

ArchiveTeam Warriorを始めよう！ お手軽Webアーカイブ ボランティア

stepney141 (ネクストチューター)

セキュリティ・キャンプ全国大会2025 LTセッション

2025年 8月 14日

※ほぼ同じ内容は <https://stepney141.hatenablog.com/entry/2025/02/17/182148> で読めます

お前は誰？



素性

- 東京大学大学院 総合文化研究科
- 研究テーマ：ゲーム情報学, 完全情報ゲームの複雑性推定
 - セキュリティは特に専門分野ではありません

趣味

- ボードゲームエンジンの開発
- 関数電卓プログラミング
- 魔術（オカルティズム）の歴史
- Webアーカイブ

スマホをぶん投げたくなる
瞬間といえバ？

コレ!!!!!!



ちょっと真面目な話

消滅したサイト … Yahoo! ジオシティーズ (2019)



サービス終了のお知らせ

Yahoo!ジオシティーズは終了しました

2019年3月31日をもちましてYahoo!ジオシティーズのサービス提供を終了いたしました。長らくご愛顧いただき誠にありがとうございました。

ホームページをお持ちのお客様につきましては、2020年3月31日までFTPによるファイルダウンロードのみご利用可能となっております。ホームページやドメインの移行方法などは[サービス終了のお知らせ](#)をご確認ください。

Yahoo! JAPAN

消滅したサイト … 米国連邦政府の各種データベース (2025)

