

Day 3 — Web App & Content OSINT (step-by-step)

Goal: discover hidden files, unpublished documents, versioned pages, and sensitive content exposed on web servers (and historically) without exploiting anything.

Main tools: `gobuster` or `dirsearch` (directory brute force), browser Google Dorks, Burp Suite Community (or `curl/wget`), Wayback Machine (archive.org), `wget/curl` for safe download, SecLists wordlists.

Folder: `day3_webcontent/` with subfolders `scans/`, `evidence/`, `downloads/`

0) Setup

- Create folders:
 - `mkdir -p day3_webcontent/scans day3_webcontent/evidence day3_webcontent/downloads`
- Install wordlists (SecLists):
 - on Linux: `sudo apt install seclists` or `git clone https://github.com/danielmiessler/SecLists.git`
- Install tools:
 - `gobuster` (Go) or `dirsearch` (Python)
 - Burp Suite Community (Java) optional
 - `wget / curl` available everywhere

1) Collect robots.txt and sitemap.xml (first look)

These often point to interesting directories/files.

PowerShell:

```

# TLS 1.2 (if needed)
[Net.ServicePointManager]::SecurityProtocol =
[Net.SecurityProtocolType]::Tls12

# Ensure folder exists
New-Item -ItemType Directory -Path "day3_webcontent\scans" -Force

# Download robots.txt from a real site
Invoke-WebRequest -Uri "https://owasp.org/robots.txt" -OutFile
"day3_webcontent\scans\robots_owasp.txt"

# Download sitemap.xml from the same site (if available)
Invoke-WebRequest -Uri "https://owasp.org/sitemap.xml" -OutFile
"day3_webcontent\scans\sitemap_owasp.xml"

```

```

PS C:\Users\Nidhi> [Net.ServicePointManager]::SecurityProtocol = [Net.SecurityProtocolType]::Tls12
PS C:\Users\Nidhi> > New-Item -ItemType Directory -Path "day3_webcontent\scans" -Force
PS C:\Users\Nidhi> Directory: C:\Users\Nidhi !
Mode                LastWriteTime        Length Name
----                -----        ----- 
d-----  25-09-2025      13:54            scans

PS C:\Users\Nidhi> > Invoke-WebRequest -Uri "https://owasp.org/robots.txt" -OutFile "day3_webcontent\scans\robots_owasp.txt"
PS C:\Users\Nidhi> > Invoke-WebRequest -Uri "https://owasp.org/sitemap.xml" -OutFile "day3_webcontent\scans\sitemap_owasp.xml"

```

2) Directory brute-force (safe, non-destructive)

Important: Use only on domains you control. Keep rate low to avoid DoS.

Install: `pip install git+https://github.com/maurosoria/dirsearch.git`

```

python C:\tools\dirsearch\dirsearch.py -u https://example.com -w
C:\tools\SecLists-master\SecLists-master\Discovery\Web-Content\commo
n.txt -e php,html,txt,log,sql -o "C:\Users\Nidhi [REDACTED]"
[REDACTED]\day3_webcontent\scans\dirsearch_jaide.txt"

```

```

PS C:\Users\Nidhi [REDACTED] python C:\tool\dirsearch.py -u https://example.com -w C:\tools\SecLists-master\SecLists-master\Discovery\Web-Content\common.txt -e php,html,txt,log,sql -o "C:\Users\Nidhi [REDACTED]\dirsearch_juice.txt"
dir5_d5c5ch v0.4.3
Extensions: php, html, txt, log, sql | HTTP method: GET | Threads: 25 | Wordlist size: 4749
Target: https://example.com/
[14:12:37] Scanning:
[14:12:57] 404 - 4108 - /1996
[14:12:57] 404 - 4108 - /14
[14:12:57] 403 - 3638 - /2001
[14:12:57] 403 - 3638 - /2004
[14:12:57] 403 - 3638 - /2010
[14:12:57] 403 - 3638 - /2019
[14:12:58] 403 - 3618 - /25
[14:12:59] 403 - 3628 - /500
[14:13:24] 403 - 3738 - /cardinalform
[14:13:24] 404 - 4128 - /carbuyaction
[14:13:24] 403 - 3728 - /carthandler
[14:14:03] 200 - 1KB - /index.html
[14:14:41] 403 - 3628 - /reg
[14:14:41] 403 - 3678 - /regional
[14:14:43] 403 - 3668 - /request
[14:14:44] 403 - 3698 - /requests
[14:14:50] 404 - 4108 - /shared
[14:14:50] 404 - 4108 - /share
[14:14:50] 404 - 4108 - /ship
[14:14:50] 403 - 3638 - /shop
[14:14:50] 403 - 3748 - /shop_closed
[14:14:53] 404 - 4108 - /sitecore
[14:14:53] 404 - 4108 - /sitemages
[14:14:53] 404 - 4108 - /sitemaps
[14:14:53] 403 - 3638 - /skip
[14:14:53] 404 - 4108 - /sk
[14:14:54] 403 - 3778 - /skin1_original
[14:14:54] 403 - 3628 - /smf
[14:14:54] 403 - 3658 - /smiles
[14:14:54] 403 - 3668 - /small
[14:14:55] 403 - 3698 - /slimstat
[14:14:56] 403 - 3618 - /sp
[14:14:57] 404 - 4128 - /ssl_check
[14:14:57] 404 - 4128 - /sslvpn
[14:14:57] 404 - 4128 - /sso
[14:14:57] 404 - 4128 - /staff_directory
[14:14:57] 403 - 3638 - /sony
[14:14:57] 403 - 3658 - /star
[14:14:57] 403 - 3668 - /start
[14:14:57] 403 - 3668 - /start

