8%80%85%E3%83%9A%E3%83%BC%E3%82%B8%E3%81%AE%E5%89%8A%E9%99%A4%E4%BE%9D%E9%A0%BC

# nowiki

# T13432

Disallow: /wiki/Bruker:

Disallow: /wiki/Bruker%3A

Disallow: /wiki/Brukerdiskusjon

Disallow: /wiki/Wikipedia:Administratorer

Disallow: /wiki/Wikipedia%3AAdministratorer

Disallow: /wiki/Wikipedia-diskusjon:Administratorer

Disallow: /wiki/Wikipedia-diskusjon%3AAdministratorer

Disallow: /wiki/Wikipedia:Sletting

Disallow: /wiki/Wikipedia%3ASletting

Disallow: /wiki/Wikipedia-diskusjon:Sletting

Disallow: /wiki/Wikipedia-diskusjon%3ASletting

#

# plwiki

# T10067

Disallow: /wiki/Wikipedia:Strony\_do\_usuni%C4%99cia

Disallow: /wiki/Wikipedia%3AStrony\_do\_usuni%C4%99cia

Disallow: /wiki/Wikipedia:Do\_usuni%C4%99cia

Disallow: /wiki/Wikipedia%3ADo\_usuni%C4%99cia

Disallow: /wiki/Wikipedia:SDU/

Disallow: /wiki/Wikipedia%3ASDU/

Disallow: /wiki/Wikipedia:Strony\_podejrzane\_o\_naruszenie\_praw\_autorskich

Disallow: /wiki/Wikipedia%3AStrony\_podejrzane\_o\_naruszenie\_praw\_autorskich

#

# ptwiki:

# T7394

Disallow: /wiki/Wikipedia:Páginas\_para\_eliminar/

Disallow: /wiki/Wikipedia:P%C3%A1ginas\_para\_eliminar/

Disallow: /wiki/Wikipedia%3AP%C3%A1ginas\_para\_eliminar/

Disallow: /wiki/Wikipedia\_Discussão:Páginas\_para\_eliminar/

Disallow: /wiki/Wikipedia\_Discuss%C3%A3o:P%C3%A1ginas\_para\_eliminar/

Disallow: /wiki/Wikipedia\_Discuss%C3%A3o%3AP%C3%A1ginas\_para\_eliminar/

#

# rowiki:

# T14546

Disallow: /wiki/Wikipedia:Pagini\_de\_%C5%9Fters

Disallow: /wiki/Wikipedia%3APagini\_de\_%C5%9Fters

Disallow: /wiki/Discu%C5%A3ie\_Wikipedia:Pagini\_de\_%C5%9Fters

Disallow: /wiki/Discu%C5%A3ie\_Wikipedia%3APagini\_de\_%C5%9Fters

#

# ruwiki:

Disallow: /wiki/%D0%A1%D0%BF%D0%B5%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5:Search

Disallow: /wiki/%D0%A1%D0%BF%D0%B5%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5%3ASearch

#

# svwiki:

# T12229

Disallow: /wiki/Wikipedia%3ASidor\_f%C3%B6reslagna\_f%C3%B6r\_radering

Disallow: /wiki/Wikipedia:Sidor\_f%C3%B6reslagna\_f%C3%B6r\_radering

Disallow: /wiki/Wikipedia:Sidor\_föreslagna\_för\_radering

Disallow: /wiki/Användare

Disallow: /wiki/Anv%C3%A4ndare

Disallow: /wiki/Användardiskussion

Disallow: /wiki/Anv%C3%A4ndardiskussion

Disallow: /wiki/Wikipedia:Skyddade\_sidnamn

Disallow: /wiki/Wikipedia%3ASkyddade\_sidnamn

# T13291

Disallow: /wiki/Wikipedia:Sidor\_som\_bör\_raderas

Disallow: /wiki/Wikipedia:Sidor\_som\_b%C3%B6r\_raderas

Disallow: /wiki/Wikipedia%3ASidor\_som\_b%C3%B6r\_raderas

#

# zhwiki:

# T7104

Disallow: /wiki/Wikipedia:删除投票/侵权

Disallow: /wiki/Wikipedia:%E5%88%A0%E9%99%A4%E6%8A%95%E7%A5%A8/%E4%BE%B5%E6%9D%83

