

What WebStripper Does

WebStripper downloads web pages and sites to your hard disk then provides you with an easy way to manage and browse the downloaded sites.

How WebStripper Works

WebStripper finds files on the target site by downloading HTML pages and searching them for links to other pages and files. This means that, even if you only want say graphics files from a site, WebStripper must also download the HTML pages in order to find the links to the files. It also means that WebStripper can't find files which aren't linked to by other files on the site.

Once all the files on the site have been downloaded WebStripper 'parses' the files and replaces any links with links to the newly downloaded files. This means that you can browse the downloaded sites using any browser and you can even copy them to a removeable drive or CD-R to save disk space or view them on another machine.

Free for Personal Use

WebStripper Free can be used for free by private individuals. If you are using it within a business you must pay for and upgrade to WebStripper Pro. This applies to both release and beta versions.

The Main Window

The main window is divided into four sections.

The Advertising Banner (Free Version Only)

WebStripper is free for personal use thanks to the adverts shown here. If an advert interests you just click on it, as you would an image in your browser, to see more information. If you'd rather not see the adverts you can [buy](#) a version without adverts.

The Toolbar

This is divided into sections according to the function of the buttons. The left hand section controls the browser. You can also use the back and forward buttons to move around the file lists. The right hand section is for adding a new site, changing options for an existing site and starting a download.

The Site Tree

Displays an explorer-like tree of categories and sites. Click on a **site** to see details of the site inside it. Click on a site to browse it or view details of files. Right click on a **category** to add and delete categories and sites. You can re-arrange items using drag and drop.

The Browser and File List.

The main part of the site has a number of tabs for browsing the site and the files in it. The **Browser** tab shows the site itself. **File list** shows all the files in the site and **Whats new** shows any new or updated files found during the last update. On both of these pages there is a **File types** box for selecting what type of files to list. The **Errors** page shows any files which could not be found or had problems for other reasons. The second line of each entry give details.

You can use the **Back** and **Fwd** buttons to move around both within the browser and between list pages.

For example if you double click a file in the file list to browse it you can use the **Back** button to return to the file list ready to select another file. The **Browser** page also has **Prior** and **Next** button which move you through the file list or whats new list one item at a time - they apply to the last list displayed. You can, for example, select File types: Text in the file list and then use the Prior and Next buttons to move back and forth through each HTML page in turn.

On any of the list pages you can change the width of each column by positioning the cursor over the break between columns in the header and holding the left mouse button down while dragging the column to its new width. Click on a column heading to sort on that column. Click again to change the sort order. You can also rearrange the order of the columns by dragging a column heading and dropping it in a new position.

Site Management

WebStripper includes a number of features to make managing sites easier.

Adding a site

Select a category in the site tree and click on the [new site](#) button.

Changing site properties

Select a site and click of the [properties](#) button.

Deleting a site

Select a site and click the [delete site](#) button. This brings up the [delete site](#) window in which you can select whether to delete the site from the WebStrippers listing, the files from disk or both..

Moving a Site in the Tree

Use drag and drop to move sites around the site tree.

Repairing and Reparsing

Repairing

In rare circumstances an unrecoverable error may occur while downloading a site (eg running out of hard disk space). If this happens you will be left with many files on your hard disk which are not browsable - they usually have the .temp extension. The files still need to be parsed and given their proper filenames.

You can do this using the **Repair** command on the **Utilities** menu.

If the site was downloaded by V1.1 or later (or V2 beta 4 or later) WebStripper reads a file containing recovery information which was written while downloading. For earlier versions WebStripper searches the sites folder on the hard disk and attempts to re-construct the filenames (and the original web addresses) from them.

Reparsing

Some of the options in WebStripper affect what information is written into the downloaded files when they are parsed - most notably the [Dial Protection](#) settings. If you change these settings they only affect newly downloaded pages. To change the settings in an already downloaded site use the **ReParse** item on the **Utilities** menu.

Related Topics:

[Importing sites](#)

Adding a Site

To start a new site click the **New Site** button in the main window. If you have a browser open the address of the current page will be grabbed from the most recently used browser window. The site will be placed in the currently selected category in the site tree.

The New Site window contains a number of pages of settings. You can navigate between pages using either the **Back** and **Next** buttons or by clicking directly on the links to the left of the window. You can click the **Finish** button when you have set all the options you want to change (you don't have to go through all the pages). Checking the **Advanced** check box brings up some pages of more advanced options.

On the [address page](#) either keep the address WebStripper has found or type in your own. WebStripper will create a site name and folder for you or you can change these to your own choices.

