Team “Turn All References Blue” (TARB)

Stephen Balbach
Mark Graham
Max Doerr
Jake Orlowitz
James Hare
Dr. Sawood Alam
Turn All **References** Blue

In the future nearly everything ever written or recorded, and all related resources, will be a click, gesture or a thought away.

The goal of the “Turn All References Blue” project is to contribute to making that a reality, **starting with Wikipedia**.
Link Rot: When Good URLs go Bad

Live Web (404)

 Archived Version
Content Drift: Change Happens

April 2, 2020

Strategic National Stockpile is the nation’s largest supply of life-saving pharmaceuticals and medical supplies for use in a public health emergency severe enough to cause local supplies to run out.

When state, local, tribal, and territorial responders request federal assistance to support their response efforts, the stockpile ensures that the right medicines and supplies get to those who need them most during an emergency. Organized for scalable response to a variety of public health threats, this repository contains enough supplies to respond to multiple large-scale emergencies simultaneously.

April 6, 2020

The Strategic National Stockpile's role is to supplement state and local supplies during public health emergencies. Many states have products stockpiled, as well. The supplies, medicines, and devices for life-saving care contained in the stockpile can be used as a short-term stopgap buffer when the immediate supply of adequate amounts of these materials may not be immediately available.
We get URLs cited in Wikipedia via the EventStreamAPI

Event Platform/EventStreams

EventStreams is a web service that exposes continuous streams of structured event data. It does so over HTTP using chunked transfer encoding following the Server-Sent Events (SSE) protocol. EventStreams can be consumed directly via HTTP, but is more commonly used via a client library.

EventStreams provides access to arbitrary streams of data, including MediaWiki RecentChanges, it replaced RCSStream, and may in the future replace irc.wikimedia.org. EventStreams is backed by Apache Kafka.

Note: Often 'SSE’ and EventSource are used interchangeably. This document refers to SSE as the server-side protocol, and EventSource as the client-side interface.

Contents [hide]
1 When not to use EventStreams
2 Examples
   2.1 JavaScript
   2.2 Python
   2.3 Command-line
3 API
   3.1 Stream selection
   3.2 Historical Consumption
   3.3 Response Format
   3.4 Filtering
4 Architecture
   4.1 SSE vs. WebSockets/Socket.IO
   4.2 KafkaSSE
   4.3 Kafka
5 WiMF Administration
We Archive Millions of Wikipedia Sourced URLs/day

### Wikipedia-Eventstream

<table>
<thead>
<tr>
<th>Date</th>
<th>Items</th>
<th>Count</th>
<th>Size</th>
<th>B/uri</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-08-07</td>
<td>7</td>
<td>1,298,950</td>
<td>82,731,858,081</td>
<td>63.7KB</td>
</tr>
<tr>
<td>2023-08-06</td>
<td>8</td>
<td>1,410,651</td>
<td>81,116,676,520</td>
<td>57.5KB</td>
</tr>
<tr>
<td>2023-08-05</td>
<td>6</td>
<td>824,281</td>
<td>60,862,226,293</td>
<td>73.8KB</td>
</tr>
<tr>
<td>2023-08-04</td>
<td>6</td>
<td>960,376</td>
<td>62,092,605,426</td>
<td>64.7KB</td>
</tr>
<tr>
<td>2023-08-03</td>
<td>7</td>
<td>1,076,207</td>
<td>82,631,879,715</td>
<td>76.8KB</td>
</tr>
<tr>
<td>2023-08-02</td>
<td>6</td>
<td>890,475</td>
<td>81,443,293,754</td>
<td>91.5KB</td>
</tr>
<tr>
<td>2023-08-01</td>
<td>8</td>
<td>1,070,500</td>
<td>81,339,625,595</td>
<td>76.0KB</td>
</tr>
<tr>
<td>2023-07-31</td>
<td>8</td>
<td>1,405,654</td>
<td>83,193,245,880</td>
<td>59.2KB</td>
</tr>
<tr>
<td>2023-07-30</td>
<td>12</td>
<td>2,305,861</td>
<td>132,761,583,245</td>
<td>57.6KB</td>
</tr>
<tr>
<td>2023-07-29</td>
<td>9</td>
<td>1,390,846</td>
<td>93,290,117,260</td>
<td>67.1KB</td>
</tr>
</tbody>
</table>

