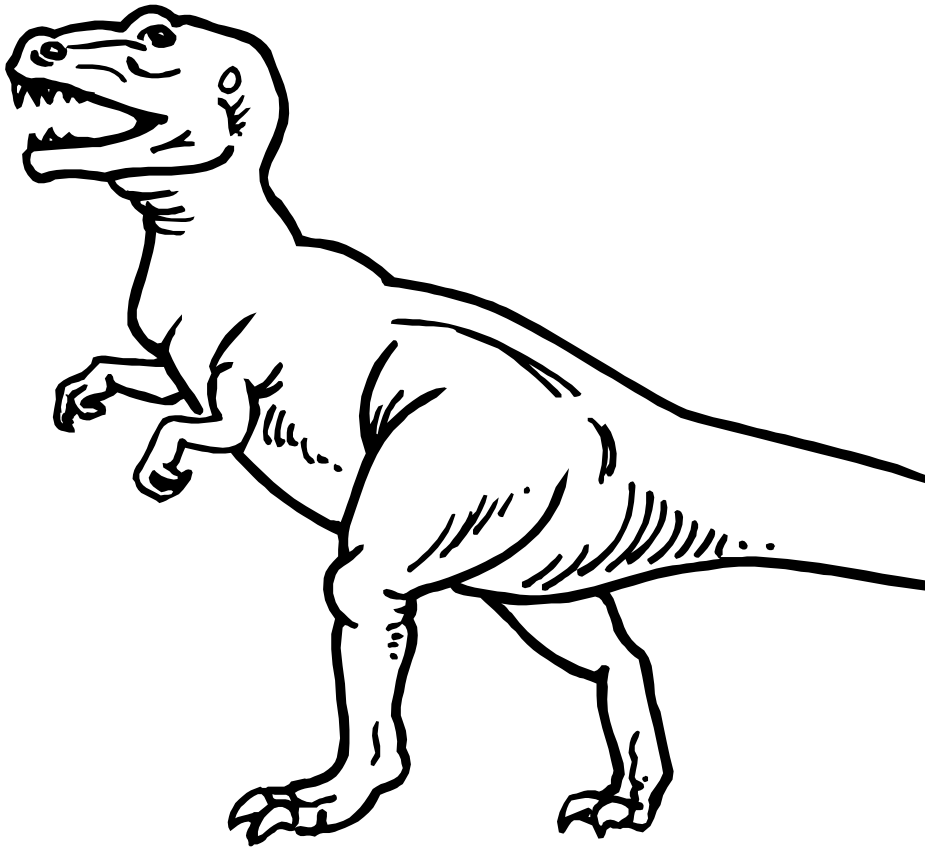
EMBEDDED COMPONENTS, INC.

Software Tools for Intelligent Marketing

ECI's WebScan

SOFTWARE TOOLS FOR INTELLIGENT MARKETING

ECI's WebScan



Embedded Components, Inc.
233 East Red Oak Drive • Suite i
Phone 408.390.1895 • info@EmbeddedComponents.com

Table of Content

Revision Details

ECI's WebScan User Guide, Version 1.0

Last updated: October 2, 2006

Landing page for ECI's WebScan product:
http://www.embeddedcomponents.com/test/webscan.php

Chapter

1

Use ECI's WebScan to Speed Read Web Sites or to Find Bad Links

ECI's WebScan is a new cognitive marketing tool. It helps you take control of large volumes of Internet content using a real-time scan and display web browser tool.

ECI's WebScan is a service that runs on our web server that is accessible from most any type browser on your computer. ECI has a landing page reserved for this tool that describes the latest version and features. You can launch ECI's WebScan from this landing page as well.

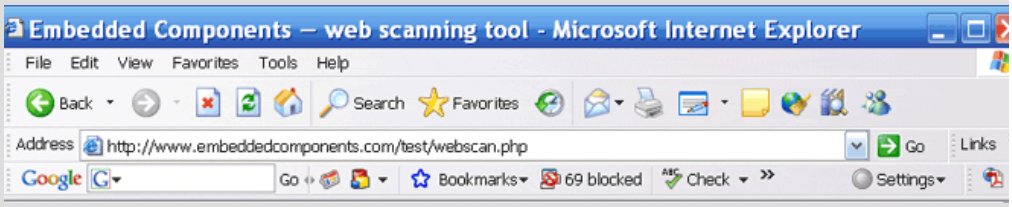


Figure 1 - www.embeddedcomponents.com/test/webscan.php

Once ECI's WebScan is launched, you define a starting web page and select focus controls to define your desired scan using the form presented. Submit your request and watch the progress of the scanning session in real-time. A typical scan will display the base address, the page address of the content being scanned, the page title, and other pertinent details such as page faults, number of redirects, and any reasons for skipping the scan of this page.

What others are saying

> "I use Webscan to speed read other web sites. A fast read is possible because the web page titles are shown along with page links during the scanning process. I use it to summarize our own web site's bad links as they appear over time too." - a quote from a delighted marketer.