Disallow: /wiki/Wikipedia:删除投票和请求

Disallow: /wiki/Wikipedia:%E5%88%A0%E9%99%A4%E6%8A%95%E7%A5%A8%E5%92%8C%E8%AF%B7%E6%B1%82

Disallow: /wiki/Category:快速删除候选

Disallow: /wiki/Category:%E5%BF%AB%E9%80%9F%E5%88%A0%E9%99%A4%E5%80%99%E9%80%89

Disallow: /wiki/Category:维基百科需要翻译的文章

Disallow: /wiki/Category:%E7%BB%B4%E5%9F%BA%E7%99%BE%E7%A7%91%E9%9C%80%E8%A6%81%E7%BF%BB%E8%AF%91%E7%9A%84%E6%96%87%E7%AB%A0

#

# sister projects

#

# enwikinews:

# T7340

Disallow: /wiki/Portal:Prepared\_stories/

Disallow: /wiki/Portal%3APrepared\_stories/

#

# itwikinews

# T11138

Disallow: /wiki/Wikinotizie:Richieste\_di\_cancellazione

Disallow: /wiki/Wikinotizie:Sospette\_violazioni\_di\_copyright

Disallow: /wiki/Categoria:Da\_cancellare\_subito

Disallow: /wiki/Categoria:Da\_cancellare\_subito\_per\_violazione\_integrale\_copyright

