

ArchiveTeam Warriorで始める 分散並列Webアーカイブ入門

stepney141

Kernel/VM探検隊@関西 No.12

2026年 5月 30日

お前は誰？



素性

- 基盤系SE見習い (新卒研修中) @某大手SIer
- ゲームAI, ゲーム状態の到達可能性問題 @某大学院
- Twitter (現 X): “ゆみや” @stepney141

趣味

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

今日は自己防衛の話をします



皆さんの悲しい瞬間といえは？

悲しい瞬間の例



悲しい瞬間の例



このサイトにアクセスできません

etudiant.epita.fr のサーバーの IP アドレスが見つかりませんでした。

次をお試しください

- 接続を確認する
- [プロキシ、ファイアウォール、DNS の設定を確認する](#)

ERR_NAME_NOT_RESOLVED

詳細

再読み込み

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>

2025年, 世界から消滅したオンラインサービス(一部)

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 Gourme^[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 1024p) sizes of non-CC-BY. 2025-05-05: Skype will shut down.^[106] 2006-06-01: Hayabusa Owners Group^[IA+Wiki+today+MemWeb] (Hayabusa.ORG), due to a sharp decline in and 2025-05-01^[109] A few days before that, a new announcement was made that the site would stay online for 2025-05: Dailymotion^[IA+Wiki+today+MemWeb] will delete "inactive" videos.^{[110][112]} 2025-05: Note.com^[IA+Wiki+today+MemWeb] will "most likely" shut down.^[113] 2025-04-30: CODE - (Coromoo) 2025-04-30: openStar^[IA+Wiki+today+MemWeb], a generic sharing website, will shut down citing a "rise in fa 2025-04-30: Xyz2^[IA+Wiki+today+MemWeb], an NFT marketplace, will shut down.^[115] 2025-04-30: Yoggast 2025-04-27: Ovan^[IA+Wiki+today+MemWeb] 2025-04-22: fannio 2025-04-19: favelog 2025-04-19: Skatet 2025-04-03: octoid 2025-04-02 15:00: R 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 01.20] • End of the year: horaro.org^[IA+Wiki+today+MemWeb], an event sched at least through 2026.^[21] 2025-12-end: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] 2025-12-31: www.buffoon.net/ (Bufftoon), 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.^[28] 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.^[29] - End of October: Launchpad^[IA+Wiki+today+MemWeb] will shut down its mailing lists. The data will be removed. There will likely be a read-only date in August. Canonical sent us a list of mailing lists.^[30] 2025-10-28: Dropbox is shutting down its password manager (Dropbox Passwords).^[31] 2025-10-10: Ex-Christan^[IA+Wiki+today+MemWeb], a forum for former Christians, will shut down.^[32] - July to September: iTeos Therapeutics^[IA+Wiki+today+MemWeb] will wind down operations in 2025 third quarter.^[33] - Early September: OSzone.net^[IA+Wiki+today+MemWeb] will close after being made read-only on 2025-08-01.^[34] 2025-09-30: AOL will shut down its dial-up internet service and its associated software AOL Shield Browser and AOL Dialer.^{[35][36]} 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.^[37] 2025-09-30: The Samsung Galaxy Watch App Store for Tizen OS watches will stop allowing downloads of previously purchased content.^[38] 2025-09-30 11:00 JST (UTC+9): gooランキング (goo ranking)^[IA+Wiki+today+MemWeb] will be discontinued.^[39] 2025-09-17: oshiete.goo.ne.jp^[IA+Wiki+today+MemWeb] Japanese Q&A site "oshiete.goo.ne.jp" will be terminated.