On the [Range page](#) **Single page** will bring down a single page, whole site will bring down a whole section of a site or the whole site - check **Sub-pages only** to only bring down only pages in sub-folders of the current page, uncheck it to bring down the whole site Use the **Advanced** button for some more [advanced range settings](#). Note that the Range settings only apply to HTML pages, WebStripper will bring down images necessary to make up the page from anywhere on the server or other servers. **Type of file** gives you the option to ignore image files for a faster download. There are also some [advanced settings](#).

The **Strip Now** page gives you the choice of downloading straight away or doing it manually later.

Downloading

Download Status Window

The Download Status window shows the progress of your downloads.

The **Threads** page shows each thread which has been created for downloading. You can adjust the number of threads using the 'spinner'. New threads will be created immediately, but old threads won't be removed until they have finished downloading their currently active file. **Bytes received** shows the amount of data received this session. Click the **Details** button to see the individual threads.

The **Sites** page shows statistics for individual sites. Select from the drop down list of active sites for more information. Bytes received shows data received this session.

The **Parsing** page shows the progress of parsing the current site and the number of sites waiting to be parsed.

Errors shows any errors that have occurred since the **Clear errors** button was last clicked. **AutoLog** details any queued operations since the **Clear log** button was last clicked (Queued operations are in the Pro version only).

The **Pause & Parse** button stops downloading and tells WebStripper to parse the sites making them browseable. **Quick Pause** forces WebStripper to stop downloading immediately and not to parse the files. This means they won't be browseable until they have been parsed either by separately parsing them or by resuming the download. See also: [What WebStripper Does](#) and [View paused Sites Window](#).

Downloading Action

For each site WebStripper stores an '**action**' which is one of:

Strip: The website will be downloaded as though no files have already been downloaded.

Update: The website will be checked for new or updated file and only these will be downloaded.

Resume: The site has been paused and WebStripper will pick up where it left off.

WebStripper automatically gives each site the appropriate action but you can change this from the Sites, Action menu. The current action is displayed on the Go button.

More Than One Site

You can start as many sites downloading at one time as you like. WebStripper creates one set of threads and allocates each thread to any site which has files waiting to be downloaded. Each site will be parsed and become available for browsing as it is finished.

Queueing Sites: Scheduling and AutoUpdate

(WebStripper Pro only)

Queueing is a very powerful feature for automating the operation of WebStripper. Scheduling a site will download or update it at a particular time. Using AutoUpdate WebStripper will visit a site at regular intervals to download the latest updates. You can set as many sites to be downloaded at one time as you like. WebStripper will allocate it's downloading threads to whichever site needs them.

To set a site as Queued go into the [Site Wizard](#) (New Site or Site Properties window) for a site, check the **Advanced** check box and change the settings on the [Schedule](#) or [Update](#) pages.

Once you have set the options you can shut down WebStripper and it will automatically start up at the scheduled date and time, perform the download and, if you've told it to, close down afterwards. There is no need to leave WebStripper running (check the options below for more information).

You can see which queued operations WebStripper has performed from the **AutoLog** page of the download status window. If the window isn't currently being shown use the menu item **Downloads, Download Status Window**. To see which sites are queued for download see the menu item **Download, [View queued sites](#)**.

Options

The [Options window](#) (menu item Options, Options, Queued sites) contains options for Queued sites.

If **Start WebStripper automatically if tasks are due** is set then there is no need to leave WebStripper running - it will automatically start up when a task is due.

If you are using a dial up connection there is an option to connect automatically or to ask first. Set it to connect automatically if you want WebStripper to download sites during the night or early in the morning. If you're using a dial up connection it's best to set all your sites to be downloaded at the same time.

The next set of options set various timeouts. We recommend that you leave one or more timeout switched on just in case a site causes WebStripper problems. Finally there are options for what you want WebStripper to do once it has finished downloading. If you set WebStripper to close the internet connection it will do this as soon as the sites have finished downloading (and before the sites have finished being parsed.)

The Site Wizard

The site wizard windows displays itself in a number of ways including the **New Site** windows and the **Site Preferences** window. The Site Wizard displays different sets of pages depending on what information it is collecting.

Navigate around the Site Wizard using the **Back** and **Next** buttons or by directly selecting a page from the list of pages on the left. You do not need to visit every page in turn, just those where you want to change the defaults and you can click the **Finish** button whenever you are done.

The **Advanced** check box on the bottom left displays or hides a number of pages with more advanced settings.

For more information see each individual page:

[New Site: Address](#)

[Site Properties: Address](#)

[Import Site: Address](#)

[Range](#)

[Advanced Range](#)

[Advanced Type](#)

[Filters](#)

[Schedule](#)

[Update](#)

[Options](#)

[Strip Now?](#)

WebStripper Options

You can change the way WebStripper operates with the options window. At the main screen select the menu item **Options|Options**.