```

3) Crawl site to download visible content (non-recursive first)

Use `wget` carefully — recursive crawl can download a lot.

```

PS C:\Users\Nidhi [REDACTED] Invoke-WebRequest -Uri
"https://owasp.org/www-project-juice-shop/" -OutFile
"day3_webcontent/downloads/index.html"

```

The screenshot shows the OWASP Juice Shop homepage. At the top, there's a navigation bar with links for 'PROJECTS', 'CHAPTERS', 'EVENTS', 'ABOUT', and a search icon. Below the navigation is a main header with the OWASP logo and links for 'Store', 'Donate', and 'Join'. The main content area features a large banner for the OWASP Foundation. Below the banner, there's a section titled 'OWASP Juice Shop' with a navigation menu including 'Main', 'Overview', 'News', 'Challenges', 'Learning', 'CTF', 'Ecosystem', and 'Supporters'. A prominent yellow house icon is displayed, along with social media links for GitHub, LinkedIn, and Twitter. A GitHub badge indicates 'release v19.0.0' and 'GitHub 12k'. Below the main content, there's a 'About Us' section with a detailed description of the application's purpose and a note about cookie usage. A sidebar on the right contains sections for 'Project Information', 'Classification' (Flagship Project, Tool), 'Audience' (Builder, Breaker, Defender), and 'Installation' (Accept, From Source, Packaged (GitHub/SourceForge)).

Record: saved documents (PDFs, DOCX), HTML pages, resource URLs. Put filenames into day3_webcontent/scans/download_list.txt.

4) Google Dorking — targeted searches to find exposed docs

Use these dorks in Google (and Bing). Replace <https://owasp.org/www-project-juice-shop/> or try without domain to find loose exposures.

site:owasp.org filetype:pdf

Google search results for "site:owasp.org filetype:pdf". The results include:

- OWASP Combined Financial Statements December 31, 2016 (PDF)
- owasp Combined Financial Statements December 31, 2016 (PDF)
- owasp Combined Financial Statements December 31, 2019 (PDF)
- owasp the owasp foundation, inc. and affiliate combined financial ... (PDF)

site:owasp.org filetype:xls OR filetype:xlsx OR filetype:csv

Google search results for "site:owasp.org filetype:xls OR filetype:xlsx OR filetype:csv". The results include:

- owasp Profit and Loss Detail (XLS)

*In order to show you the most relevant results, we have omitted some entries very similar to the 1 already displayed.
If you like, you can repeat the search with the omitted results included.*

site:owasp.org intext:"password" OR intext:"confidential" OR intext:"Internal Use Only"

Google site:owasp.org intext:"password" OR intext:"confidential" OR intext:"Internal Use Only"

All Mode All Images Videos Short videos Shopping Forums More Tools

owasp https://owasp.org / latest > 04-Authentication_Testing :

Testing for Vulnerable Remember Password

WSTG - Latest Testing for Vulnerable Remember Password Summary Credentials are the most widely used authentication technology.

OWASP https://owasp.org / 04-Authentication_Testing > 07-Test... :

Testing for Weak Password Policy

The simplest and cheapest of these is the introduction of a strong password policy that ensures password length, complexity, reuse and aging.