### All Collections

<table>
<thead>
<tr>
<th>Date</th>
<th>Items</th>
<th>Count</th>
<th>Size</th>
<th>B/uri</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-08-07</td>
<td>5,934</td>
<td>8.6B</td>
<td>57.7TB</td>
<td>6.7KB</td>
</tr>
<tr>
<td>2023-08-06</td>
<td>6,337</td>
<td>9.4B</td>
<td>61.6TB</td>
<td>6.5KB</td>
</tr>
<tr>
<td>2023-08-05</td>
<td>6,241</td>
<td>9.6B</td>
<td>61.4TB</td>
<td>6.4KB</td>
</tr>
<tr>
<td>2023-08-04</td>
<td>6,863</td>
<td>7.5B</td>
<td>65.5TB</td>
<td>8.0KB</td>
</tr>
<tr>
<td>2023-08-03</td>
<td>5,741</td>
<td>3.4B</td>
<td>55.0TB</td>
<td>16.2KB</td>
</tr>
<tr>
<td>2023-08-02</td>
<td>5,888</td>
<td>8.2B</td>
<td>56.1TB</td>
<td>6.8KB</td>
</tr>
<tr>
<td>2023-08-01</td>
<td>4,816</td>
<td>6.0B</td>
<td>45.7TB</td>
<td>7.6KB</td>
</tr>
<tr>
<td>2023-07-31</td>
<td>4,317</td>
<td>3.1B</td>
<td>40.4TB</td>
<td>13.0KB</td>
</tr>
<tr>
<td>2023-07-30</td>
<td>4,253</td>
<td>2.8B</td>
<td>37.6TB</td>
<td>13.2KB</td>
</tr>
<tr>
<td>2023-07-29</td>
<td>4,372</td>
<td>2.4B</td>
<td>41.3TB</td>
<td>17.1KB</td>
</tr>
</tbody>
</table>
How 1 seed URL can result in 31,259 Archived URLs

1 seed page

338 embeds

174 outlinks

30,746 outlink embeds

Results from cnn.com on June 17, 2021
We also Archive Millions of Outlinked URLs
IABot has run on 215 out of 334 Wikipedia language editions, and 58 additional wikis.
Progress Fixing Broken Links

- Fixed more than 19 million formerly broken links across 305 Wikipedia languages
- More than 8 million fixed in the last 2 years alone!
Better World Books

The Internet Archive creates a wish list of books based on Wikipedia citations
More than **1.8 million links** to books from [archive.org](http://archive.org) across have been added to ~**60** *Wikipedia language* editions

(1 million added by our software and 800,000 added by [Wikipedia Editors](https://www.wikipedia.org))
Our software has added more than 141,000 links to academic papers and articles across ~60 Wikipedia language editions.
In the Lab: Reference Explorer
Next Up: Finding/Fixing “Soft 404s”

Sorry, this page could not be found.

Sorry! The page you are looking for doesn’t exist or no longer exists. Some options for how to find it:

1. If you typed in a URL... make sure the spelling, capitalization, and punctuation are correct. Then try reloading the page.
2. Try and find it in the sitemap.
3. Start over again at the homepage.
4. Search for it:
   Search
   q
What the Hachette v. Internet Archive Decision Means for Our Library

Posted on August 17, 2003 by Chris Freeland

Our library is still strong, growing, and serving millions of patrons. But the publishers' attack on basic library practices continues.

Upcoming Events

- AUG 2003 10:00 am - 11:00 am
  - Book Talk: Moving Theory into Practice

- OCT 2003 10:00 am - 11:30 am
  - Library Leaders Forum: VIRTUAL

- OCT 2003 05:00 pm - 07:00 pm

Latest Posts

- What the Hachette v. Internet Archive Decision Means for Our Library
- Internet Archive Responds to Recent PR Industry Lawsuit Targeting Online Media
- Our Fight Is Far From Over
- ENH Rollout: How to Improve the Path Toward a Better Web
- Winning Together: The Story of Historic Libraries and the Internet Archive

[Link to relevant page]