Connection

Retries. Sometimes an error will occur whilst downloading a page. This option defines how many times WebStripper should attempt to strip a page before giving up. Recommended value: 99

Timeout. You may have noticed in your browser that sometimes a page 'sticks' and refuses to download fully and you have to press the reload button. This option to tell WebStripper how long to wait (in seconds) before pressing it's internal 'reload' button. Recommended value: 180 seconds.

Max number of parallel downloads. WebStripper makes downloading faster by stripping more than one file at once. This value is the number of files to strip at once. The recommended value depends on the speed of your internet connection and the type of sites you download. With a 33.6k connection use a value around five, at 56k about eight is a good start. If the sites you strip have many small files (say a largely text only site) or you have a slow internet connection you may want to increase these figures.

Alternatively if you'd prefer WebStripper to operate unobtrusively in the background while you use the web for other things you may want to set this figure to a lower value.

Use a dial up connection. Set this option if you use a dial up connection, clear it if you have a permanent connection.

Default connection. This is the dial up connection WebStripper will use.

Silent connection. Use this option to bypass the dial up confirmation boxes.

Disconnect when done. Does exactly what it says: hangs up the phone when stripping is finished. Use this option for unattended use. See also the disconnect when done button on the [main window](#).

Use a proxy server. If you access the internet through a proxy server you will need to set these options. **Proxy** is the name of the proxy server (e.g. 'proxy.myisp.net'). **Proxy port** will usually be 80. If your proxy server requires a **username** and **password** enter these in the boxes provided otherwise leave these boxes blank.

Firewall

If you connect through a firewall enter your settings here.

New Site

These options set the initial values for when a new site is added. **Default root folder** is the folder under which your stripped site will be placed. The other options are the same as those in the [New site](#) window.

Listen to clipboard. Turns on or off the [clipboard monitor](#).

Browsing

Dial protection. Changes [dial protection](#) settings. **None:** Turns off Dial protection. **When page loads:** Prevents dial up when a page is loaded in your browser. **When following a link:** Asks for confirmation before following a link to a page which was not downloaded.

If **Use built in browser** is checked WebStripper will browse sites within it's own window. Uncheck this if you prefer to use an external browser. **Home button moves to root page of current site** affects the Home button within WebStripper. If **Back and Forward only affect browser** is checked the Back and Forward buttons will only move around within the browser, not between tabs. Check this if you use

Internet Explorer 3.

Queued Sites (WebStripper Pro only)

These options affect sites which are set for scheduled downloading or to be autoupdated. For more information see [here](#).

General

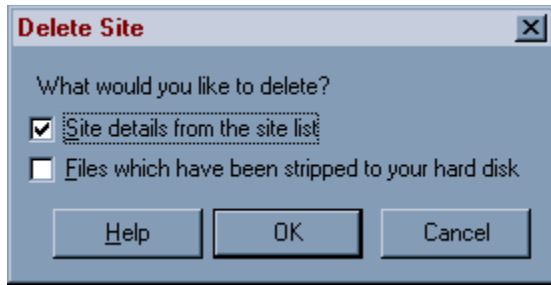
Deleted items go to recycle bin. When you delete a site files will be sent to the recycle bin if this option is checked.

Send windows file notify messages. If this option is on WebStripper will send messages to Windows Explorer whenever a file is created, deleted or renamed. This keeps any open Explorer windows up to date, but can also slow down you machine.

Show agent as. Every time you visit a web page your browser (or WebStripper) sends details about itself. This doesn't normally cause problems but occasionally a site will only work with specific browsers. If this is a problem you can change the agent string which WebStripper sends. Either select one of the default options or enter a string of your own. Please only use this feature when absolutely necessary - it's always best if browser identify themselves correctly to the server.

WebStripper periodically visits the solent software website to check for news about WebStripper and related products. To stop seeing news items uncheck **Show news**. Occasionally WebStripper may add special items to the site tree, such as special offers on other products. In WebStripper Pro you can turn this feature off by unchecking **Show special tree items**.

Delete Site



Confirm whether you want to delete the site and what you want to delete.

Check **Site details from list** to remove the site from the site list in WebStripper. The stripped files on your disk are not affected.

Check **Files which have been stripped to your hard disk** to physically delete files from your hard disk. If the option **Deleted files go to recycle bin** is checked in the [options window](#) the files will be sent to the recycle bin and can be undeleted later, otherwise they are removed permanently.

Import Site

This feature can be used to:

- Add a site which has been deleted from the site tree (Provided the downloaded files are still on disk).
- Import a site which was downloaded by another offline browser.

