Programming Assignment #1 - URL parser

What’s a URL:

According to the developers at Mozilla Firefox, a URL is an Uniform Resource Locator that helps the user navigate through the whole wide web, known as the Internet. Each URL will point to a unique resource. An URL contains several basic parts. An Protocol, which defines the method for the browser to request the resource, and the only protocol we use today are http, https, ftp, and ftps. The Domain, which will indicate what Web server to request. Domains or known as authority, contain 3 parts, all separated by a period. The last part can only contain either com, net, edu, biz, gov, or edu. A :// has to be between the Domain and Protocol or else the URL is not a valid URL. Next is the port, which is separated by the domain with a colon and is a number between 1 and 65535, a port indicate the gate used to access the website. This is optional on a URL. The next part is the path, which is the path to a spefeic part of a website. This has to end with a .html or .htm. And the last part is the parameter, this provides the web server extra variables to know and is separate by the path by a question mark.

Assignment’s Purpose:

We have developed a parser to check if a URL is a valid URL. This checks if all of the components of URL is there and neatly display each part of the URL. This parser will also check for various errors, like if the protocol is not either https, ftp, or ftps or checks if the domain as the correct amount of parts. If the parser runs into an error in the URL, the parser will neatly display an error message and what part of the URL is invalided.

How to use parser:

The follow instructions are for in terms of running this parser through shell commands.

1. Open your shell terminal and navigate to the parser folder, in our case we would type cd /home/user/PA01\_Jesse\_Truong. The parser requires the code to be compile, in the event that the parser is not compiled, enter the following command **exactly**
   1. **g++ URLparser.cpp -o URLCompile**
2. In the folder there is an executable file with permissions to execute for all users. To run the parser, enter the following command **exactly**
   1. **bash parserExecutable**
3. The script should prompt the following words, “Please enter a valid url: “, and just enter any URL you want, need it be correct or incorrect.
4. If the URL is correct, the script should print the Protocol, Domain, Port, File Path, and Parameters in the URL. The script would stop after.
5. If the URL is incorrect, the script should print “Invalid URL with following erroneous components” and it would print what parts that failed the parser’s algorithm. The script would stop after.

URL Errors:

The parser can detect the following errors

* Missing :// that is require in an URL
* Protocol is not a valid Protocol
* Too many or too little domain arguments
* The last Domain has an invalid domain name
* The port number is not between 1 and 65535
* Path is missing either .html or htm at the end of the path

**Useful knows:**

**The main code is called URLparser.cpp**

**The Executable only works if the compiled code is called URLCompile**

**The Executable is called parserExecutable**