OWASP https://owasp.org / www-community / attacks > Passwor... :

Password Spraying Attack

Password spraying is a type of brute force attack. In this attack, an attacker will brute force logins based on list of usernames with default passwords on the ...

OWASP https://owasp.org / latest > 04-Authentication_Testing :

Testing for Weak Password Change or Reset Functionalities

The password reset process provides an alternative mechanism to access a user's account, and so should be at least as secure as the usual authentication process ...

site:owasp.org intitle:"index of" "parent directory"

Google site:owasp.org intitle:"index of" "parent directory"

All Mode All Images Videos Shopping Short videos Forums More Tools

No results found for site:owasp.org intitle:"index of" "parent directory".

Results for site:owasp.org intitle: index of parent directory (without quotes):

OWASP https://owasp.org / OWASP_Testing_Guide_v4 PDF :

Testing Guide

by T Share — The Open Web Application Security Project (OWASP) is a worldwide free and open community focused on improving the security of application software.
224 pages

OWASP https://owasp.org / assets / OWASP_Code_Revie... PDF :

CODE REVIEW GUIDE

It is common knowledge that more secure software can be produced and developed in a more cost effective way when bugs are detected early on in the systems ...
220 pages

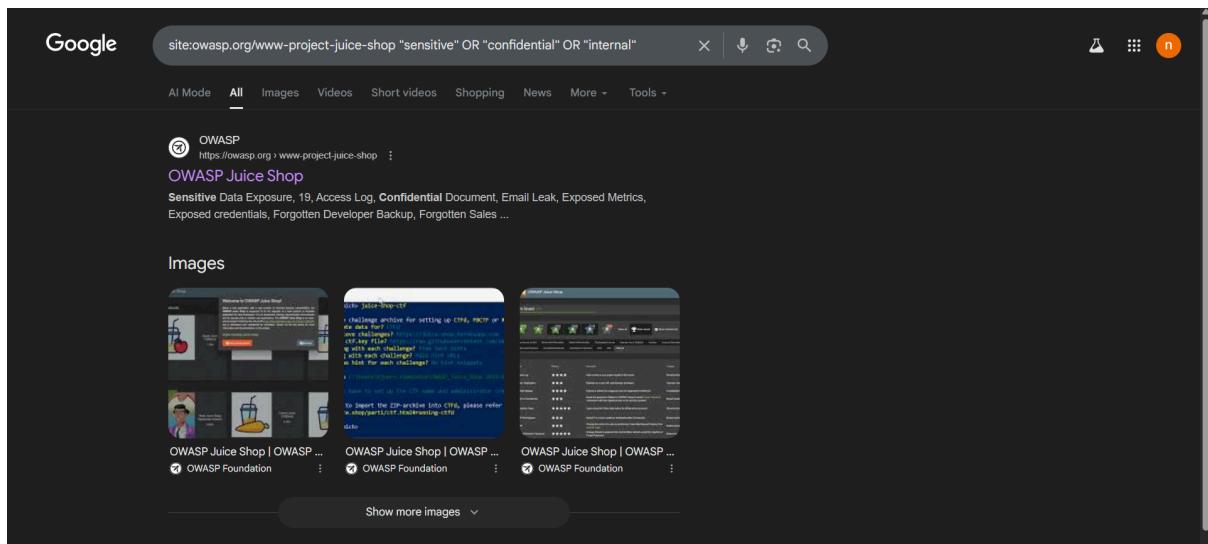
Missing: intitle: | Show results with: intitle:

OWASP https://owasp.org / assets / archive / OWASP_Tes... PDF :

