**What is a URL?**

Imagine all files on the Internet are located in a megalopolis, each of them lucky to be living in their own home. Given the scale of the Internet, in the resulting settlement, there will be an unimaginable number of blocks and streets. For example, you have an interesting article that you want to share with your friend who lives in one of the houses in the megalopolis. How do you explain where exactly to find the article? That's right, you need to come up with a single standard by which you could name all the addresses in the city, and then give your friend the street name and house number, just like in a real city!

All documents on the Internet have a personal address. For example, the URL of the JetBrains website looks like this:

https://www.jetbrains.com/

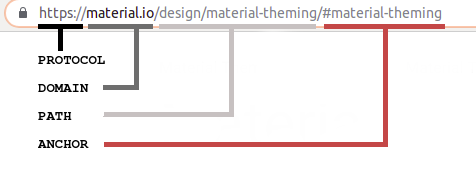
Web pages, images, videos, and other documents that can be stored on your computer also have addresses. To make them look the same on the Internet, in 1990 the creators of the World Wide Web developed a special standard that defines what addresses should look like. That standard is called a **URL,** which stands for **Uniform Resource Locator**. It represents the standardized way of recording file addresses on the Internet.

The standard has one specific feature: not all characters can be used in URLs. The list of allowed characters includes:

* Latin alphabet (or English alphabet symbols);
* An Internationalized Resource Identifier (IRI), a form of URL that includes Unicode characters;
* Numbers;
* Reserved characters with special meaning !#$&'()\*+,/:;=?@[];
* Unreserved characters: -\_.~.

**Basic URL Structure**

Here is an example of a URL address:



URL address has a certain structure based on the following template:

<scheme>://<login>:<password>@<host>:<port>/<path>?<request parameters>#<anchor>

Now let's look at this template in more detail:

* <scheme> is a way of exchanging data with a resource. You are probably most familiar with http and https URL schemes, but there are others;
* <login>and<password> are prefixes that transmit authentication data for some protocols, if necessary;
* <host> is the domain name or IP address where the site is located. **Domain** is the name of the site, IP is its address in a network;
* <port> is required for connection within the specified host. The official default port for HTTP connections is 80, and the alternative is 8080, but it is possible to use any other ports too. The default setting for HTTPS is 443;
* <path> indicates the exact address of a particular file or page within a domain;
* <request parameters> are parameters transmitted to the server. Depending on request parameters, the site may slightly change its display. For example, it is possible to sort the items of a list in a different order;
* <anchor>allows you to connect to a specific part of a web page or document.

This is the general structure of most URLs. Most often, when accessing web pages and documents located on a web server, most of the parameters are not mandatory and are set automatically.

When you just want to see a particular page on the Internet with your browser, the URL template looks a lot easier:

<scheme>://<host>

For example, it can be recorded in a form:

https://www.google.com

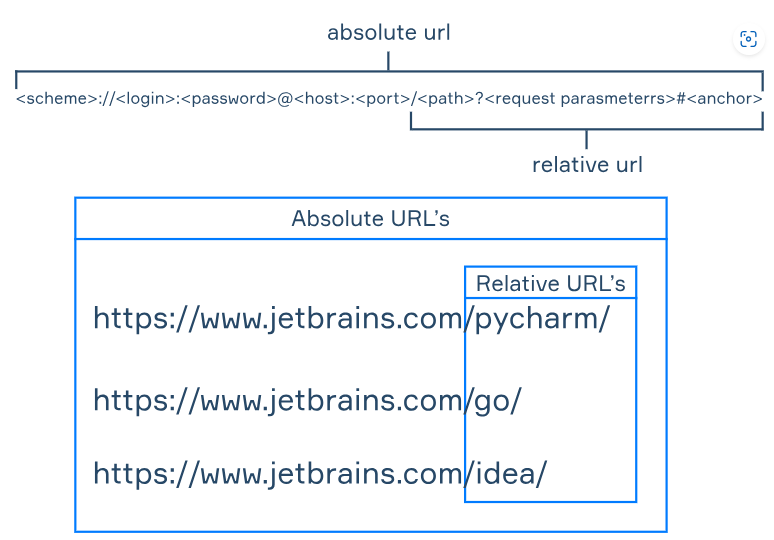
This simplification was created to make life easier for ordinary Internet users, but most programmers need to know the complete template, and now you do.

**Absolute and relative URLs**

As we know, a URL consists of several parts, and when you're browsing through the same site, some elements of it stay the same. Whichever IDE you want to read about on JetBrains, the scheme and host parts of a URL always match [https://www.jetbrains.com](https://www.jetbrains.com/pycharm/). For example, let's look at these links:

* <https://www.jetbrains.com/pycharm/> about PyCharm
* <https://www.jetbrains.com/go/> about GoLand
* <https://www.jetbrains.com/idea/> about IntelliJ IDEA

The new information in each link is its <path>. There exists another way to locate resources on the same site by only <path>?<request parameters>#<anchor>. The complete web address is called an **absolute** URL, while its shorter version is called a **relative** URL. These terms are shown in the picture below.