Disallow: /wiki/Wikinotizie:Storie\_in\_preparazione

#

# enwikiquote:

# T17095

Disallow: /wiki/Wikiquote:Votes\_for\_deletion/

Disallow: /wiki/Wikiquote%3AVotes\_for\_deletion/

Disallow: /wiki/Wikiquote\_talk:Votes\_for\_deletion/

Disallow: /wiki/Wikiquote\_talk%3AVotes\_for\_deletion/

Disallow: /wiki/Wikiquote:Votes\_for\_deletion\_archive/

Disallow: /wiki/Wikiquote%3AVotes\_for\_deletion\_archive/

Disallow: /wiki/Wikiquote\_talk:Votes\_for\_deletion\_archive/

Disallow: /wiki/Wikiquote\_talk%3AVotes\_for\_deletion\_archive/

#

# enwikibooks

Disallow: /wiki/Wikibooks:Votes\_for\_deletion

#

# working...

Disallow: /wiki/Fundraising\_2007/comments

#

#

#

#----------------------------------------------------------#

#

#

#

#

# Localisable part of robots.txt for en.wikipedia.org

#

# Edit at https://en.wikipedia.org/w/index.php?title=MediaWiki:Robots.txt&action=edit

# Don't add newlines here. All rules set here are active for every user-agent.

#

# Please check any changes using a syntax validator

# Enter https://en.wikipedia.org/robots.txt as the URL to check.

#

# https://phabricator.wikimedia.org/T16075

Disallow: /wiki/MediaWiki:Spam-blacklist

Disallow: /wiki/MediaWiki%3ASpam-blacklist

Disallow: /wiki/MediaWiki\_talk:Spam-blacklist

Disallow: /wiki/MediaWiki\_talk%3ASpam-blacklist

Disallow: /wiki/Wikipedia:WikiProject\_Spam

Disallow: /wiki/Wikipedia\_talk:WikiProject\_Spam

#

# Folks get annoyed when XfD discussions end up the number 1 google hit for

# their name.

# https://phabricator.wikimedia.org/T16075

Disallow: /wiki/Wikipedia:Articles\_for\_deletion

Disallow: /wiki/Wikipedia%3AArticles\_for\_deletion

Disallow: /wiki/Wikipedia:Votes\_for\_deletion

Disallow: /wiki/Wikipedia%3AVotes\_for\_deletion

Disallow: /wiki/Wikipedia:Pages\_for\_deletion

Disallow: /wiki/Wikipedia%3APages\_for\_deletion

Disallow: /wiki/Wikipedia:Miscellany\_for\_deletion

Disallow: /wiki/Wikipedia%3AMiscellany\_for\_deletion

Disallow: /wiki/Wikipedia:Miscellaneous\_deletion

Disallow: /wiki/Wikipedia%3AMiscellaneous\_deletion

Disallow: /wiki/Wikipedia:Categories\_for\_discussion

Disallow: /wiki/Wikipedia%3ACategories\_for\_discussion

Disallow: /wiki/Wikipedia:Templates\_for\_deletion

Disallow: /wiki/Wikipedia%3ATemplates\_for\_deletion

Disallow: /wiki/Wikipedia:Redirects\_for\_discussion

Disallow: /wiki/Wikipedia%3ARedirects\_for\_discussion

Disallow: /wiki/Wikipedia:Deletion\_review

Disallow: /wiki/Wikipedia%3ADeletion\_review

Disallow: /wiki/Wikipedia:WikiProject\_Deletion\_sorting

Disallow: /wiki/Wikipedia%3AWikiProject\_Deletion\_sorting

Disallow: /wiki/Wikipedia:Files\_for\_deletion

Disallow: /wiki/Wikipedia%3AFiles\_for\_deletion

Disallow: /wiki/Wikipedia:Files\_for\_discussion

Disallow: /wiki/Wikipedia%3AFiles\_for\_discussion

Disallow: /wiki/Wikipedia:Possibly\_unfree\_files

Disallow: /wiki/Wikipedia%3APossibly\_unfree\_files

#

# https://phabricator.wikimedia.org/T12288

Disallow: /wiki/Wikipedia\_talk:Articles\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3AArticles\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Votes\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3AVotes\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Pages\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3APages\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Miscellany\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3AMiscellany\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Miscellaneous\_deletion

Disallow: /wiki/Wikipedia\_talk%3AMiscellaneous\_deletion

Disallow: /wiki/Wikipedia\_talk:Templates\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3ATemplates\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Categories\_for\_discussion