※ この話へ興味のある人へ → <https://govdiff.com/> を見よう



<https://www.bloomberg.co.jp/news/articles/2025-08-13/T0WWN0GP9VD000>

今年、世界から消滅した/する予定のオンラインサービス

2025-05-31: The Samsung Galaxy Watch App Store for Tizen OS watches will stop allowing downloads of new content.^[90] 2025-05-30 12:00 PDT (UTC-7): Fauna^[IA+Wiki+today+MemWeb] a hosted database service.^{[100][101]} 2025-05-29: Nexon will discontinue MapleStory 2^[IA+Wiki+today+MemWeb] and shut down the last servers in Korea.^[102] 2025-05-28: Japanese food service/app/7 Livdoor Gourmet^[IA+Wiki+today+MemWeb] will shut down its service gradually until termination on May 28th.^[103] 2025-05-21: Penny Arcade^[IA+Wiki+today+MemWeb] has decided to divest resources from its forums^[IA+Wiki+today+MemWeb]. The community migrated to Coin Return^[IA+Wiki+today+MemWeb], but the content did not. 2025-05-21 is the date PAs contract with Vanilla runs out.^[104] 2025-05-20: The Hopscotch Forum^[IA+Wiki+today+MemWeb] will be permanently closed, after being made read-only on December 21, 2024. 2025-05-19: Twitch will implement a 100-hour storage limit on highlights and uploads and will delete the least popular content for accounts exceeding that limit.^[105] 2025-05-16: Read.cu^[IA+Wiki+today+MemWeb] and Posts.cu will shut down, following switching to read-only. 2025-05-15: Flickr will restrict downloads of the original and large (larger than 1024pi) sizes of non-CC-BY. 2025-05-05: Skype will shut down.^[106] 2006-06-01: Hayabusu Owners Group^[IA+Wiki+today+MemWeb] (Hayabusu.ORG), due to a sharp decline in a 2025-05-01^[109] A few days before that, a new announcement was made that the site would stay online fo 2025-05: Dailymotion^[IA+Wiki+today+MemWeb] will delete "inactive" videos.^{[110][112]} 2025-05: Nete.com^[IA+Wiki+today+MemWeb] will most likely shut down.^[113] 2025-04-30: コーエー (Coroem) 2025-04-30: openSt^[IA+Wiki+today+MemWeb], a genetic sharing website, will shut down citing a rise in fa 2025-04-30: XZY2^[IA+Wiki+today+MemWeb], an NFT marketplace, will shut down.^[115] 2025-04-30: Yoggsta 2025-04-27: Ovan^[IA+Wiki+today+MemWeb] 2025-04-22: fannbo 2025-04-19: favoloo 2025-04-19: Skatpet 2025-04-03: octoid 2025-04-02 15:00 P of creators' content 2025-04-01: plai.a 2025-04-01: utme.u discontinued.^[124] - April: Snuffishop G - Unknown: http://rocket3.net/^[IA+Wiki+today+MemWeb], a Japanese w - Unknown: arcavie^[IA+Wiki+today+MemWeb] In April 2025, the model's profitability, external pressure, and operational fatigue—effectively exit. - Unknown: https://webb-site.com/^[IA+Wiki+today+MemWeb] (David W site.com/articles/shutdown.asp^[IA+Wiki+today+MemWeb] - End of the year: NS Community^[IA+Wiki+today+MemWeb], the forums 1.120 - End of the year: horaro.org^[IA+Wiki+today+MemWeb], an event schea at least through 2026.^[121] 2025-12-ends: https://www.greun.ru/^[IA+Wiki+today+MemWeb] Oren i 2025-12-31: Toy collector site shmax is retiring its forums^[IA+Wiki+today+MemWeb] 2025-12-17: ㅂㅂ^[IA+Wiki+today+MemWeb] (Buttoon), a Korean we 2025-12-01: Evbox Everon^[IA+Wiki+today+MemWeb] Engle stops serv 2025-11-30 23:59 JST (UTC+9): KCN Kyoto will delete kint-tv.ne.jp home pages.^[127] 2025-11-18: gooブログ^[IA+Wiki+today+MemWeb], one of Japan's largest blog services run by NTT since 2004, will be shutting down 2025-11-18. New blogs can't be created after 2025-07-01. Posting new articles will no longer be possible after 2025-10-01.^[128] 2025-11: Allegro Archive^[IA+Wiki+today+MemWeb] 2025-08 will become read-only, redirect to the main site 2025-11 and shut down 2026-03.