Notes:

If you import a site from another offline browser you can only update the site if the site was downloaded using the same directory structure and filenames as used on the server. This depends on the offline browser you used. WebStripper imports every file in the directory you specify, even those which are not linked from the site. You cannot import just a subsection of the files.

This window works is very similar to the [Site Wizard](#). Jump to the pages you want to set and press Finish when done. The folder you specify is the folder containing the site.

The other pages are: [Address](#) and [Range](#).

Related Topics

[Repairing and Reparsing](#)

New Site: Address

On this page you enter the address of the website. The address may have been filled in for you if you are using the [clipboard monitor](#) or the [address grabber](#). If you want to change this address just overwrite it.

In the first box type the web **address** (URL) of the site, e.g. www.solentsoftware.com. You don't need to enter the <http://> part of the address, though it won't hurt if you do. Notice that the contents of the next two boxes have been filled in for you. You can either leave these with the values WebStripper has filled in, or change them if you prefer.

The **name** of the site is any name you care to give the site. It's the way you'll identify the site in the main WebStripper [site list](#).

The **folder** is the disk folder the site will be stripped into. You should always use a new folder for each site - Most sites contain files with similar names which will be overwritten if you use the same folder more than once. (When WebStripper creates a folder name for you it makes sure the folder is new - you are recommended to leave this value unchanged). To change the initial value of the root folder (c:\Stripped) use the [options window](#).

Some sites require a username and password. If this is the case fill in the details under '**logging in required**'.

Site Wizard: Range

In the **range** section you can specify how far WebStripper is to look for files. The range applies only to HTML files. Graphics or other files which are necessary to complete the display of the page will may be fetched from anywhere. Single page fetches just the page specified. Whole Site fetches every page on the site (or the subsection of the site if Sub-pages only is specified). **Custom** means you have set a more complex range using the [Advanced](#) button.

Sub-pages only forces WebStripper to only look in the specified folder and sub-folders of it. Use this option if you only want to strip a section of a website. (again graphics etc may be fetched from elsewhere).

If **Fetch graphics from other servers** is checked WebStripper will fetch graphics and other items needed to make up a page from anywhere on the internet. If unchecked only items on the same server will be downloaded.

Type of files lets you control which files are downloaded (and therefore the speed of downloading). **All files** fetches any file necessary for the display of the page. **Ignore large files** ignores any file over 100Kb. **Text only** only strips HTML and text files. **Custom** will be checked when you select a more complex option using the [Advanced](#) button.

Site Wizard: Advanced Range

This page tells WebStripper how far to look for files. The range only affects HTML files. Any other files which are necessary to display the page completely will be fetched from anywhere on the server, or on other servers.

Page depth refers to the number of links between pages and is not based on the folders in which the pages are stored. The starting page (eg index.html) has a depth of one. If this links to a page, Contents.html, which links to another page, Dogs.html, then Contents.html has a depth of two and Dogs.html has a depth of three.

Index.html (depth 1) -> Contents.html (depth 2) -> Dogs.html (depth 3)

Setting the page depth to two will retrieve Index.html and Contents.html but not Dogs.html

The first section gives options for the [primary server](#). If you check **Only sub-pages** WebStripper will strip files from the specified folder and any subfolders of that folder only (this only applies to web pages - not sub-items like graphics). If the option is unchecked any files from any folder on the server will be stripped. You can limit the [page depth](#) of a search by checking **Up to a depth of n pages** and entering a value in the box or strip to any depth by checking all pages.

If you only want to retrieve pages from the [primary server](#) check **Don't visit other servers**. If you choose to visit other servers check **Up to a depth of n servers** and enter the number of [servers](#) in the box.

If you chose to visit other servers you can then specify the [depth](#) of pages to strip from other servers by entering a value in **Up to a depth of n pages** or, to strip all pages check **All pages**.

Site Wizard: Advanced Type

This page contains options to speed up the stripping of the site by only stripping certain file types. You can tell WebStripper to ignore certain file types by unchecking options in the **File types to download** section or tell it to not strip files above a certain size by checking the **Don't strip files larger than** option. To change the file size enter a new value in the box.

Use the **Don't strip files smaller than** option to avoid downloading small files. This is useful on sites with lots of small graphics which you'd rather not waste time downloading. This option does not affect HTML files which are always downloaded however small.

Site Wizard: Strip Now?

If you want to strip the site immediately select Yes. If you'd rather wait till later, select No.

Site Preferences: Address

Address is the web address (URL) of the site or page. You don't need to give the http:// part of the address, though it won't hurt if you do.

The **name** of the site is any name you care to give the site. It's the way you'll identify the site in the main WebStripper [site list](#).