Disallow: /wiki/Wikipedia\_talk%3ACategories\_for\_discussion

Disallow: /wiki/Wikipedia\_talk:Deletion\_review

Disallow: /wiki/Wikipedia\_talk%3ADeletion\_review

Disallow: /wiki/Wikipedia\_talk:WikiProject\_Deletion\_sorting

Disallow: /wiki/Wikipedia\_talk%3AWikiProject\_Deletion\_sorting

Disallow: /wiki/Wikipedia\_talk:Files\_for\_deletion

Disallow: /wiki/Wikipedia\_talk%3AFiles\_for\_deletion

Disallow: /wiki/Wikipedia\_talk:Files\_for\_discussion

Disallow: /wiki/Wikipedia\_talk%3AFiles\_for\_discussion

Disallow: /wiki/Wikipedia\_talk:Possibly\_unfree\_files

Disallow: /wiki/Wikipedia\_talk%3APossibly\_unfree\_files

#

Disallow: /wiki/Wikipedia:Copyright\_problems

Disallow: /wiki/Wikipedia%3ACopyright\_problems

Disallow: /wiki/Wikipedia\_talk:Copyright\_problems

Disallow: /wiki/Wikipedia\_talk%3ACopyright\_problems

Disallow: /wiki/Wikipedia:Suspected\_copyright\_violations

Disallow: /wiki/Wikipedia%3ASuspected\_copyright\_violations

Disallow: /wiki/Wikipedia\_talk:Suspected\_copyright\_violations

Disallow: /wiki/Wikipedia\_talk%3ASuspected\_copyright\_violations

Disallow: /wiki/Wikipedia:Contributor\_copyright\_investigations

Disallow: /wiki/Wikipedia%3AContributor\_copyright\_investigations

Disallow: /wiki/Wikipedia:Contributor\_copyright\_investigations

Disallow: /wiki/Wikipedia%3AContributor\_copyright\_investigations

Disallow: /wiki/Wikipedia\_talk:Contributor\_copyright\_investigations

Disallow: /wiki/Wikipedia\_talk%3AContributor\_copyright\_investigations

Disallow: /wiki/Wikipedia\_talk:Contributor\_copyright\_investigations

Disallow: /wiki/Wikipedia\_talk%3AContributor\_copyright\_investigations

Disallow: /wiki/Wikipedia:Protected\_titles

Disallow: /wiki/Wikipedia%3AProtected\_titles

Disallow: /wiki/Wikipedia\_talk:Protected\_titles

Disallow: /wiki/Wikipedia\_talk%3AProtected\_titles

Disallow: /wiki/Wikipedia:Articles\_for\_creation

Disallow: /wiki/Wikipedia%3AArticles\_for\_creation

Disallow: /wiki/Wikipedia\_talk:Articles\_for\_creation

Disallow: /wiki/Wikipedia\_talk%3AArticles\_for\_creation

Disallow: /wiki/Wikipedia\_talk:Article\_wizard

Disallow: /wiki/Wikipedia\_talk%3AArticle\_wizard

#

# https://phabricator.wikimedia.org/T13261

Disallow: /wiki/Wikipedia:Requests\_for\_arbitration

Disallow: /wiki/Wikipedia%3ARequests\_for\_arbitration

Disallow: /wiki/Wikipedia\_talk:Requests\_for\_arbitration

Disallow: /wiki/Wikipedia\_talk%3ARequests\_for\_arbitration

Disallow: /wiki/Wikipedia:Requests\_for\_comment

Disallow: /wiki/Wikipedia%3ARequests\_for\_comment

Disallow: /wiki/Wikipedia\_talk:Requests\_for\_comment

