

We're now on LinkedIn: For news and resources from Google Search on making your site discoverable, [follow us on LinkedIn](https://www.linkedin.com/showcase/googlesearchcentral/) (<https://www.linkedin.com/showcase/googlesearchcentral/>).

Our new search index: Caffeine

It's been a while since we published this blog post. Some of the information may be outdated (for example, some images may be missing, and some links may not work anymore).

Wednesday, June 09, 2010

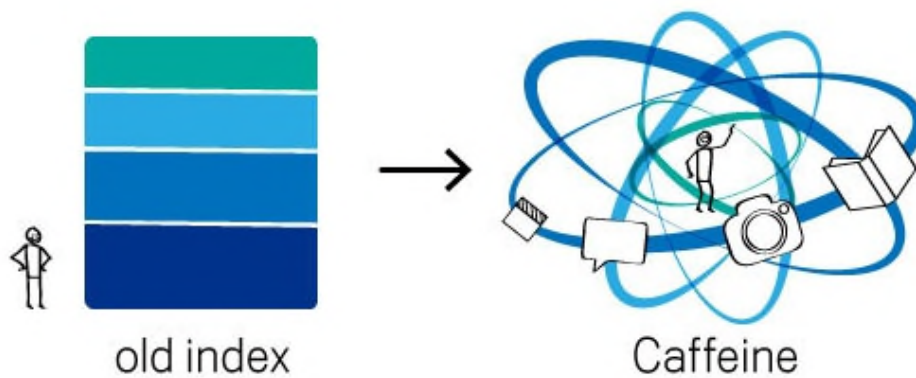
(Cross-posted on the [Official Google Blog](https://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html) (<https://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>))

Today, we're announcing the completion of a new web indexing system called Caffeine. Caffeine provides 50 percent fresher results for web searches than our last index, and it's the largest collection of web content we've offered. Whether it's a news story, a blog or a forum post, you can now find links to relevant content much sooner after it is published than was possible ever before.

Some background for those of you who don't build search engines for a living like us: when you search Google, you're not searching the live web. Instead you're searching Google's index of the web which, like the list in the back of a book, helps you pinpoint exactly the information you need. (Here's a [good explanation](https://www.google.com/search/howsearchworks/) (<https://www.google.com/search/howsearchworks/>) of how it all works.)

So why did we build a new search indexing system? Content on the web is blossoming. It's growing not just in size and numbers but with the advent of video, images, news and real-time updates, the average webpage is richer and more complex. In addition, people's expectations for search are higher than they used to be. Searchers want to find the latest relevant content and publishers expect to be found the instant they publish.

To keep up with the evolution of the web and to meet rising user expectations, we've built Caffeine. The image below illustrates how our old indexing system worked compared to Caffeine:



Our old index had several layers, some of which were refreshed at a faster rate than others; the main layer would update every couple of weeks. To refresh a layer of the old index, we would analyze the entire web, which meant there was a significant delay between when we found a page and made it available to you.

With Caffeine, we analyze the web in small portions and update our search index on a continuous basis, globally. As we find new pages, or new information on existing pages, we can add these straight to the index. That means you can find fresher information than ever before—no matter when or where it was published.

Caffeine lets us index web pages on an enormous scale. In fact, every second Caffeine processes hundreds of thousands of pages in parallel. If this were a pile of paper it would grow three miles taller every second. Caffeine takes up nearly 100 million gigabytes of storage in one database and adds new information at a rate of hundreds of thousands of gigabytes per day. You would need 625,000 of the largest iPods to store that much information; if these were stacked end-to-end they would go for more than 40 miles.

We've built Caffeine with the future in mind. Not only is it fresher, it's a robust foundation that makes it possible for us to build an even faster and comprehensive search engine that scales with the growth of information online, and delivers even more relevant search results to you. So stay tuned, and look for more improvements in the months to come.

Posted by Carrie Grimes, Software Engineer

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](https://developers.google.com/site-policies) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.