The **folder** is the disk folder the site will be stripped into. You cannot change this option.

Some sites require a username and password. If this is the case fill in the details under '**logging in required**'.

Import Site: Address

In the first box type the web **address** (URL) of the site, e.g. www.solentsoftware.com. You don't need to enter the `http://` part of the address, though it won't hurt if you do. Notice that the contents of the next two boxes have been filled in for you. You can either leave these with the values WebStripper has filled in, or change them if you prefer.

The **name** of the site is any name you care to give the site. It's the way you'll identify the site in the main WebStripper [site list](#).

The **folder** is the disk folder containing the stripped site. You must enter a folder which contains a site which has been previously downloaded with WebStripper.

Some sites require a username and password. If this is the case fill in the details under '**logging in required**'.

Site Wizard: Filters

(WebStripper Pro only)

This page lets you set [filters](#) for a site. The list at the top shows the filters you have created in the order they will be evaluated (Unless you use [advanced filtering](#)).

Use the **New**, **Edit** and **Delete** buttons to create, change or remove filters. Use **Move up** and **Move down** to change the order of the filters.

To use [Advanced Filtering](#) tick the check box and put your expression in the box at the bottom.

The **Test** button brings up a special [Test Filters](#) window. You can leave the Test Filters window open while you edit the filters.

See also:

[Filter Editor](#)

[About Filters](#)

Filter Editor

Use this windows to add or edit filters. See the article on [filters](#) for more information on the options.

Test Filters Window

Use the Test Filters window to check the operation of your filters. It works for both normal and advanced filters.

Set up your filters on the filters page of the [Site Wizard](#). Click the Test button to show the window. Enter a complete web address in the address box (ie <http://somewhere.com/file.html>). If your filters use mime types put a value in the mime type box. If your filters use size put a value in the size box.

The line at the bottom shows whether the file would be downloaded using your filters. If the results aren't what you want move the test filters window to one side and go back to the site wizard. Edit the filters. The Test Filters window will be constantly updated.

Site Wizard: Schedule

(WebStripper Pro only)

Use this page if you want WebStripper to download the site at a specific date and time. For more information on how this feature works see [queueing sites](#).

To download as soon as you click the **Finish** button select **Now**. Check **Only when I press to Go button** if you want to start the download manually. To set a scheduled date and time select **On this day** and choose an appropriate date and time. If you use a dial up connection it is best to set all sites to download at the same date and time.

Site Wizard: Update

(WebStripper Pro only)

The options on this page can be used to get WebStripper to update a site at regular intervals. For more information on how this feature works see [queueing sites](#).

Date

If **Only when I press the Go button** is checked WebStripper won't automatically update the site. To download on the same days each week (eg. weekdays only) check **On selected days** and check each day you want the site updated. To download on the same day each month check the **On the nth of every month** option.

Time

To download the site once each day set the **At this time** option. To download the site at regular intervals throughout the day, say every hour between 9am and 5pm use the **Every n minutes...** options.

Site Wizard: Options

This page contains some advanced options for fine tuning the way WebStripper works. You should leave them alone unless you really want to change them.

When updating use quick update (default: checked). When you update a site with this item checked WebStripper will only check the HTML files for updates. This is the best setting for sites which rarely delete content - WebStripper will find any new or changed pages and also find any news graphics etc linked to them. Uncheck this item if you have a site which changes regularly (eg a new site) and you want to delete the old content when it is removed from the server (see 'When updating delete files no longer on the server' below).

Change file extension to match type (default: checked). When a server returns a file it also returns a file type. Normally this matches the files extension. However, when the file is returned from a server side script (ie a database driven site) the returned file type won't match the extension. It is impossible to know which files are returned from scripts and which are straight files so WebStripper checks every file and if it's file extension doesn't match the file type it changes the file type appropriately. Sometimes this can cause problems, mainly due to the registry on a machine not being set up properly. If you experience problems with file extensions being changed inappropriately try unchecking this item.

When updating delete files no longer on the server (default: unchecked). Use this option for sites with changing content where you want to conserve disk space by deleting old files.

Copy site (Don't ReParse) (default: unchecked). Normally WebStripper reparses a site after it has been downloaded to change any links into relative links (instead of links onto the internet which were in the original files). If this option is checked WebStripper will not make any changes to the downloaded files. Use this if you want to edit and re-upload a site. Bear in mind that most of the internet is copyright protected and you should obtain permission from the copyright owner before re-using material.

View Paused Sites

This window lists all sites which are currently paused.

For each site you can **Remove** it from the paused sites list, **Parse** it (if it has been paused without parsing, or **Resume** the download. There are also buttons for performing one of the above operations on all paused sites.