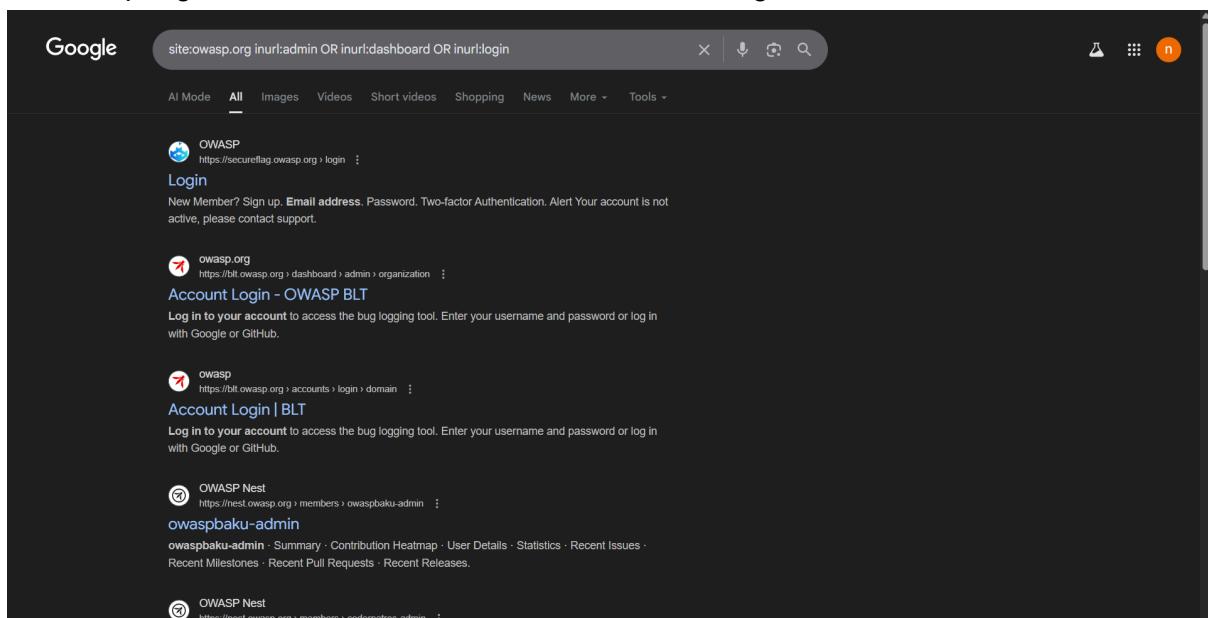
OWASP Testing Guide v2

... intitle operator is possible to find pages that contain "index of. Index of /backup/. N. Parent Directory. 21-Jul-2004 17:48 ~ Test: "Login to Webmin" inurl ...
272 pages

site:owasp.org/www-project-juice-shop "sensitive" OR "confidential" OR "internal"



site:owasp.org inurl:admin OR inurl:dashboard OR inurl:login

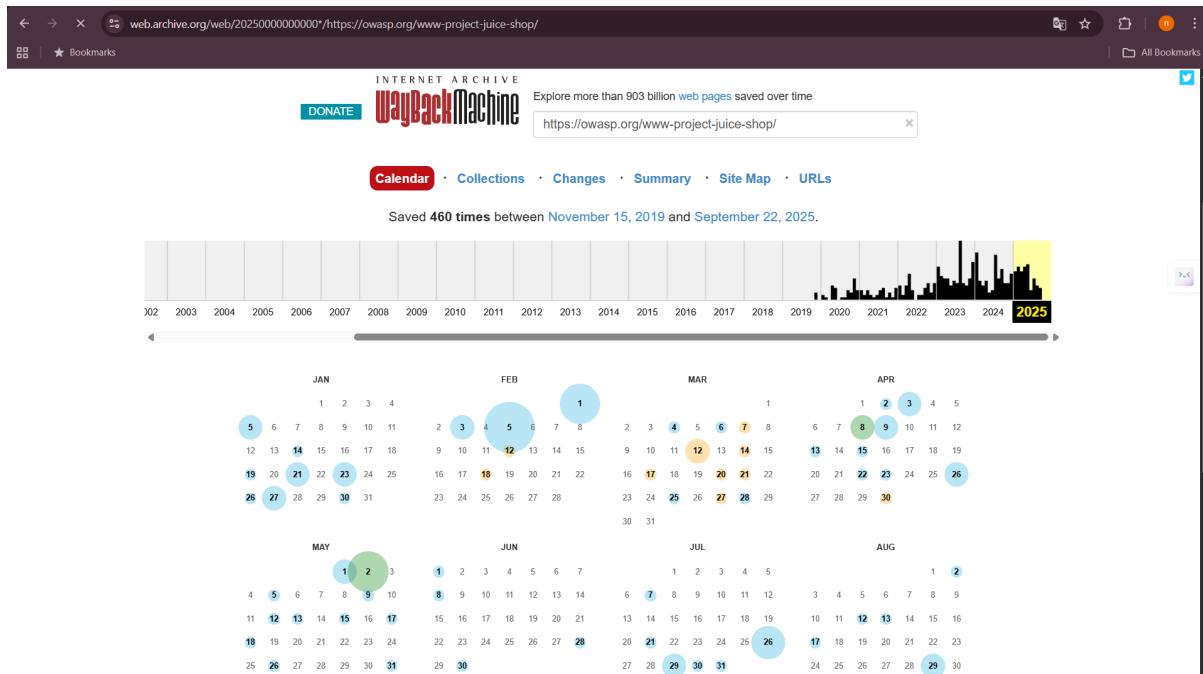


5) Wayback Machine (archive.org) — historical content

Use the Wayback web UI or the `waybackpy` / `waybackpack` tools.