Disallow: /wiki/Wikipedia\_talk%3ARequests\_for\_comment

Disallow: /wiki/Wikipedia:Requests\_for\_adminship

Disallow: /wiki/Wikipedia%3ARequests\_for\_adminship

Disallow: /wiki/Wikipedia\_talk:Requests\_for\_adminship

Disallow: /wiki/Wikipedia\_talk%3ARequests\_for\_adminship

#

# https://phabricator.wikimedia.org/T14111

Disallow: /wiki/Wikipedia:Requests\_for\_checkuser

Disallow: /wiki/Wikipedia%3ARequests\_for\_checkuser

Disallow: /wiki/Wikipedia\_talk:Requests\_for\_checkuser

Disallow: /wiki/Wikipedia\_talk%3ARequests\_for\_checkuser

#

# https://phabricator.wikimedia.org/T15398

Disallow: /wiki/Wikipedia:WikiProject\_Spam

Disallow: /wiki/Wikipedia%3AWikiProject\_Spam

#

# https://phabricator.wikimedia.org/T16793

Disallow: /wiki/Wikipedia:Changing\_username

Disallow: /wiki/Wikipedia%3AChanging\_username

Disallow: /wiki/Wikipedia:Changing\_username

Disallow: /wiki/Wikipedia%3AChanging\_username

Disallow: /wiki/Wikipedia\_talk:Changing\_username

Disallow: /wiki/Wikipedia\_talk%3AChanging\_username

Disallow: /wiki/Wikipedia\_talk:Changing\_username

Disallow: /wiki/Wikipedia\_talk%3AChanging\_username

#

Disallow: /wiki/Wikipedia:Administrators%27\_noticeboard

Disallow: /wiki/Wikipedia%3AAdministrators%27\_noticeboard

Disallow: /wiki/Wikipedia\_talk:Administrators%27\_noticeboard

Disallow: /wiki/Wikipedia\_talk%3AAdministrators%27\_noticeboard

Disallow: /wiki/Wikipedia:Community\_sanction\_noticeboard

Disallow: /wiki/Wikipedia%3ACommunity\_sanction\_noticeboard

Disallow: /wiki/Wikipedia\_talk:Community\_sanction\_noticeboard

Disallow: /wiki/Wikipedia\_talk%3ACommunity\_sanction\_noticeboard

Disallow: /wiki/Wikipedia:Bureaucrats%27\_noticeboard

Disallow: /wiki/Wikipedia%3ABureaucrats%27\_noticeboard

Disallow: /wiki/Wikipedia\_talk:Bureaucrats%27\_noticeboard

Disallow: /wiki/Wikipedia\_talk%3ABureaucrats%27\_noticeboard

#

Disallow: /wiki/Wikipedia:Sockpuppet\_investigations

Disallow: /wiki/Wikipedia%3ASockpuppet\_investigations

Disallow: /wiki/Wikipedia\_talk:Sockpuppet\_investigations

Disallow: /wiki/Wikipedia\_talk%3ASockpuppet\_investigations

#

Disallow: /wiki/Wikipedia:Neutral\_point\_of\_view/Noticeboard

Disallow: /wiki/Wikipedia%3ANeutral\_point\_of\_view/Noticeboard

Disallow: /wiki/Wikipedia\_talk:Neutral\_point\_of\_view/Noticeboard

Disallow: /wiki/Wikipedia\_talk%3ANeutral\_point\_of\_view/Noticeboard

#

Disallow: /wiki/Wikipedia:No\_original\_research/noticeboard

Disallow: /wiki/Wikipedia%3ANo\_original\_research/noticeboard

Disallow: /wiki/Wikipedia\_talk:No\_original\_research/noticeboard