^[40] - By 2025-09-11: Elistasi forum^[IA+Wiki+today+MemWeb] is being shut down in favour of Discord.^[41]	2025-03-31: sakakudo100.g^[IA+Wiki+today+MemWeb], a book info service, will end on March 31, 2025.^[150] 2025-03-31: LINE's notification service LINE Notify^[IA+Wiki+today+MemWeb] closes.^[151] 2025-03-31: 魔の5.4.4 (Maho no 5.4.4)^[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 7273A (Kakuyoku)^[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: Splice^[IA+Wiki+today+MemWeb], a spectrum analysis and sharing website, will shut down.^[155] 2025-03-31: 7273A (Kakuyoku)^[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 webhosters' old free (Star Server Free) service as it migrates to all customers new service.^[157] - ref A241), 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.^[142] ...^[IA+Wiki+today+MemWeb] will shut down due to "rising operational costs".^[143] ...^[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.^[145] ...^[IA+Wiki+today+MemWeb] closes.^[144] - inate its metaverse platform Nexon Town^[IA+Wiki+today+MemWeb].^[146] - inactive accounts on its Q&A platform Zaypt^[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] >1: Q&A platform Retrospring 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] - avage: Shooze^[IA+Wiki+today+MemWeb], a Savage Fitness enthusiast forum, will shutdown.^[144] - etchy.paly.net^[IA+Wiki+today+MemWeb], the paly Sketch LIVE streaming service will be discontinued on August 27, 2025.^[145] - oo.g), Google's URL Shortener, will stop serving existing links.^[146] Other websites that use Firebase Dynamic Links will be affected.^[147] - ewarmpo Festival^[IA+Wiki+today+MemWeb] This year will be the last iteration of this festival, as the organizer has run out of money.^[148] - news Media Corporation^[IA+Wiki+today+MemWeb], operator of a couple dozen local newspapers^[IA+Wiki+today+MemWeb] in the USA, closed with no advance notice.^[149] - er: Dark Horse Digital^[IA+Wiki+today+MemWeb], the digital store by comics publisher Dark Horse, will shut down, following their discontinuation of sales on 2025-02-24. - nger guaranteed after 2025-03-30.^{[150][151]} - Dark Horse Frontier is closing the forums for their game Planet Coaster.^[IA+Wiki+today+MemWeb] ^[52] - Proge^[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] Archiwel 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.org (i.e. ArchiveWeb page). 2025-07-28: https://umakelab.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://hetratto.com/^[IA+Wiki+today+MemWeb].^[60] 2025-07-01: Happy Kids^[IA+Wiki+today+MemWeb] elementary school will close.^[61] 2025-07-01: Tivola Squ^[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 Pijp^[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 Anz^[IA+Wiki+today+MemWeb] and Sticking Coderz^[IA+Wiki+today+MemWeb] elementary schools will probably merge.^[66] 2025-07-01: Vjeshchool Wawur^[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: Yale-NUS^[IA+Wiki+today+MemWeb] will be closed and merged with NUS.^[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

