Snapshot of Japan 2016-2018

An introduction to Stanford’s web archive on contemporary Japan

Regan Murphy Kao
Stanford University Libraries
March 21, 2019
Overview

- Why archive?
- What was archived?
- How was it archived?
- Conclusion: successes and challenges
Why archive?

- To preserve critical writings and artifacts
- To preserve broad spectrum of conversations
Wayback Machine

Explore more than 341 billion web pages saved over time

Tools
Wayback Machine Availability API
Build your own tools.
WordPress Broken Link Checker
Banish broken links from your blog.
404 Handler for Webmasters

Archive-It enables you to capture, manage, and search collections of digital content without any technical expertise or hosting facilities. Visit Archive-It to build and browse the collections.

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.
Archiving by subject specialists

- Single theme
- Event
- Type (blogs, news media, etc.)
Project’s aims?

- Provide a glimpse at a wide range of topics perceived as important at the moment of archiving
- Preserve a diversity of voices
- Compliment not duplicate other efforts
Topics archived

- **Politics & economics (54):** political parties, financial institutions, political leaders, TPP, security
- **Employment & labor issues (46):** labor unions, industrial cooperatives, irregular employment for youth, at-home work
- **Social issues (156):** aging society, Okinawa, support for women, separate surnames for married couples, food destruction, baby drop box, social withdrawal, international help, LGBTQ, influential social figures, revitalization of rural areas, preservation of old towns, 2020 Tokyo Olympics, blog rankings (politics)
- **Child care & pregnancy (29):** wait lists for child care, authorized nursery schools, egg freezing services
- **Academic & humanities (41):** higher education, religious organizations, university departments (law, economics, sociology)
- **Environment & nature (75):** resources, natural energy, anti-nuclear power generation, global warming, nuclear energy-related institutes
- **Science & technology (41):** robots, advanced materials, research institutes
- **Internet & computers (30):** net harassment, cyber security
Example of archived site

Topics & News

2017.02.15 グループ労組春闘F.A.Xニュースを更新しました
2017.02.14 春闘F.A.Xニュースを更新しました
2017.02.13 民主化闘争情報を更新しました
2017.02.07 機関紙を更新しました
2017.02.01 地域公共交通の課題と展望を更新しました
2017.01.24 JR九州「豊肥本線」の被災現場を視察
2017.01.17 青年・女性委員会機関紙を更新しました
2016.12.09 与党「2017税制改正大綱」にJR連合からの要望事項が反映！
2016.11.22 石井国土交通大臣への要請行動！
2016.11.22 2017税制改正に向け、高市総務大臣へ要請行動！（11月21日）
例：引用サイト

例：引用サイト
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>druid</td>
<td>sourceld</td>
<td>til:nonSctil:title</td>
<td>til:langu</td>
<td></td>
</tr>
<tr>
<td>mp548fy3403</td>
<td></td>
<td>自由民主党</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>qr548ck9856</td>
<td></td>
<td>民進党</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>my152ng8371</td>
<td></td>
<td>公明党</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>rz006pq6865</td>
<td></td>
<td>日本共産党</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>xh105zy8161</td>
<td></td>
<td>日本維新の会</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>sd199yp6141</td>
<td></td>
<td>自由党</td>
<td>jpn</td>
<td></td>
</tr>
<tr>
<td>kn444gv1036</td>
<td></td>
<td>社会民主党</td>
<td>jpn</td>
<td></td>
</tr>
</tbody>
</table>
How were they archived? What was the process?

- Identify sites to archive
- Send notification letters & respond to inquiries
- Archive-It crawls & quality assurance
- Six month embargo period
- Create metadata
- Ingest into Stanford Digital Repository
- Make archived sites discoverable in SearchWorks & WorldCat
Typical inquiries

- Confirming the URL of the site to be archived.
- What other sites were chosen?
- Why was our site chosen?
- We plan to renovate our site, will that cause a problem?
- What will the archived site look like? Please provide an example.
- Will our archived site be viewable after the six month embargo?
- Where will it be possible to view the archived site? Is the below site Stanford’s web archive (swap.stanford.edu)?
- We have never received such a notification before. Are you sending it because of some requirement?
- What exactly is “web crawling” (frequency, time, crawler name)?
- You mentioned a six month embargo, but can you tell us the exact date when the archived site will be published? Once it is published, could you send us the address to view it?
Notification letters and project webpage

Snapshot of Japan 2016-2018

An archive of approximately 500 websites touching on central issues of the time, including politics & economics, employment & labor issues, social issues, child care & pregnancy, academic & humanities, environment & nature, science & technology, internet & computers.

The aim of the collection is to give a preserve ephemeral websites that fall between the cracks of other collections and give an important, if complex, perspective on this three year period.

Description

The subcategories are as follows:
Conclusion: challenges

- Time-consuming approach
- Technical issues
- Aiming for accurate picture for posterity, but notification letters may have led to some distortion
- Some site owners perceived selection as endorsement (used Wayback Machine for some more heated topics)
Conclusion: successes

- Provided for posterity a glimpse of issues that were considered central during the two-year period.

- Archived sites that wouldn’t have been preserved.
Reflecting on the project

While the river’s current is ceaseless, the passing water constantly changes. The bubbles that form on its surface, form and burst, never lasting long. People and their dwellings are the same. (2)


(2) My translation of the opening lines of the Hōjōki.
More information

Thank you very much!

REGAN MURPHY KAO
EAST ASIA LIBRARY
STANFORD UNIVERSITY
REGANMK@STANFORD.EDU