Disallow: /wiki/Wikipedia\_talk%3ANo\_original\_research/noticeboard

#

Disallow: /wiki/Wikipedia:Fringe\_theories/Noticeboard

Disallow: /wiki/Wikipedia%3AFringe\_theories/Noticeboard

Disallow: /wiki/Wikipedia\_talk:Fringe\_theories/Noticeboard

Disallow: /wiki/Wikipedia\_talk%3AFringe\_theories/Noticeboard

#

Disallow: /wiki/Wikipedia:Conflict\_of\_interest/Noticeboard

Disallow: /wiki/Wikipedia%3AConflict\_of\_interest/Noticeboard

Disallow: /wiki/Wikipedia\_talk:Conflict\_of\_interest/Noticeboard

Disallow: /wiki/Wikipedia\_talk%3AConflict\_of\_interest/Noticeboard

#

Disallow: /wiki/Wikipedia:Long-term\_abuse

Disallow: /wiki/Wikipedia%3ALong-term\_abuse

Disallow: /wiki/Wikipedia\_talk:Long-term\_abuse

Disallow: /wiki/Wikipedia\_talk%3ALong-term\_abuse

Disallow: /wiki/Wikipedia:Long\_term\_abuse

Disallow: /wiki/Wikipedia%3ALong\_term\_abuse

Disallow: /wiki/Wikipedia\_talk:Long\_term\_abuse

Disallow: /wiki/Wikipedia\_talk%3ALong\_term\_abuse

#

Disallow: /wiki/Wikipedia:Wikiquette\_assistance

Disallow: /wiki/Wikipedia%3AWikiquette\_assistance

#

Disallow: /wiki/Wikipedia:Abuse\_reports

Disallow: /wiki/Wikipedia%3AAbuse\_reports

Disallow: /wiki/Wikipedia\_talk:Abuse\_reports

Disallow: /wiki/Wikipedia\_talk%3AAbuse\_reports

Disallow: /wiki/Wikipedia:Abuse\_response

Disallow: /wiki/Wikipedia%3AAbuse\_response

Disallow: /wiki/Wikipedia\_talk:Abuse\_response

Disallow: /wiki/Wikipedia\_talk%3AAbuse\_response

#

Disallow: /wiki/Wikipedia:Reliable\_sources/Noticeboard

Disallow: /wiki/Wikipedia%3AReliable\_sources/Noticeboard

Disallow: /wiki/Wikipedia\_talk:Reliable\_sources/Noticeboard

Disallow: /wiki/Wikipedia\_talk%3AReliable\_sources/Noticeboard

#

Disallow: /wiki/Wikipedia:Suspected\_sock\_puppets

Disallow: /wiki/Wikipedia%3ASuspected\_sock\_puppets

Disallow: /wiki/Wikipedia\_talk:Suspected\_sock\_puppets

Disallow: /wiki/Wikipedia\_talk%3ASuspected\_sock\_puppets

#

Disallow: /wiki/Wikipedia:Biographies\_of\_living\_persons/Noticeboard

Disallow: /wiki/Wikipedia%3ABiographies\_of\_living\_persons/Noticeboard

Disallow: /wiki/Wikipedia\_talk:Biographies\_of\_living\_persons/Noticeboard

Disallow: /wiki/Wikipedia\_talk%3ABiographies\_of\_living\_persons/Noticeboard

Disallow: /wiki/Wikipedia:Biographies\_of\_living\_persons%2FNoticeboard

Disallow: /wiki/Wikipedia%3ABiographies\_of\_living\_persons%2FNoticeboard

Disallow: /wiki/Wikipedia\_talk:Biographies\_of\_living\_persons%2FNoticeboard

Disallow: /wiki/Wikipedia\_talk%3ABiographies\_of\_living\_persons%2FNoticeboard