When ECI's WebScan is complete, a summary shows the number of web pages scanned, how many levels were traversed, elapsed time, total size of files read, any error messages indicating that the scan did not complete for some reason, and a table listing any bad, suspicious, or skipped pages found during the scan.

Figure 2 - Sample ECI WebScan Result

Chapter

2

Enter the form data needed to start and focus a new web scan session

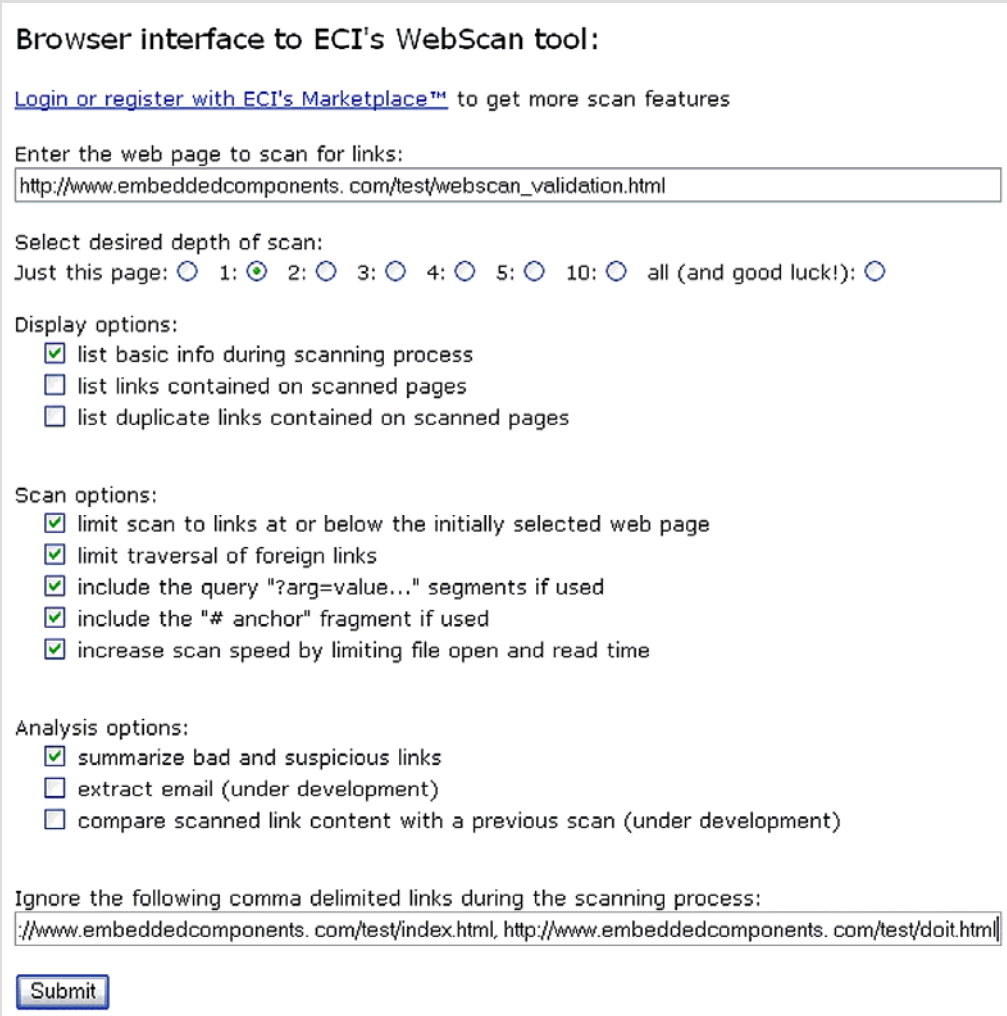


Figure 3 – ECI's WebScan sample data entry form used to focus your scan

Login or register with ECI's Marketplace

Refer to "Figure 3 – ECI's WebScan sample data entry form used to focus your scan" for this and subsequent sections of this chapter.

ECI's WebScan tool is a service offered to members of ECI's Embedded Components and Tools Marketplace. If you are not already logged into ECI's Marketplace, then a message and link suggests that you do so, before attempting a web scan. Click on the "Login or register" link if it is present. If you are already logged in, then this message will not be present. The login process will send you back to this page when finished. You may skip the login process but ECI's WebScan will only operate in a reduced memory model. All features still work but you will only be able to scan a limited number of web links.

Enter the top level, or base, web page to initiate your scan

The entry should be in the form of a Uniform Resource Locator (URL). Valid protocols are: http, https, or ftp. Valid web pages should be either text based or pdf files. Text pages might typically use an html, htm, asp, or php file extension. Sometimes a set of page qualifies are included or an anchor fragment is used. The complete URL format accepted by ECI's WebScan is shown here:

```
http://username:password@www.mypage.com/anypath/newlink.html?arg=qualifier&arg=qualifier#anchor
```


Enter desired depth to scan

ECI's WebScan will begin scanning from the initial page. New links are harvested during the scanning process. The newly harvested links from the initial page will then be scanned. Each of these pages is in turn scanned with new, non-duplicate, links harvested. A depth of zero is defined as just the initial page, a depth of one is defined as all links found on the first page. While a depth of two would be all the links harvested during the depth one scan, and so on. A practical consideration during depth selection is how much time and memory you want to consume.