If the site is removed it's action will still be 'resume' but you will not be reminded about it at startup any more.

View Queued Sites

This window lists all queued sites along with the time the site will next be downloaded and the action to be performed (**AutoUpdate** or **Scheduled** download).

The **Unqueue** button removes the queueing from the site (ie. the settings for the site will be changed to 'Only when I press the Go button').

Site Wizard: MultiAdd

Raw Site Copy

Normally WebStripper downloads a site with the aim of making it browseable on your PC, rather than over the net. This means it has to make some changes to the code of the site. If you want to download a site without making any changes to the code you can use the **New Site|Raw Site Copy** option. Use this, for example, to download a site before uploading it to another server.

The other options for this feature are the same as for [New Site](#).

Primary Server and Other Servers

The **primary server** is the server named in the site address you give WebStripper. E.g. in the address <http://www.solentsoftware.com/webstripper/index.html> the www.solentsoftware.com part is the name of the primary server.

Other server refers to any server other than this one. e.g. if the above page referenced a page <http://www.elsewhere.com/index.html> then, since the www.elsewhere.com part is different, this is on another server.

Page Depth

This refers to the number of hops between pages. For example if page A references page B then page B has a depth of 1. If page B then references page C then page C has a depth of 2. However if page C is also referenced by page A then C has a depth of 1.

Note that this doesn't apply to sub-items within a page, such as graphics, which are always at the same depth as the page itself.

Server Depth

This refers to the number of hops between servers. For example if a page on server a.com references a page on server b.com then server b.com has a depth of 1. If any page on server b.com then references a page on c.com then server c.com has a depth of 2.

However if server c.com is also referenced by a page on a.com then c.com has a depth of 1.

Clipboard Monitor

The clipboard monitor listens to the clipboard. If you cut or copy a web address it will ask you if you would like to add the site for stripping.

To turn the feature off go to [Options|options](#) and go to the New Site page.

Address Grabber

When you select New Site the address grabber fetches the address from the last browser window and adds uses that as the site to strip. If you wanted a different address you can overtype the grabbed address.

Dial Protection

Some webpages, when viewed in your browser, will try to open an internet connection. A connection will also be opened when you click on a link to a file that was not downloaded. You can use Dial Protection to make your browser prompt you before opening a connection.

There are three levels of Dial Protection. **None** leaves the pages as they are - no Dial Protection.. **'When page loads'** stops your browser opening a connection when you open the page (Common causes for a page to open a connection are graphic files which were not downloaded and page counters). **'As above plus when following a link'** also forces your browser to ask you whether you want to open a connection when you click on a link to another page.

Dial Protection works by inserting code into your downloaded web pages. If you change the settings they will only be used for pages you download after changing the settings. To change the dial protection settings for a site which has already been downloaded:

- Change the dial protection settings.
- Use the menu option **Utilites, Re-Parse** to change the settings within each page.

The settings can be changed from the [Options](#) window.

Filters

(WebStripper Pro only)

The [Range](#) options give you basic options for setting where you want WebStripper to look for files but sometimes this doesn't provide enough control. On these occasions you need to use filters. Using filters gives you very accurate control over what is and what is not downloaded based on the server, path, file name, file extension, [query string](#), [mime type](#) and file size.

When testing filters you should bear in mind that WebStripper always has to download HTML pages in order to find links to other pages. These will only be downloaded from directories you have told WebStripper to look in

A filter consists of four elements:

- The [type](#) defines whether the files will be included or excluded, or are required.
- [Field](#) is the part of the web address, mime type or file size that the filter tests.
- Match is the kind of test to apply. It can be equals, not equals, contains, doesn't contain, is larger than or is smaller than (the last two only apply when the field is size).
- Value is the string to match or a file size in Kb.

The Effect of Range and Type Settings

When you use filters WebStripper only looks at the files which have already been selected by the Range and Type settings. If you set the Type to text only then the filters will not start downloading non-text files. Filters just add an extra layer of selection to the download process. This is important to bear in mind if you filter on the server. You must tell WebStripper to visit other servers in the Advanced Range page.

Advanced Filters

The normal filters are usually all that is needed but if you want really fine control over which files are downloaded you can use [advanced filters](#).

Filter Types

(This applies only to normal filtering - [advanced filtering](#) is handled slightly differently).

A filter can be one of three types. Each filter is tested in order from top to bottom as shown in the filter list. The type of filter and it's placement in the list can be used to combine filters to produce different results.