Browser:

- Open <https://web.archive.org>
- Enter https://web.archive.org/web/20190111000000*/https://owasp.org/www-project-juice-shop/ → browse snapshots
→ look for older pages, PDFs, comments, leaked configs.



6) Inspect downloaded documents for sensitive content

- For PDFs: open and search for keywords (`password`, `credential`, `confidential`, `AWS`, `API_KEY`).

```
Select-String -Path "C:\Users\Nidhi\████████\day3_webcontent\evidence\OWASP Juice Shop _ OWASP Foundation.html" -Pattern
```

```
"password|apiKey|token|Authorization|mailto:|key=" -CaseSensitive:$false
```

```
PS C:\Users\Nidhi\████████> Select-String -Path "C:\Users\Nidhi\████████\day3_webcontent\evidence\OWASP Juice Shop _ OWASP Foundation.html" -Pattern "password|apiKey|token|Authorization|mailto:|key=" -CaseSensitive:$false
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:822:<a href="mailto:björn.kimminich@owasp.org">Björn Kimminich</a> and is developed,
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1059:<td><small>CAPTCHA Bypass, Extra Language, Multiple Likes, Reset Morty's Password</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1075:<td><small>Björn's Favorite Pet, Change Bender's Password, GDPR Data Erasure, Login Björn, Password Strength, Reset Bender's Password, Reset Björn's Password, Reset Jim's Password, Two Factor Authentication</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1123:<td><small>Access Log, Confidential Document, Email Leak, Exposed Credentials, Forgotten Developer Backup, Forgotten Sales Backup, GDPR Data Theft, Leaked Access Logs, Leaked Unsafe Product, Login Amy, Login MC SafeSearch, Meta Geo Stalking, Misplaced Signature File, NFT Takeover, Reset Uvogin's Password, Retrieve Blueprint, Visual Geo Stalking</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1177:<td><small>Bully Chatbot, CAPTCHA Bypass, Extra Language, Login Support Team, Password Strength, Reset Morty's Password
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1254:<td><small>Björn's Favorite Pet, Leaked Access Logs, Leaked Unsafe Product, Local File Read, Login Amy, Login MC SafeSearch, Meta Geo Stalking, Reset Bender's Password, Reset Björn's Password, Reset Jim's Password, Reset Morty's Password, Reset Uvogin's Password, Supply Chain Attack, Visual Geo Stalking, Vulnerable Library
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1287:<td><small>Bonus Payload, DOM XSS, Forged Feedback, Login Admin, Login Bender, Login Jim, Password Strength, Privacy Policy, Reflected XSS, Score Board, View Basket
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1412:<td style="min-width: 190px"><a href="https://web.archive.org/web/20250528084243/https://demo.owasp.juice.shop/#/hacking-instructor?challenge=Password%20Strength" target="_blank">Password Strength</a></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1487:<td><small>Reset Morty's Password</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1493:<td><small>Björn's Favorite Pet, Password Strength, Reset Bender's Password, Reset Björn's Password, Reset Jim's Password</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1523:<td><small>Access Log, Confidential Document, Exposed Metrics, NFT Takeover, Reset Uvogin's Password</small></td>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1559:<li><small><a href="https://web.archive.org/web/20250528084243/https://cheatsheetsseries.owasp.org/cheatsheets/Authorization_Cheat_Sheet.html" target="_blank">Authorization Cheat Sheet</a></small></li>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1599:<li><small><a href="https://web.archive.org/web/20250528084243/https://cheatsheetsseries.owasp.org/cheatsheets/Forgot_Password_Cheat_Sheet.html" target="_blank">Forgot Password Cheat Sheet</a></small></li>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:1611:<li><small><a href="https://web.archive.org/web/20250528084243/https://cheatsheetsseries.owasp.org/cheatsheets/JSON_Web_Token_for_Java_Cheat_Sheet/a></small></li>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:2049:<li><a href="mailto:björn.kimminich@owasp.org">Björn Kimminich</a></li>
day3_webcontent\ evidence\OWASP Juice Shop _ OWASP Foundation.html:2050:<li><a href="mailto:jannik.hollenbach@owasp.org">Jannik Hollenbach</a></li>
```