文化圏の喪失

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

「データにアクセスする自由」の喪失

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非営利団体
- 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”



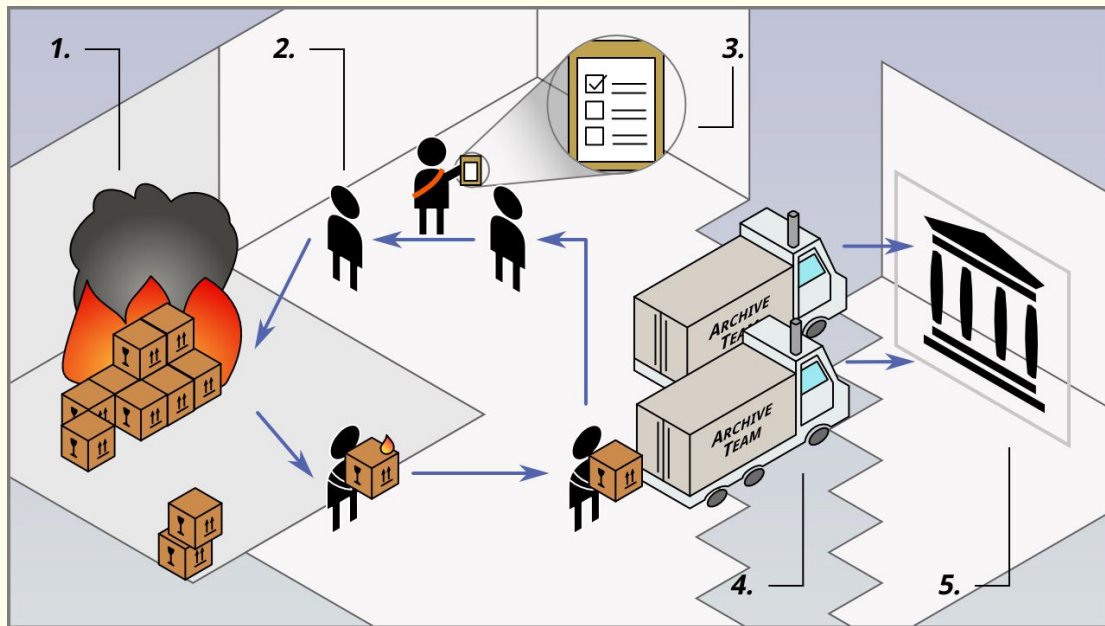
Archive Team Warrior とは？

- 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で大部分が保存されたWebサイトの例
 - Yahoo! ジオシティーズ
 - FC2 WEB
 - Goo辞書 etc…
- クライアントの VM / Docker 環境が整備されていて、誰でも簡単に動かせるようになっている



Warriorによる並列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？

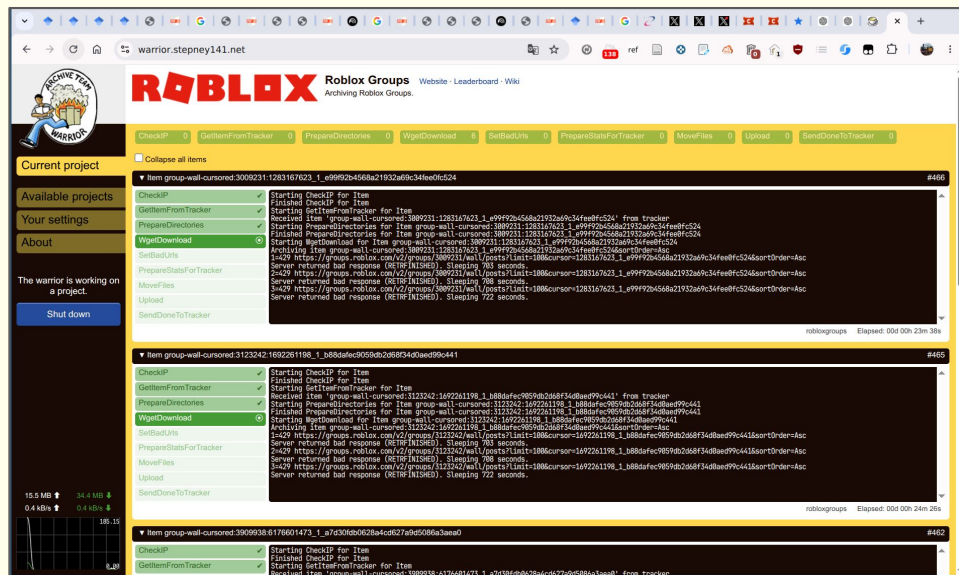
Docker Composeで動かします

ネットに繋がる計算機がある人は今すぐ動かそう！

```
services:
  archiveteam-watchtower:
    container_name: archiveteam-watchtower
    image: nickfedor/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
  networks:
    - warrior-net

archiveteam-warrior:
  container_name: archiveteam-warrior
  image: atdr.meo.ws/archiveteam-warrior-dockerfile
  environment:
    - DOWNLOADER=your_name # 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
  networks:
    - warrior-net

networks:
  warrior-net:
    name: warrior-net
```



<https://gist.github.com/stepney141/83306db94ee735b3155b9ed38dd6f891>

とはいえWarriorにも限界はあり, 先行きは不透明

- 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
 - 私の実体験: モデレータがバッチ流し忘れてたせいで全て消滅
- コンセプト的にみんなが動かしてくれないと意味がない
 - cf. 日本の個人サイトは海外アクセスを遮断してることが多い (GeoBlock)
- 生成AIへ対抗して任意のWebクローラを締め出す動き
- Internet Archiveの検索機能の貧弱さ
 - データを有益に活用するためには情報の適切な整理・検索が必須であるが...
- そもそもInternet Archiveって持続可能ですか？
 - cf. 2024年の大規模サイバー攻撃事案, 2026年のストレージ価格暴騰