Each new depth scanned can greatly increase the number of links found. The depth count can also be looked at as the minimum number of "clicks" needed to get to a particular web page. Thus, if a link is first found at depth two, then a user would have to click on two pages before landing on this page.

You can request all links be scanned. In this case, scanning will continue until no new links have been found. Typically a scan of all links is only practical if you also use some of the focusing options discussed later, such as "limit scanning of foreign links", and "only scan links below the initial web page".

Enter display options during scan

There are several display options to consider. The default option lists basic information during the scanning process. Basic information includes base page, link being scanned, title of the page, and various decisions ECI's WebScan makes during the scanning process. Decisions include: skipping the page as a result of unsupported file type, reporting on number of page redirections during the scan, breaking really large files into subfiles during the search for new links, or aborting the page for one of several reasons. The scanning of a page may be aborted because of time constraints, bad link, wrong file type, or your memory limit has been exceeded.

The basic info display option is useful as a speed reading tool. Each page is presented in your web browser in real-time during the scanning process. This is not the normal behavior for a browser. Normally a web browser only displays the results when the complete page is ready for display. ECI discovered how to configure page displays so information can be shown as it is discovered by our server. As you watch the scanning progress you can learn very quickly the content and quality of the web site. See "Figure 2 - Sample ECI WebScan Result" for a sample of what the display looks like.

Other options may be selected to increase the amount of information displayed to include duplicate links. ECI's WebScan harvests new links during the scanning process. If you would like to see what these links are, then select this option. Without this option being selected, only the problem links will be displayed as their turn in the scanning process comes up. Duplicate links are links that have already been found during the scanning process. Duplicate links are not harvested or rescanned. If you would like to see the number of duplicate links during the scanning process, then select this option. Otherwise these links will simply be ignored.

Focus and control the scan

The default settings for scan options set the stage for a scan that meets most user needs.

Limit scan to links at or below the initially selected web page: This option insures that the direction of scan stays within the same focus of the initial web page. Links found that do not start with the same domain and path as the initial web page will be ignored.

Limit traversal of foreign links: This option insures that scanning does not branch out in all directions onto all web sites referenced during the scanning process. Reasons for deselecting this option might include speed reading a web site's strategic relationships as they become exposed through cross-linking. You might also be interested in insuring the quality of the web pages your own company is linked to insure that bad links are not easily found by your own customers as they traverse your partner or supplier links.

Include the query: This option allows you to disable the query string that sometimes is appended to the web page being scanned. Sometimes the query string creates a problem with ECI's WebScan search process such as a session ID that changes.

Include the anchor: This option allows you to disable the anchor fragment attached to some web links. One reason to disable this fragment would be to increase scanning speed.

Increase scan speed: This option enables a time-out parameter during opening a reading of new web pages. Some pages take a long time to open while some pages that do not exist anymore require long wait periods before final result is know. Most pages do not take a long time so scanning speed can be drastically improved at the expense of aborting some pages that would otherwise be normal.

Analysis options

Currently there is only one analysis option. In addition to real time scanning and reading of links, ECI's WebScan can generate a table showing various quality issues and possible problems discovered during the scan. See "Figure 2 - Sample ECI WebScan Result" for an example table.

The table summarizing bad and suspicious links can include the following issues:

- Bad link
- Bad embedded link within a PDF file
- Forbidden link
- Malformed link
- Domain not found
- Link skipped because of a time-out condition
- Link skipped because it is not a text or pdf formatted file
- Other page opening errors
- Other page errors

Ignore selected web pages

Sometimes there is a bad link or server-side program that creates a problem for ECI's WebScan so that scanning may not complete properly. Use this data entry field to list these problem pages. In this way scanning can be successful under most any circumstance.

Start the scan

Once all form fields have been set to your preferred values, click the submit button to initiate the scan. Scanning will stop when all links have been scanned to the depth you select, or when an error condition is encountered.

The most likely error condition would be that all available memory has been consumed. Contact ECI to request for memory or other focus factors to meet your needs. Requests will be considered for community sake and under sponsorship.

If scanning stops for an out of reserved memory condition, the report will still report on links scanned so far. Some errors may occur that have not yet been trapped or handled gracefully by ECI's WebScan. In this case an error message and php program line number will print and the scan will simply stop without any report of activity. We would be interested in receiving reports on such conditions as part of our on-going effort to improve ECI's WebScan.

Chapter

3

Where to go for more information

Get the service you need by contacting ECI

Contact ECI to get more information on using ECI's WebScan tool.

info@EmbeddedComponents.com

Try ECI's WebScan landing page for terms of use, user profile, and specific version details:

http://www.embeddedcomponents.com/test/webscan.php