

Get WIRED for just \$10

SUBSCRIBE NOW

KLINT FINLEY BUSINESS 11.03.2019 07:00 AM

The Internet Archive Is Making Wikipedia More Reliable

The operator of the Wayback Machine allows Wikipedia's users to check citations from books as well as the web.



PHOTOGRAPH: ALEXANDER SPATARI/GETTY IMAGES

WIKIPEDIA is the arbiter of truth on the internet. It's what settles arguments at bars. It supplies answers for the

information snippets you see on your Google or Bing search results. It's the first stop for nearly everyone doing online research.

The reason people rely on Wikipedia, despite its imperfections, is that every claim is supposed to have citations. Any sentence that isn't backed up with a credible source risks being slapped with the dreaded "citation needed" label. Anyone can check out those citations to learn more about a subject, or verify that those sources actually say what a particular Wikipedia entry claims they do—that is, if you can find those sources.

It's easy enough when the sources are online. But many Wikipedia articles rely on good old-fashioned books. The [entry on Martin Luther King Jr.](#), for example, cites 66 different books. Until recently, if you wanted to verify that those books say what the article says they say, or if you just wanted to read the cited material, you'd need to track down a copy of the book.

Now, thanks to a new initiative by the [Internet Archive](#), you can click the name of the book and see a two-page preview of the cited work, so long as the citation specifies a page number. You can also borrow a digital copy of the book, so long as no else has checked it out, for two weeks—much the same way you'd borrow a book from your local library. (Some groups of authors and publishers have [challenged](#) the archive's practice of allowing users to borrow unauthorized scanned books. The [Internet Archive says](#) it seeks to widen access to books in “balanced and respectful ways.”)

So far the Internet Archive has turned 130,000 references in Wikipedia entries in various languages into direct links to 50,000 books that the organization has scanned and made available to the public. The organization eventually hopes to allow users to view and borrow every book cited by Wikipedia, with the ultimate goal being to digitize every book ever published.

“Our goal is to be a library that’s useful and reachable by more people,” says Mark Graham, director of the Internet Archive's Wayback Machine service.

If successful, the Internet Archive's project would be a boon to students, journalists, or anyone who wants to check the references of a Wikipedia entry. Google Books also has a massive collection of digitized print books, but it tends to only show small snippets of a text.

"I've tried to verify Wikipedia pages by searching blurbs in Google Books but it's an unpredictable link, and you often don't have enough surrounding context to evaluate the use," says Mike Caulfield, a digital literacy expert and director of blended and networked learning at Washington State University Vancouver. "The ability to read a page or two of context around a quote is crucial to both editors trying to protect the integrity of articles, and to readers who need to get to that next step of verification."

Sign Up Today



**Want More
Stories From
the Future?**

Sign up for the Daily newsletter and never miss the best of WIRED.

You could, of course, verify the information the traditional way by tracking down a physical copy of a book. But students working late into the night on term papers, or reporters on tight deadlines, might not have time to order a book on Amazon or wait for a library book to become available. In other cases, books might be hard to come by. The Wikipedia entry on the internment of Japanese-Americans during World War II, for example, cites hard-to-find titles, says Internet Archive director of partnerships Wendy Hanamura. But thanks to the Internet Archive's Digital Library of Japanese-American Incarceration, created with the Seattle-based organization Densho, many of those rare books are now available online.

The Internet Archive embarked on its effort to weave digital books into Wikipedia after the 2016 election. "No matter who you wanted to be president, I would say almost everyone would agree the whole process was a train wreck," Internet Archive founder Brewster Kahle

said in a [speech in San Francisco last week](#). From fake news and inauthentic social media campaigns waged by foreign nations to concerns about voting systems themselves being rigged, there were plenty of ways that technology and information systems failed the public. So Kahle convened a group of people to discuss how to improve the information ecosystem. One issue that came up was the fragility of Wikipedia citations. Books and academic journals supply some of the best, most reliable information for Wikipedia editors, but those sources frequently are either unavailable online or are behind paywalls. And even freely available internet content often disappears.

The Internet Archive was in a unique position to help solve this problem. The organization's Wayback Machine service has archived 387 billion webpages since 2001. It's also been digitizing physical books and other analog media, and has now scanned 3.8 million books. It has millions more books warehoused.

Graham and company created the InternetArchiveBot, a tool that scans Wikipedia for broken links and automatically adds links to versions archived in the Wayback Machine. Because automatic editing tools require special permission to use, Graham has to work with the Wikipedia communities that manage versions of the encyclopedia in different languages. "All told, we've edited 14 million links; more than 11 million point to Internet Archive," he says.

Adding links to books is similar but more challenging. "If a book has an ISBN number and an entry has a traditional citation format, it's pretty easy," Graham explains. But not all books have ISBN numbers, and many Wikipedia citations aren't properly formatted. For instance, some only cite the book and not a specific page number. There can also be differences between different editions of a book.

Of course, the Internet Archive hasn't scanned all the books cited by Wikipedia yet. It's working hard to digitize collections from libraries around the world, along with donations from companies like Better World Books. Graham says the organization scans more than 1,000 books per day. But it has plenty more work to do.

More Great WIRED Stories

- YouTubers must unionize, no matter what Google says
- Could an astronaut lost in space use gravity to get around?
- The best jobs are in government. No, really
- WIRED25: Stories of people who are racing to save us
- The plan to boost drone batteries with a teensy jet engine
- 👁 Prepare for the deepfake era of video; plus, check out the latest news on AI
- 📱 Torn between the latest phones? Never fear—check out our iPhone buying guide and favorite Android phones

1 Year of WIRED for \$10.
Plus a free tote!



SUBSCRIBE

TOPICS

WIREDCAR

INTERNET

BOOKS