^[129] - End of October: Launchpad^[IA+Wiki+today+MemWeb] will shut down its mailing lists. The content will be removed. There will likely be a read-only date in August. Canonical sent us a list of mailing lists.^[130] 2025-10-28: Dropbox is shutting down its password manager (Dropbox Passwords).^[131] 2025-10-10: Ex-Christan^[IA+Wiki+today+MemWeb], a forum for former Christians, will shut down.^[132] - July to September: iTeos Therapeutics^[IA+Wiki+today+MemWeb] will wind down operations in 2025 third quarter.^[133] - Early September: OSzone.net^[IA+Wiki+today+MemWeb] will close after being made read-only on 2025-08-01.^[134] 2025-09-30: AOL will shut down its dial-up internet service and its associated software AOL Shield Browser and AOL Dialer.^{[135][136]} 2025-09-30: @niftyホームページサービス (LaCooCan)^[IA+Wiki+today+MemWeb], a Japanese web hosting service, will discontinue its free plan.^[137] 2025-09-30: The Samsung Galaxy Watch App Store for Tizen OS watches will stop allowing downloads of previously purchased content.^[138] 2025-09-30 11:00 JST (UTC+9): gooランキング (goo ranking)^[IA+Wiki+today+MemWeb] will be discontinued.^[139] 2025-09-17: oshiete.goo.ne.jp^[IA+Wiki+today+MemWeb] Japanese Q&A site "oshiete.goo.ne.jp" will be terminated.^[140] - By 2025-09-11: Elistasi forum^[IA+Wiki+today+MemWeb] is being shut down in favour of Discord.^[141]	2025-03-31: sagakudo100.g^[IA+Wiki+today+MemWeb], a book info service, will end on March 31, 2025.^[146] 2025-03-31: LINE's notification service LINE Notify^[IA+Wiki+today+MemWeb] closes.^[151] 2025-03-31: 魔法の国A/C (Maho no Kuni) ^[IA+Wiki+today+MemWeb], a Japanese web novel site active since 1999, will shut down. A work export function was launched on 2024-09-26; users are encouraged to import their works to 魔法の国A/C (Maho no Kuni) ^[IA+Wiki+today+MemWeb], another site owned by the same publisher.^[152] 2025-03-31: NTT trials will remove public access to home pages.^[153] 2025-03-31: So-net Blog OSS Blog, the former Sony Communications blog hosting service since 2004, will end its services on 31 March 2025.^[154] 2025-03-31: Spills.jp^[IA+Wiki+today+MemWeb], a speedrun analysis and sharing website, will shut down.^[155] 2025-03-31: 7-11 ^[IA+Wiki+today+MemWeb], formerly Shomashi Japan, and OSDN^[IA+Wiki+today+MemWeb], formerly SourceForge.JP, After a change in operating company, the - ended, and the servers will now be shut down.^[156] - Japanese webhost, is ending access to all webistes hosted its old free (Star Server Free) service as it migrates to all customers new service.^[157] - net A&T), a Japanese internet radio station, closes.^[158] ...^[IA+Wiki+today+MemWeb] (Yahoo! Partner), Yahoo! Japan's dating platform, will shut down.^[140] ...^[IA+Wiki+today+MemWeb] ...^[IA+Wiki+today+MemWeb] Due to the lack of a new administrator, "Words You Shouldn't Search For" Wiki will close on 2025/03/20.^[145] ...^[IA+Wiki+today+MemWeb] will shut down due to "losing operational costs".^[142] ...^[IA+Wiki+today+MemWeb] and the Microsoft^[IA+Wiki+today+MemWeb] platform will close due to the UK Online Safety Act taking effect the next day.^[143] ...^[IA+Wiki+today+MemWeb] closes.^[144] - inate its metaverse platform Nexon Town^[IA+Wiki+today+MemWeb] (145) - inactive accounts on its Q&A platform Zeggyta Qnq^[IA+Wiki+today+MemWeb] (146) - once largest Spanish-speaking videogame forum, will shut down.^[147] ...^[IA+Wiki+today+MemWeb] will also shut down its remaining Korean servers on March 6th.^[148] ...