#

Disallow: /wiki/Wikipedia:Content\_noticeboard

Disallow: /wiki/Wikipedia%3AContent\_noticeboard

Disallow: /wiki/Wikipedia\_talk:Content\_noticeboard

Disallow: /wiki/Wikipedia\_talk%3AContent\_noticeboard

#

Disallow: /wiki/Template:Editnotices

Disallow: /wiki/Template%3AEditnotices

#

Disallow: /wiki/Wikipedia:Arbitration

Disallow: /wiki/Wikipedia%3AArbitration

Disallow: /wiki/Wikipedia\_talk:Arbitration

Disallow: /wiki/Wikipedia\_talk%3AArbitration

#

Disallow: /wiki/Wikipedia:Arbitration\_Committee

Disallow: /wiki/Wikipedia%3AArbitration\_Committee

Disallow: /wiki/Wikipedia\_talk:Arbitration\_Committee

Disallow: /wiki/Wikipedia\_talk%3AArbitration\_Committee

#

Disallow: /wiki/Wikipedia:Arbitration\_Committee\_Elections

Disallow: /wiki/Wikipedia%3AArbitration\_Committee\_Elections

Disallow: /wiki/Wikipedia\_talk:Arbitration\_Committee\_Elections

Disallow: /wiki/Wikipedia\_talk%3AArbitration\_Committee\_Elections

#

Disallow: /wiki/Wikipedia:Mediation\_Committee

Disallow: /wiki/Wikipedia%3AMediation\_Committee

Disallow: /wiki/Wikipedia\_talk:Mediation\_Committee

Disallow: /wiki/Wikipedia\_talk%3AMediation\_Committee

#

Disallow: /wiki/Wikipedia:Mediation\_Cabal/Cases

Disallow: /wiki/Wikipedia%3AMediation\_Cabal/Cases

#

Disallow: /wiki/Wikipedia:Requests\_for\_bureaucratship

Disallow: /wiki/Wikipedia%3ARequests\_for\_bureaucratship

Disallow: /wiki/Wikipedia\_talk:Requests\_for\_bureaucratship

Disallow: /wiki/Wikipedia\_talk%3ARequests\_for\_bureaucratship

#

Disallow: /wiki/Wikipedia:Administrator\_review

Disallow: /wiki/Wikipedia%3AAdministrator\_review

Disallow: /wiki/Wikipedia\_talk:Administrator\_review

Disallow: /wiki/Wikipedia\_talk%3AAdministrator\_review

#

Disallow: /wiki/Wikipedia:Editor\_review

Disallow: /wiki/Wikipedia%3AEditor\_review

Disallow: /wiki/Wikipedia\_talk:Editor\_review

Disallow: /wiki/Wikipedia\_talk%3AEditor\_review

#

Disallow: /wiki/Wikipedia:Article\_Incubator

Disallow: /wiki/Wikipedia%3AArticle\_Incubator

Disallow: /wiki/Wikipedia\_talk:Article\_Incubator

Disallow: /wiki/Wikipedia\_talk%3AArticle\_Incubator

#

Disallow: /wiki/Category:Noindexed\_pages

Disallow: /wiki/Category%3ANoindexed\_pages

#

# User sandboxes for modules and Template Styles are placed in these subpages for testing

#

Disallow: /wiki/Module:Sandbox

Disallow: /wiki/Module%3ASandbox

Disallow: /wiki/Template:TemplateStyles\_sandbox

Disallow: /wiki/Template%3ATemplateStyles\_sandbox

# Adminship and arbitration discussions are blocked from search engines, so blocking newer adminship processes (as of 2024) as well

