The Internet is not forever after all: CNET deletes old articles to game Google

Aurich Lawson

reader comments

63 with

CNET, one of the great-granddaddies of tech news on the web, has been having a rough year. First, its AI-written articles sparked drama, then layoffs rocked the publication. And now, Gizmodo reports that the 28-year-old site has been deleting thousands of its old articles in a quest to achieve better rankings in Google searches.

The deletion process began with small batches of articles and dramatically increased in the second half of July, leading to the removal of thousands of articles in recent weeks. Although CNET confirmed the culling of stories to Gizmodo, the exact number of deleted articles has not been disclosed.

“Removing content from our site is not a decision we take lightly. Our teams analyze many data points to determine whether there are pages on CNET that are not currently serving a meaningful audience. This is an industry-wide best practice for large sites like ours that are primarily driven by SEO traffic. In an ideal world, we would leave all of our content on our site in perpetuity. Unfortunately, we are penalized by the modern Internet for leaving all previously published content live on our site,” Taylor Canada, CNET’s senior director of marketing and communications, told Gizmodo.

SEO (search engine optimization) is the practice of attempting to purposely achieve higher rankings in search engine results by changing a website's content. Proponents of SEO techniques believe that a higher rank in Google search results can significantly affect visitor count, product sales, or ad revenue. Many companies go to extremes trying to please Google's ranking algorithm.

Advertisement

One theory of improving page rank involves a practice called "content pruning." Gizmodo obtained an internal memo from CNET which states that removing old URLs "sends a signal to Google that says CNET is fresh, relevant and worthy of being placed higher than our competitors in search results." However, before deleting an article, CNET reportedly maintains a local copy, sends the story to The Internet Archive's Wayback Machine, and notifies any currently employed authors that might be affected at least 10 days in advance.

However, some experts say that CNET's extreme example of content pruning is misguided. The website Search Engine Land notes that while Google once advised (in 2011) that removing "low-quality pages" could potentially increase rank, it also says that Google has never advised people to delete content simply because it is old. In fact, on Tuesday, Google's SearchLiaison X account tweeted, "Are you deleting content from your site because you somehow believe Google doesn't like "old" content? That's not a thing! Our guidance doesn't encourage this. Older content can still be helpful, too. Learn more about creating helpful content."

The long, ongoing decay of the web

At one point, it was perceived as common knowledge that "the Internet is forever," meaning that whatever you put online will always stay there. Our informal searches through Google Books and Google suggest that the phrase originated around

2005 but became very popular in the 2008-2009 social networking boom era.

As time has passed, however, it’s become increasingly clear that the Internet is transitory. Link rot threatens content on the web every day, and content found online is far from permanent. A 2021 Harvard University study examined hyperlinks in over 550,000 New York Times articles from 1996 to 2019 and discovered that 25 percent of links to specific pages were inaccessible. If it weren’t for The Internet Archive, for instance, many early websites would be completely lost.

Advertisement

Enlarge / A screenshot of a PCWorld article that is missing an image.
Ars Technica

Causes of link rot include website shutdowns, server migrations, shifts to new content management systems, and more. Now we can add another culprit to the list: content pruning for SEO. It is perhaps another sign of how bad things have become with Google’s search results—full of algorithmically generated junk sites—that publications like CNET are driven to such extremes to stay above the sea of noise.

Even if websites don’t pull down content completely, certain archives can be compromised over time in other ways. Over the past decade, a plague of copyright trolls threatened many publications with lawsuits for using images in a manner that would likely constitute fair use if tested in court. But trials are expensive, so the trolls won by receiving cash settlements. In response, many websites removed old images from articles instead of sorting millions of them individually. Archives on some of IDG’s websites, such as PCWorld and Macworld, have been affected by this image culling.

Not all websites disregard their archives or fall into the SEO trap, thankfully. For example, on Ars Technica, you can still find articles written 25 years ago (and many in their original format), and Ars’ search function still works remarkably well.

From time immemorial, the protection of historical content has required making many copies without authorization, regardless of the cultural or business forces at play, and that has not changed with the Internet. Archivists operate in a parallel IP universe, borrowing scraps of reality and keeping them safe until shortsighted business decisions and copyright protectionism die down. Hopefully, despite the link rot, future historians can piece together an accurate history of our fragile digital era.

https://arstechnica.com/?p=1960032