^[IA+Wiki+today+MemWeb] will be shut down following its acquisition by Getty Images in 2024.^[149] ...^[IA+Wiki+today+MemWeb] ...^[IA+Wiki+today+MemWeb] will delete almost all data due to the UK Online Safety Act 2023.^[150] ...^[IA+Wiki+today+MemWeb] will shut down (again). Their social media accounts will remain online.^[151] - ix has discontinued its forums feature and will delete data from existing sites.^[152] 59 EDT (UTC-4): The old Dungeons & Dragons Online forums^[IA+Wiki+today+MemWeb] that had been made read-only in 2023 will be shut down.^[143] ...^[IA+Wiki+today+MemWeb] a Savage Fismes enthusiast forum, will shutdown.^[144] ...^[IA+Wiki+today+MemWeb] the play Sketch LIVE streaming service will be discontinued on August 27, 2025.^[145] ...^[IA+Wiki+today+MemWeb] will stop serving existing links.^[146] Other websites that use Firebase Dynamic Links will be affected.^[147] ...^[IA+Wiki+today+MemWeb] This year will be the last iteration of this festival, as the organizer has run out of money.^[148] ...^[IA+Wiki+today+MemWeb], operator of a couple dozen local newspapers^[IA+Wiki+today+MemWeb] in the USA, closed with no advance notice.^[149] ...^[IA+Wiki+today+MemWeb], the digital store by comics publisher Dark Horse, will shut down, following their discontinuation of sales on 2025-02-24.^[150] ...^[IA+Wiki+today+MemWeb] - game developer Frontier is closing the forums for their game Planet Coaster 2^[IA+Wiki+today+MemWeb] (52) ...^[IA+Wiki+today+MemWeb], a Japanese web hosting service, will shut down.^[53] ...^[IA+Wiki+today+MemWeb] will be closed.^[54] 2025-07-31: E.g^[IA+Wiki+today+MemWeb], a Spanish game forum will close.^[55] Archiveweb was unable to save it due to use of cloudflare with TLS fingerprinting, although some users have reported success with manual archiving attempts via archive.today and Wayback.com (i.e. Archiveweb page). 2025-07-28: https://umakaba.com/^[IA+Wiki+today+MemWeb] will shut down. It became read-only on 2025-06-30.^[56] 2025-07-08: Olitch will shut down their user content web hosting. While the announcement is phrased as a 'change', this is effectively a full shutdown.^[57] 2025-07-08: Pocket^[IA+Wiki+today+MemWeb], by Mozilla, will shut down.^{[58][59]} 2025-07-04: Neosping^[IA+Wiki+today+MemWeb] is shutting down in favor of its successor, https://netratto.com/^[IA+Wiki+today+MemWeb] (60) 2025-07-01: Happy Kids^[IA+Wiki+today+MemWeb] elementary school will close.^[61] 2025-07-01: Tinkle Squid^[IA+Wiki+today+MemWeb] elementary school will close doors.^[62] 2025-07-01: Microsoft Maps App will be removed from the Microsoft Store by July 2025.^[63] 2025-07-01: OBS De Pijpeler^[IA+Wiki+today+MemWeb] will close doors.^[64] 2025-07-01: PerTitle Art^[IA+Wiki+today+MemWeb], a Japanese social network for sharing AI pictures, will shut down.^[65] 2025-07-01: Sticking Anzai^[IA+Wiki+today+MemWeb] and Sticking Coderz^[IA+Wiki+today+MemWeb] elementary schools will probably merge.^[66] 2025-07-01: Vrijeshool Wavur^[IA+Wiki+today+MemWeb] elementary school will close, due to lack of schoolchildren.^[67] 2025-07-01: Backdrops.org^[IA+Wiki+today+MemWeb], a WordPress farm or "tjion-commercial, antifascist, antiracist, privacy-oriented blog platform for radical activists" will shut down.^[68] 2025-07-01: Fakepost^[IA+Wiki+today+MemWeb], by Mozilla, will shut down. 2025-06-30: Yae-Nj^[IA+Wiki+today+MemWeb] will be closed and merged with NJS.^[69] 2025-06-30: History West Midlands^[IA+Wiki+today+MemWeb] will be closed. 2025-06-30: ES Shochiku Tokyo^[IA+Wiki+today+MemWeb] will be closed.^[70]
--	--

