Wayback Machine Adds 160 Billion Indexed Pages In A Year, Surpasses 400 Billion Indexed Pages

May 12, 2014 at 9:40am ET by Barry Schwartz

The Internet Archive announced that the Wayback Machine, a huge internet archive of web pages dating back to 1996, has surpassed 400 billion pages indexed.

In January 2013, a little over a year ago, the Wayback Machine said they had 240 billion URLs indexed and since then, they have added another 160 billion URLs! That brings up the indexed page count by the Wayback Machine to over 400 billion URLs.

On Friday, the Internet Archive announced this on their blog and said the indexed pages date back from late 1996 up until a few hours ago. Then they shared some of their history:

- 2006 – Archive-It is launched, allowing libraries that subscribe to the service to create curated collections of valuable web content.
- March 25, 2009 – The Internet Archive and Sun Microsystems launch a new datacenter that stores the whole archive and serves the Wayback Machine. This 3 Petabyte data center handled 500 requests per second from its home in a shipping container.
- June 15th, 2011 – The HTTP Archive becomes part of the Internet Archive, adding data about the performance of websites to our collection of web site content.
- May 28, 2012 – The Wayback Machine is available in China again, after being blocked for a few years without notice.
- October 26, 2012 – Internet Archive makes 80 terabytes of archived web crawl data from 2011 available for researchers, to explore how others might be able to interact with or learn from this content.
- October 2013 – New features for the Wayback Machine are launched, including the ability to see newly crawled content an hour after we get it, a “Save Page” feature so that anyone can archive a page on demand, and an effort to fix broken links on the web starting with WordPress.com and Wikipedia.org.
- Also in October 2013 – The Wayback Machine provides access to important Federal Government sites that go dark during the Federal Government Shutdown.

Related Topics: Channel: Consumer | Internet Archive | Search Engines: Academic | Search Engines

Sponsored

About The Author: Barry Schwartz is Search Engine Land's News Editor and owns RustyBrick, a NY-based web consulting firm. He also runs Search Engine Roundtable, a popular search blog on very advanced SEM topics. Barry's personal blog is named Cartoon Barry and he can be followed on Twitter here. For more background information on Barry, see his full bio over here. See more articles by Barry Schwartz.

Connect with the author via: Email | Twitter | Google+ | LinkedIn

SearchCap: Get all the top search stories emailed daily!

Search Engine Land

Also on Marketing Land: Social · Display · Analytics · Email · Content · Mobile · Video · Industry · Strategy

Get Our News, Everywhere!

Daily Email: EMAIL ADDRESS Subscribe

Got a tip? Tell us here!

Attend Our Conferences

Join us at one of our SMX or MarTech events:

United States
- SMX Advanced - June 11-12, 2014
- MarTech Boston - August 19-20, 2014
- SMX East - Sept. 30 - Oct. 1, 2014
- SMX Social Media Marketing - Nov. 19-20, 2014
- SMX West - March 2015

Europe
- SMX London - May 13-14, 2014
- SMX Paris - June 16-17, 2014
- SMX Milan - Nov. 13-14, 2014
- SMX Israel - January 2015
- SMX Munich - March 2015

Australia & China
- SMX Sydney - May 27-28, 2014
- SMX Beijing - December 10-11, 2014

Learn more about: SMX | MarTech

Free Daily Search News Recap!

SearchCap is a once-per-day newsletter update - sign up below and get the news delivered to you!

Enter Email Address Subscribe Now

Digital Marketing Depot

Learn more about internet and search marketing with our free webinars, whitepapers and research reports at Digital Marketing Depot.

Webcasts:
- Increasing Mobile Performance Through Advanced Call Analytics - Jun 3
substantially add further to a discussion. Comments using foul language, being
disrespectful to others or otherwise violating what we believe are common sense
standards of discussion will be deleted. Comments may also be removed if they are posted
from anonymous accounts. You can read more about our comments policy here.