

getlinks.pl

Challenge

Description: Create a perl application called “getlinks.pl” using perl 5 and regular expressions that gets all the urls from tags in an html document that is loaded from a file whose path is provided as a parameter to the program.

Purpose: This challenge provides experience creating regular expressions.

Requirements:

Name of application: getlinks.pl

Programming language: perl 5 using regular expressions

The goal is to write a program that displays a list of urls from tags in an html document. The html document is loaded from a file whose path is provided as a parameter to the program.

The program getlinks.pl is to take one argument:

- Path to an html file.

The program is to display the list of urls followed by the number of urls found. One url per line is to be displayed. The number of urls is to be the last line displayed.

In a web page the following is the format for an <a> tag:

This is what the user sees for the link

The <a> tag can have other properties in it beyond href. For example:

This is what the user sees for the link

The program is to locate all of the <a> tags and get the url this is provided between the quotes after the href= in the tag. In need to accommodate that <a> tags may have other properties in them beside href.

If the user does not supply sufficient/correct command-line parameters, the program is to display correct usage of the program. DO NOT hard code the name of the program for displaying the correct usage. Use the perl variable that holds the name of the program.

The challenge is very, very, very similar to the example covered in class where we got the src url from an tag.

You should use the macrumors.html file to test your program.