- Include downloads a file if the filter matches. The file will be included even if it doesn't match any of the filters above it.
- Exclude filters specify that a file will not be downloaded. Note that the filter only applies to files which would have been downloaded by filters listed above it - it doesn't affect those below. If the first filter is an exclude filter then any file except those which match the filter will be downloaded. This is equivalent to the '-' option on an altavista search. (HTML files will still be downloaded)
- A require filter tells WebStripper to only download a file if a) It is already matched by the filters preceding it and b) it matches the require filter. This is equivalent to a '+' option on an altavista search.

Example 1:

Include when folder item equals dog

Exclude when filename contains retriever

will download files under a folder called 'dog' except those called 'retriever'.

Example 2

Include when folder item equals dog

Exclude when filename equals retriever

Include when filename equals training

As example 1 but downloads any file whose name is training even if it's not in the dog folder.

Example 3

Include when folder item equals dog

Require that filename equals training

Only download files called training in folders called dog.

Filter Fields

WebStrippers filters can work on a variety of 'fields'. Using the web address

'<http://www.solentsoftware.com/webstripper/support/search.html?type=abstract>' as an example:

- Filename is the name of the file excluding the extension, eg 'search' (see also entire filename).
- Extension is the file extension (excluding the period) eg 'html'.
- Entire filename is the filename plus extension eg 'search.html'.
- Path is the path to the file excluding the server and filename, eg '/webstripper/support/'. Generally it's easier to use folder item.
- Entire path is the entire path including server and filename but not the query string. Eg '<http://www.solentsoftware.com/webstripper/support/search.html>'
- Folder Item: Each path can contain more than one folder item. Each folder item is one folder in the path. In the above example both 'webstripper' and 'support' are folder items.
- Server is the server name, eg www.solentsoftware.com. To use this field you must tell WebStripper to visit other servers in the [Advanced Range](#) page.
- A [query string](#) is the part of an address after the '?' which is passed to a server script, eg 'type=abstract'
- [Mime type](#). When WebStripper downloads a file it retrieves information from the server about the files type. This is called a 'mime type'. You should usually use extension instead - it's faster - but mime type can be useful, for example, for html files which can have many different extensions including html, htm, asp, shtml, dhtml among others.
- Size is the size of the file in kilobytes (Kb)

For most of the fields WebStripper can decide whether or not to download a file just by looking at the address - which is very fast - but for the last two (Mime type and size) WebStripper actually has to access the server to test whether the file should be downloaded which, obviously, takes longer. It's not always easy, but avoid these if possible.

Advanced Filters

Using advanced filtering you can build complex expressions using [boolean expressions](#) (expressions combined with and, or, xor and not operators).

Before using advanced filters you should ensure that all filters are include filters. If you use exclude filters you'll probably get confused. Required filters act the same as include filters.

When you select the Use advanced filtering option each filter in the list is assigned a number. The numbers are combined with the operators (not, and, or, xor) and brackets to form expressions.

Consider the following filters:

- 1) When mimetype equals text/html
- 2) When folder item equals dogs
- 3) When folder item equals cats
- 4) When filename equals training

To download html files about dogs use the expression:

1 and 2

To download html files about cats or html files about dogs use:

(1 and 2) or (1 and 3)

To download files about dogs but not files about training use:

2 and not 4

Boolean Expressions

In a boolean expression each item has a value of true or false and the result is true or false.

Values are combined with one of the following operators:

<code>a and b</code>	Returns true if both a and b are true.
<code>a or b</code>	Returns true if either of a or b are true.
<code>a xor b</code>	Returns true if either a or b is true, but false if they are both false or both true.
<code>not a</code>	Returns true if a is false or false if a is true.

In WebStripper 'not' is evaluated first. The other operators are evaluated from left to right. You can alter the precedence by placing sub-expression in brackets.

Examples

<code>not a and b</code>	True if a is false and b is true.
<code>not (a and b)</code>	False if both a and b are true, else true.

Query String

If a file is running a script on the server to return information from a database it may have a question mark (?) followed by some parameters. The parameters are called a query string. In WebStripper the query string doesn't include the question mark.

Mime Type

Every file returned by a server has a mime type. The mime type tells WebStripper what kind of data the file contains.

Some common mime types are:

text/html	HTML files
text/plain	Plain text
image/jpeg	JPEG picture
image/gif	GIF picture
audio/wav	Wave audio file

Frequently Asked Questions (FAQ)

I get the error message 'Unable to establish a connection. I use AOL (America Online).

From the main menu select Options, Options and uncheck the item 'Use a dial up connection'.

I have a permanent net connection but WebStripper tries to use dial up networking.

At the main window go to Options:options and uncheck the item use a dial up connection. See also the next question.

When I click on a link I get asked if I want to open an internet connection. How do I stop this?

This is called [dial protection](#). You can change the settings by opening [Options|options](#) and going to the **general** page.