Disallow: /wiki/Wikipedia:Administrator\_recall

Disallow: /wiki/Wikipedia%3AAdministrator\_recall

Disallow: /wiki/Wikipedia\_talk:Administrator\_recall

Disallow: /wiki/Wikipedia\_talk%3AAdministrator\_recall

Disallow: /wiki/Wikipedia:Administrator\_elections

Disallow: /wiki/Wikipedia%3AAdministrator\_elections

Disallow: /wiki/Wikipedia\_talk:Administrator\_elections

Disallow: /wiki/Wikipedia\_talk%3AAdministrator\_elections

**Amazon robots.txt file:**

User-agent: \*

Disallow: /exec/obidos/account-access-login

Disallow: /exec/obidos/change-style

Disallow: /exec/obidos/flex-sign-in

Disallow: /exec/obidos/handle-buy-box

Disallow: /exec/obidos/tg/cm/member/

Disallow: /gp/aw/help/id=sss

Disallow: /gp/cart

Disallow: /gp/flex

Disallow: /gp/product/e-mail-friend

Disallow: /gp/product/product-availability

Disallow: /gp/product/rate-this-item

Disallow: /gp/sign-in

Disallow: /gp/reader

Disallow: /gp/sitbv3/reader

Disallow: /gp/richpub/syltguides/create

Disallow: /gp/gfix

Disallow: /gp/associations/wizard.html

Disallow: /gp/dmusic/order

Disallow: /gp/legacy-handle-buy-box.html

Disallow: /gp/aws/ssop

Disallow: /gp/yourstore

Disallow: /gp/gift-central/organizer/add-wishlist

Disallow: /gp/vote

Disallow: /gp/voting/

Disallow: /gp/music/wma-pop-up

Disallow: /gp/customer-images

Disallow: /gp/richpub/listmania/createpipeline

Disallow: /gp/content-form

Disallow: /gp/pdp/invitation/invite

Disallow: /gp/customer-reviews/common/du

Disallow: /gp/customer-reviews/write-a-review.html

Disallow: /gp/associations/wizard.html

Disallow: /gp/music/clipserve

Disallow: /gp/customer-media/upload

Disallow: /gp/history

Disallow: /gp/item-dispatch

Disallow: /gp/dmusic/order/handle-buy-box.html

Disallow: /gp/recsradio

Disallow: /gp/slredirect

Disallow: /dp/shipping/

Disallow: /dp/twister-update/

Disallow: /dp/manual-submit/

Disallow: /dp/e-mail-friend/

Disallow: /dp/product-availability/

Disallow: /dp/rate-this-item/

Disallow: /gp/registry/wishlist/\*/reserve

Disallow: /gp/structured-ratings/actions/get-experience.html

Disallow: /gp/twitter/

Disallow: /ap/signin

Disallow: /gp/registry/wishlist/

Disallow: /wishlist/

Allow: /wishlist/universal\*

Allow: /wishlist/vendor-button\*

Allow: /wishlist/get-button\*

Disallow: /gp/wishlist/

Allow: /gp/wishlist/universal\*

Allow: /gp/wishlist/vendor-button\*

Allow: /gp/wishlist/ipad-install\*

Disallow: /registry/wishlist/

Disallow: /review/common/du

Disallow: /gp/registry/search.html

Disallow: /product-reviews/B0069IY63Y

Disallow: /gp/orc/rml/

Disallow: \*/gcrnsts

Disallow: /gp/gc/widget

Disallow: /gp/dmusic/mp3/player

Disallow: /gp/entity-alert/external

Disallow: /gp/customer-reviews/dynamic/sims-box

Disallow: /review/dynamic/sims-box

Disallow: /gp/redirect.html

Disallow: /gp/twister/ajaxv2

Disallow: /ss/twister/ajax

Disallow: /b?\*node=7454917011

Disallow: /b?\*node=7454927011

Disallow: /b?\*node=7454939011

Disallow: /b?\*node=7454898011

Disallow: /gp/customer-media/actions/delete/

Disallow: /gp/customer-media/actions/edit-caption/

Disallow: /gp/dmusic/

Allow: /gp/dmusic/promotions/PrimeMusic

Allow: /gp/dmusic/promotions/AmazonMusicUnlimited

Disallow: /gp/offer-listing/

Disallow: /b?\*node=9052533011

Disallow: /lm/R1XIHQVKXSKBNJ

Disallow: /lm/R3HQ5WJSZK6QSO

Disallow: /surprise/

Disallow: /local/ajax/

Disallow: /gp/socialmedia/giveaways

Disallow: /gp/b2b-rd

Disallow: /gp/aw/so.html

Disallow: /reviews/iframe

Disallow: /gp/switch-language

Disallow: /ga/p/

Disallow: /gp/profile/

Disallow: /giveaway/host/setup/

Disallow: /ss/customer-reviews/lighthouse/

Disallow: /ospublishing/story/\*

Disallow: /gp/aw/ol/

Disallow: /gp/promotion/

Disallow: /hz/leaderboard/top-reviewers/

Disallow: /creatorhub

Disallow: /creatorhub/\*

Disallow: /slp/s$

Disallow: /-/

Allow: /-/es/

Allow: /-/en$

Allow: /-/zh\_TW/

Allow: /-/zh\_TW$

Allow: /-/he/

Allow: /-/he$

Disallow: /hz/help/contact/\*/message/$

Disallow: /gp/aw/shoppingAids/

Disallow: /rss/people/\*/reviews

Disallow: /gp/pdp/rss/\*/reviews

Disallow: /gp/cdp/member-reviews/

Disallow: /gp/aw/cr/

Allow: /gp/offer-listing/B000

Allow: /gp/offer-listing/9000

Disallow: /gp/aag

Allow: /gp/aag/main?\*seller=ABVFEJU8LS620

Disallow: /gp/pdp/profile/

Disallow: /gp/help/customer/express/c2c/

Disallow: /slp/\*/b$

Disallow: /hz/contact-us/ajax/initiate-trusted-contact/

Disallow: /hp/video/

Disallow: /gp/video/api

Disallow: /gp/video/settings

Disallow: /gp/video/library

Disallow: /gp/video/watchlist

Disallow: /gp/video/mystuff

Disallow: /gp/video/profiles

Disallow: /gp/video/search

Disallow: /gp/video/auth

Disallow: /gp/video/salp/i

User-agent: EtaoSpider

Disallow: /

User-agent: GPTBot

Disallow: /

User-agent: CCBot

Disallow: /

User-Agent: PerplexityBot

Disallow: /

Links to terms of use pages:

<https://foundation.wikimedia.org/wiki/Policy:Terms_of_Use/en>

<https://www.amazon.com/gp/help/customer/display.html?nodeId=202140280>