You should remember that it would work only on the same site, while you cannot refer to another site by a relative path.

We know that by absolute URLs we can easily find the resource through the Internet, but why do we need relative paths? Here are the main reasons for that:

* They are short, and coding is more accessible with them.
* We can easily move the site to another host because relative paths do not depend on a particular domain.
* They are a little bit faster to retrieve by a browser.

**Conclusion**

Let's sum up what you have learned about URLs in this topic:

* We can locate any resources on the Internet through a URL.
* Each URL consists of several parts, but some of them are optional.
* We can retrieve resources by an absolute URL and then browse them through relative paths.

Given the following URL:

https://freenews.org/world

What port is used here by default? Enter a port number in the answer field.

The default port used for HTTPS (which is indicated by https:// in the URL) is:

443

What is a domain?

The name of a site, its address in the global network

Select the URLs that may exist.

Let's evaluate each URL for validity:

1. ✅ [**https://www.iwant-tobeabird.com**](https://www.iwant-tobeabird.com)
   * Valid: Uses only allowed characters (letters, numbers, hyphens).
2. ✅ [**https://www.itgets9better.org**](https://www.itgets9better.org)
   * Valid: Uses letters and numbers, which are allowed.
3. ❌ [**https://www.ho(pe.com**](https://www.ho(pe.com)
   * Invalid: Parentheses ( and ) are **not allowed** in domain names.
4. ❌ **https://www.my\_mom-said!.com**
   * Invalid:
     + Underscores \_ are **not allowed** in domain names.
     + Exclamation marks ! are also **not allowed**.

### ✅ Correct selections:

* https://www.iwant-tobeabird.com
* <https://www.itgets9better.org>

What does URL stand for?

Uniform Resource Locator

Given the following URL:

https://freenews.org/world

As you know, the default port for an https connection is 443; change the URL to use port 9999 and enter the full result.

https://freenews.org:9999/world

Specify which URL matches the pattern:

<scheme>://<host>/<path>/#<anchor>

<https://www.test.com/test123/#test>

Match parts of the URL structure to their values.

| **URL Part** | **Description** |
| --- | --- |

|  |  |
| --- | --- |
| <scheme> | a way of exchanging data with a resource |

|  |  |
| --- | --- |
| <login> and <password> | these prefixes transmit authentication data for some protocols, if necessary |

|  |  |
| --- | --- |
| <host> | the domain name or IP address on which the server or site is located |

|  |  |
| --- | --- |
| <port> | required for connection within the specified host |

|  |  |
| --- | --- |
| <path> | indicates the exact address of a particular file or page |

|  |  |
| --- | --- |
| <request parameters> | parameters that are transmitted to the server |

|  |  |
| --- | --- |
| <anchor> | allows you to link to a specific part of a web page or document |

Select the URL addresses that cannot exist.

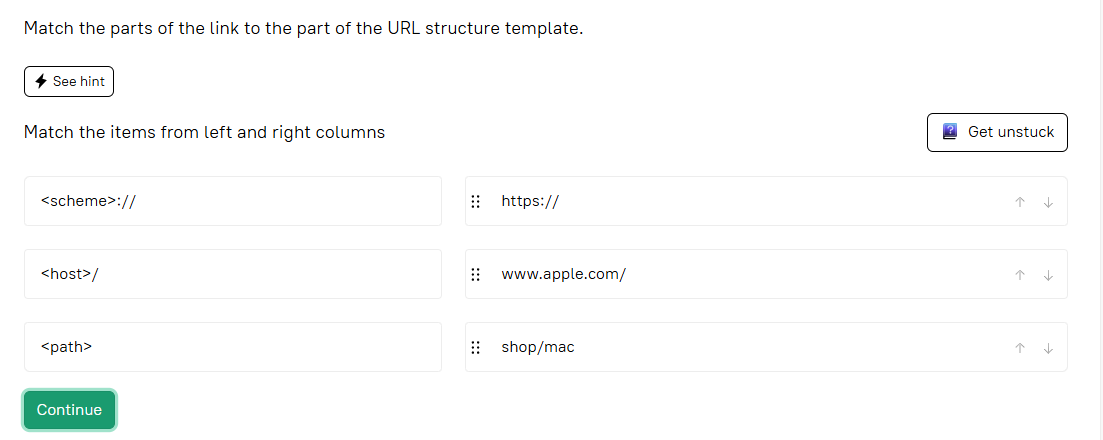
The characters **allowed** in a URL are:

✅ **the Latin alphabet**  
✅ **numbers**  
✅ **/** (forward slash)  
✅ **-** (hyphen)

❌ **<** and **>** are **not allowed** in URLs — they are reserved characters in HTML and can cause errors.  
✅ **\_** (underscore) is technically allowed in the **path** and **query** parts of a URL, though **not allowed in domain names**.

### Final correct options:

* ✅ the Latin alphabet
* ✅ numbers
* ✅ /
* ✅ -
* ✅ \_ (with note: allowed in path, not domain)



What is a URL?

A standardized way of recording the resource address on the Internet