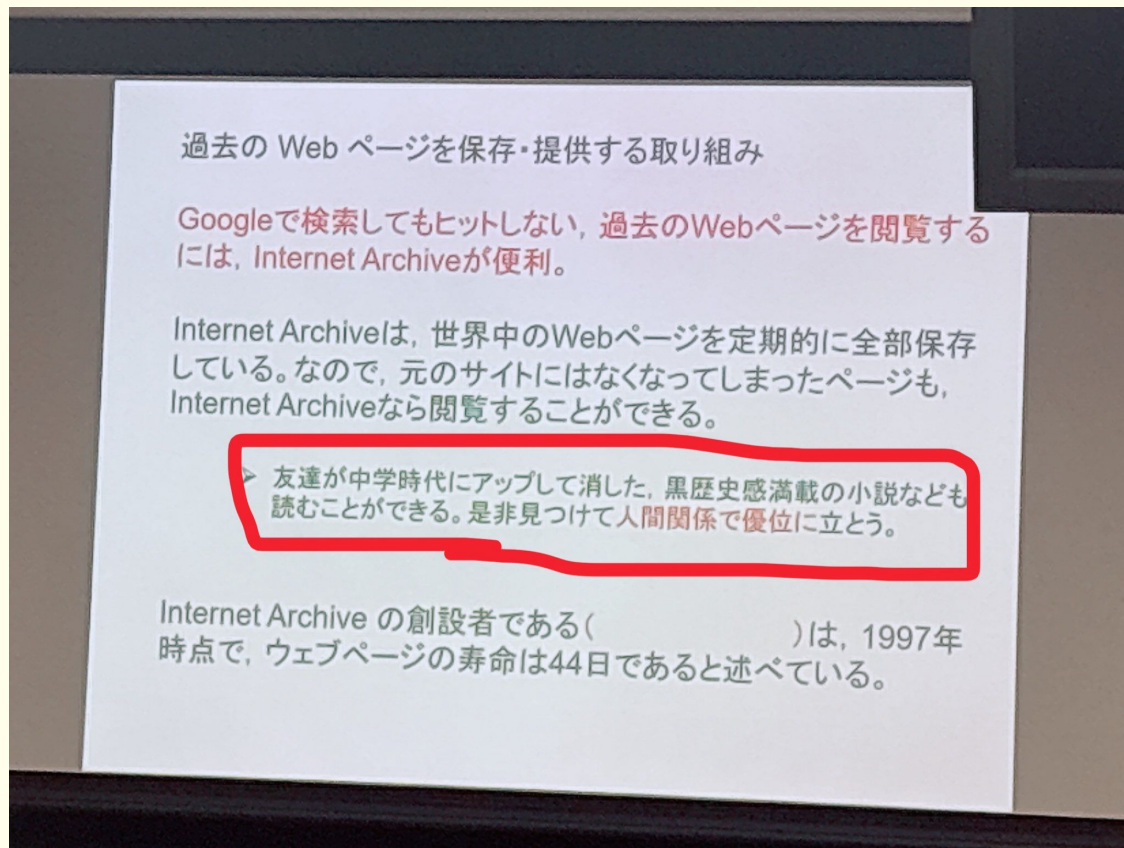
とはいえWarriorにも限界はあり, 先行きは不透明

しかし, 全く何もやらないよりは遥かにマシ！

- コンセプト的
○ cf. 日本のは
 - 生成AIへ対抗
○ データを有
 - Internet Arch
○ cf. 2024年
- 
- 誰も…
- 消防車を
呼んでいないので
ある!!!
- い (GeoBlock)
- が必須であるが...
- ジ価格暴騰

伝えたいこと

Internet Archiveに対する世間の(ナীবな)認識の例



情報を残すには自助努力が必要！！



(奇跡的に時間が余ったら)デモ
