THE DEEP WEBB AND HOW TO USE IT

There are over 130 trillion individual Web pages. But that's only the very tip of the iceberg. Beyond what popular search engines offer up, there is a universe of information that's online and discoverable -- if you have the right skills and tools. Here's how to gain access to the rest of the Web.

What's Out There on the Deep Web?

One hundred and thirty trillion is a big number. Back in 2016, Google's "How Search Works" page said the number of web pages had doubled since 2013, and in 2008 there were "only" one trillion Web pages. Clearly, none of us will ever lack for reading material. And that's only the beginning of what's available online. (Side note: the first webpage was published on August 6, 1991, by British physicist Tim Berners-Lee, who invented the World Wide Web.)

General search engines like Google, Bing, and Yahoo! index only the "surface Web," pages that have unique URLs. They don't even index (catalog) all of the surface Web. Some website owners prefer not to have their sites (or portions of them) indexed, so they use a file called robots.txt that tells search engines, "Don't include this page in your database."

Search engines choose to exclude many other surface Web pages from their indexes for a variety of reasons including relevance, legality, and violations of search optimization policies. Other pages are locked behind passwords, intended only for those who are granted access.

Beneath the surface Web lies the "Deep Web," a mass of information hundreds of times greater than the trillions of surface pages discovered by Google. By their very nature, Deep Web resources cannot be found

by the web-crawling software that search engines use to find and index pages.

Note that the "Deep Web" is not the same as the "Dark Web" where criminals lurk. On the Dark Web, everyone tries to hide their identities as well as what they're doing. Unless you're the type that wanders down dark alleys at 2 AM, you'll want to steer clear of the Dark Web. The Deep Web consists of perfectly legitimate information and its users.

Some of these Deep Web pages can be accessed only by a user clicking or manually typing a link that's not been indexed by a search engine. Other Deep Web pages can be accessed only by a user who directly enters a query in a search form. The desired data exist in a database, not on a Web page that a crawler can find by following links from other pages. The data retrieved in response to a user query is displayed as a "dynamic" Web page that lasts only until the user moves on. When you search for products on an ecommerce site like Amazon, the results are dynamic pages.

What Tools will Search the Deep Web?

To find Deep Web material via Google or Edge, try adding the term, "database" to your search query. "Plane crash database," "drug interaction database," "government grants database," and so on, will often lead to the home page of a database where you can enter search terms specific to that resource.

There are also paid tools such as LexisNexis and Factiva which professional researchers use to find information about legal and business topics. Genealogy researchers can find a wealth of free information online, but often the best sources require payment. Ancestry.com is one such example. It's also becoming more common for online newspapers and magazines to limit free content, and erect paywalls that require a subscription to view more than current headlines.

General search engines suffice for most needs. Scholars, journalists, and other serious researchers often must resort to the Deep Web.