<https://wiki.archiveteam.org/index.php/Deathwatch>

Web上に情報が残らずに何が困るのか？

知識の喪失

- 有益な知識が失われる
- ex. シンプルにもものすごい情報量の個人サイト

文化圏の喪失

- Web上コミュニティの拠り所が失われる
- ex. サブカル作品の二次創作サイト, 著名人のブログ

「データにアクセスする自由」の喪失

- **まず何よりこれが許せなくないですか？？？**

世界から悲劇をなくすために
私たちができることは？

Archive Team とは？

- Webアーカイブを専門に行う有志団体
- 世界中の様々なWebサービスの終了情報を集め,
自前のツール でクローल & 保存している
- 収集したデータをInternet Archiveに提供している

→ この「自前のツール」の1つが“Warrior”



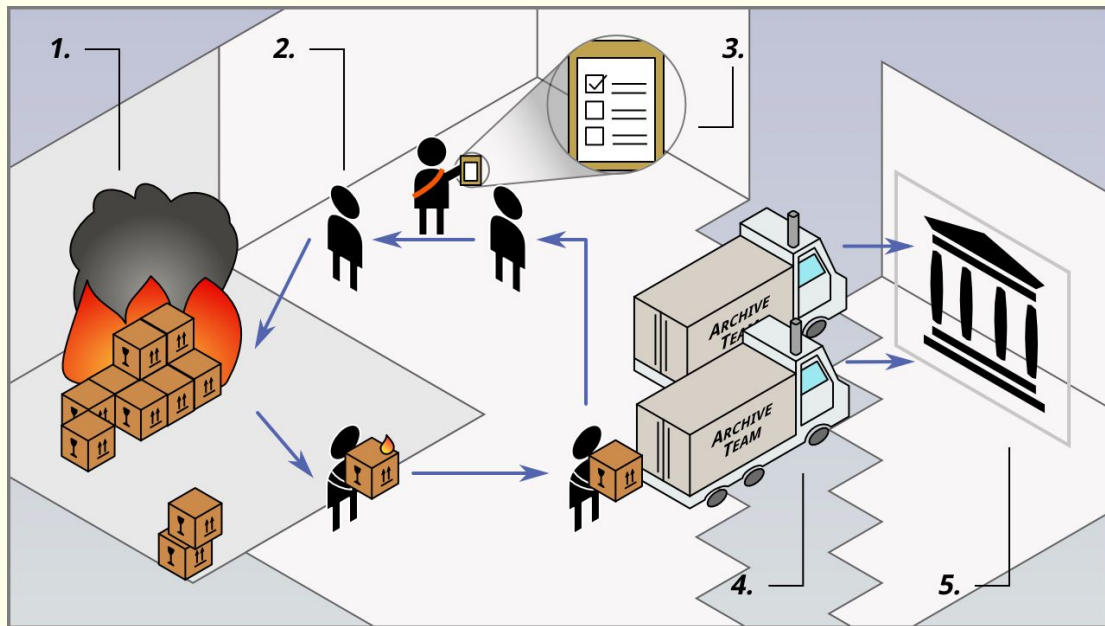
Archive Team Warrior とは？

- Archive Teamが開発・運用する分散クロール環境
- 収集したデータをInternet Archiveに提供している
- Warriorで保存されたWebサイトの例
 - Yahoo! ジオシティーズ
 - FC2 WEB
 - Goo辞書 etc…
- **誰でも VM / Docker で簡単に動かせる**



Warriorによる分散クローリングのしくみ

1. 消えそうなサイトを...
2. Warriorが保存！
3. ジョブマネージャが
タスクを割り振る
4. データを整理
5. Internet Archiveへ



どうやって動かせばいい？

Docker Composeで動かします

30億のデバイスで走るcompose.yaml

```
services:
  archiveteam-watchtower:
    container_name: archiveteam-watchtower
    image: containrrr/watchtower
    labels:
      - com.centurylinklabs.watchtower.enable=true
      - com.centurylinklabs.watchtower.scope=archiveteam-warrior
    volumes:
      - '/var/run/docker.sock:/var/run/docker.sock'
    command: '--label-enable --cleanup --interval 3600 --scope archiveteam-warrior'
    restart: unless-stopped

  archiveteam-warrior:
    container_name: archiveteam-warrior
    image: atdr.meo.ws/archiveteam/warrior-dockerfile
    environment:
      - DOWNLOADER=stepney141 # Change this to your nickname
      - SELECTED_PROJECT=auto
      - CONCURRENT_ITEMS=6
    stop_signal: SIGINT
    stop_grace_period: 5m
    labels:
      - com.centurylinklabs.watchtower.enable=true
      - com.centurylinklabs.watchtower.scope=archiveteam-warrior
    ports:
      - '127.0.0.1:8001:8001'
    restart: always
```


とはいえWarriorにも限界はある

- あくまで有志団体なのでリソースは有限 → サ終に間に合わない場合も
- 最近ではAIの影響でクロールを明示的に禁止する動きも増えてきてる
- Internet Archiveの検索機能はだいぶアレ
- そもそもWayback Machineって持続可能ですか？
 - cf. 2024年の大規模サイバー攻撃事案

とはいえWarriorにも限界はある

しかし、全く何もやらないよりは遥かにマシ！

- めくまで有念団体なのでウブースは有限→ウ終に間に合わない場合も
- 最近ではAIの影響
- Internet Arch
- そもそもWay
- cf. 2024



伝えたいこと

情報を保全することも
一種の「セキュリティ」では？

(奇跡的に時間が余ったら)デモ
