

I'm not a bot



into two major components: the scheme and the authority. A lot of people think of a URL as just a web address, but it's not quite that simple. A web address is a URL, but all URLs are not web addresses. Other services you can access on the internet--like FTP--or even locally--like MAILTO--are also URLs. The scheme portion of a URL (those letters followed by a colon) denote the protocol with which an app (like your web browser) and the server should communicate. Web addresses are the most common URL, but there are others. So, you might see schemes like: HyperText Transfer Protocol (HTTP): This is the underlying protocol of the web and determines what actions web servers and browsers should take in response to certain commands. HTTP Secure (HTTPS): This is a form of HTTP that works over a secure, encrypted layer for safer transport of information. File Transfer Protocol (FTP): This protocol is often still used for transferring files over the internet. In modern browsers, the scheme isn't technically required as part of the URL. If you enter a website like "www.howtogeek.com", your browser will automatically determine the right protocol to use. Still, some other apps (and protocols) require the use of a scheme. The authority portion of a URL (which is preceded by two slashes) is itself broken down into a bunch of parts. Let's start off with a very simple URL--the kind that would take you to the home page of a website. In this simple example, the whole "www.example.com" part is called a hostname, and it resolves to an IP address. You can also type an IP address into your browser's address bar instead of the hostname if you happen to know it. But, when parsing the hostname it helps to read it backward to understand what's going on, so here are those components: Top-Level Domain: In the example here, "com" is the top-level domain. These are the highest level in the Domain Name System (DNS) hierarchy used to translate IP addresses into simple language addresses that are easier for we humans to remember. These top-level domains are created and managed by the Internet Corporation for Assigned Names and Numbers (ICANN). The three most common top-level domains are .com, .net, and .gov. Most countries also have their own two-letter top-level domain, so you'll see domains like .us (United States), .uk (United Kingdom), .ca (Canada), and many others. There are also some additional top-level domains (like .museum) that are sponsored and managed by private organizations. In addition to these, there are also some generic top-level domains (like .club, .life, and .news). Subdomain: Since DNS is a hierarchical system, both the "www" and "example" parts of our example URL are considered subdomains. The "www" portion is a subdomain of the "com" top-level domain, and the "www" portion is a subdomain of the "example" domain. That's why you'll often see a company with a registered name like "google.com" broken out into separate subdomains like "www.google.com," "news.google.com," "mail.google.com," and so on. That's the most basic example of the authority section of a URL, but things can get more complicated. There are two other components that the authority section can contain: User Information: The authority section can also contain a username and password for the site you're accessing. It's uncommon to see this structure in URLs today, but it can happen. If present, the user info portion comes before the hostname and is followed by an @ sign. So, you might see something like "/username:password@www.example.com" if it includes the user information. Port Number: Network devices use IP addresses to get information to the right computer on a network. When that traffic arrives, a port number tells the computer the application for which that traffic is intended. The port number is another element you won't often see when browsing the web, but you might see it in network apps (like games) that require you to enter a URL. If the URL includes a port number, it comes after the hostname and is preceded by a colon. It would look something like this: "://www.example.com:8080." So, that's the scheme and authority portions of a URL, but as you might have guessed after looking at a lot of URLs while browsing the web, they can include even more stuff. There are three additional parts of a URL that you might see after the authority portion: paths, queries, and fragments. Here's how those work. Path The authority section of a URL gets your browser (or whatever app) to the right server on a network. The path that follows--which works just like a path in Windows, macOS, or Linux--gets you to the right folder or file on that server. The path is preceded by a slash, and there's a slash between every directory and subdirectory, like this: www.example.com/folder/subfolder/filename.html The last piece is the name of the file that is opened when you access the website. Although you may not see it in the address bar, that doesn't mean it isn't there. Some languages used to create web pages hide the file name and extension you're viewing. This makes the URL easier to remember and type, and gives it a cleaner look. Query The query portion of a URL is used to identify things that aren't part of a strict path structure. Most often, you'll see them used when you perform a search or when a web page delivers data through a form. The query portion is preceded by a question mark and comes after the path (or after the hostname if a path is not included). As an example, take this URL presented when we searched Amazon for the keywords "wi-fi extender": nb_sb_noss_2?url=search-alias%3Daps&field-keywords=wi-fi-extender The search form passed information to Amazon's search engine. Following the question mark, you can see there are two parts to the query: a URL for the search (that's the "url=search-alias%3Daps&field" part) and the keywords we typed (that's the "keywords=wi-fi+extender" part). That's a fairly simple example, and you'll often see URLs with additional (and more complicated) variables. For example, here's the URL when we searched Google for the keyword "howtogeek": As you can see, there's some different information there. In this case, you can see that there's additional information indicating the search language, the browser we used (Chrome), and even the version number of the browser. Fragment The final component of a URL that you might see is called a fragment. The fragment is preceded by a hash mark (#) and is used to indicate a specific location on a web page. When coding a web page, designers can create anchors for specific text like headings. When the proper fragment is used at the end of a URL, your browser will load the page and then jump to that anchor. Anchors and URLs with fragments are often used to create tables of content on web pages to make navigation easier. Here's an example. Wikipedia's page on the Renaissance is quite a long document, and it's broken up into about 11 sections, each of which has multiple subsections. But each heading on the page has an anchor included, and a table of contents at the top of the article includes links that let you jump to the different sections. Those links work by including fragments. You can also use these fragments directly in your address bar or as shareable links. Say, for example, you wanted to show somebody the section of that page that covers Russia. You could just send them this link: That "#Russia" part at the end of the URL jumps them straight to that section after loading the page. So there you have it--more than you likely ever wanted to know about how URLs work. Image Credit: Pawel Horazy/Shutterstock