For a larger and more recently updated selection of answers visit our [support website](#) and FAQ.

WebStripper Pro

WebStripper Free is for personal use only. If you want to use WebStripper in a business environment you must upgrade to WebStripper Pro.

WebStripper Pro is an enhanced version of WebStripper with the advertising removed and extra features.

Bonus features currently in WebStripper Pro:

- The ability to schedule downloads at any time of the day or night
- Filters which let you filter pages based on very fine requirements.

For the latest information on bonus features check out our site at

<http://webstripper.net/buy.html>.

Contacting Solent Software

Email. Sales queries: sales@solentsoftware.com, [Technical support](#).

WebSite. <http://webstripper.net/>

Post.

Solent Software
Daisy Villa, Brook Avenue,
Warsash,
Southampton,
Hampshire.
SO31 9HP
England.

Credits

WebStripper was conceived, designed and written by Mike Sutton.

The following tools and components were used during development.

Delphi by Inprise/Borland.

Compiler and RAD tool.

www.borland.com

HelpScribble by Jan Goyvaerts.

Help file authoring tool.

www.jgsoft.com

ICS (Internet Component Suite) by Francois Piette.

Internet access components including WinSock and HTTP.

<http://www.rtfm.be/fpiette>

RasComp by Magenta Systems.

Dial up access component.

<http://www.magsys.co.uk/delphi/>

Drag and Drop Component Suite by Angus Johnson & Anders Melander.

<http://www.melander.dk/>

Bome OneInstance by Florian Bomers.

Run only one instance of a Delphi application.

<http://www.bome.com/>

Support

On-line Help and FAQ

If you have any queries about WebStripper the first place to look is this help file. Just press the F1 key or click any of the Help buttons. The help file gives full details of each window and answers all the most [frequent asked questions \(FAQ\)](#). Use the Contents or Index buttons above for more help.

WebSite FAQ

This is where we post the most recent information. <http://webstripper.net/support/>

Mailing List

The **Wbs-Announce** mailing list gives details of the latest versions and any other important news. Visit our online [sign up page](#) for full information.

The News Page

If you think you've found a bug the chances are someone has already reported it. This page of our website contains details of any known bugs as well as the latest bug fixes and new feature announcements.

Email

If you still can't find an answer to your problem or you have found a new bug then sent an email to webstripper@solentsoftware.com. Make sure you include the following information:

- Your operating system.
- The version of WebStripper you are using. This is displayed in the About window: Select Help|About.
- The address of the website you where trying to access.
- A description of the problem. Include any information you think may be relevant including a step by step description of what you did and any messages displayed by WebStripper.

'Unable to access advertising server' message

The free for personal use version of WebStripper is funded by advertising revenue. For this to happen WebStripper needs to be able to access a server across the internet to pass back statistics about advertising shown and to download new adverts.

If you see this message WebStripper has been unable to access the advertising servers and, since the advertising pays for WebStripper, it won't let you download or update any sites.

You may see the message if some files vital to the advertising component have been removed by accident, so try reinstalling WebStripper.

If you use a proxy server then reinstall WebStripper. You will be asked for your proxy server settings during the install process. Double check that they are correct.

The other common cause of this message is if you have a firewall or proxy server which isn't configured to allow WebStrippers advertising component to access the internet or if access to the advertising servers has been blocked.

If your firewall blocks selected programs to need to set it to allow TSAdBot.exe access to the internet. If your firewall/proxy blocks sites based on servers you need to make sure it can access content.conducent.com and bootstrap.conducent.com

If that doesn't cure the problem or you don't have access to the firewall/proxy server settings you need to ask your network administrator to change the settings for you.

If this doesn't help you may continue to use WebStripper by buying [WebStripper Pro](#).

Check the internet for [updates](#) to this information.

Error Running Browser

WebStripper was unable to start up your browser. Check that you have correctly installed the browser.

Function Not Available

The function you have selected cannot be performed on a site while it is being stripped. Either wait until stripping is complete or press pause and then resume the stripping later.

Wrong File Type

The file specified in the web address is of a type which you have told WebStripper not to strip for this site. E.g.. you have told WebStripper not to strip graphics but entered a filename with an extension of .gif. Click on Site|Properties and either change the file name or the options on the **Files** page under **File types to download**.

Unrecoverable Error

Whilst stripping an error occurred from which WebStripper could not recover.

This error is very rare and usually means you have run out of disk space while WebStripper was performing a critical operation.

The nature of the error meant that WebStripper was unable to make the stripped site browsable and you will have to strip the site from fresh after freeing up some disk space.

Move the Stripped Site?

In the [Site Preferences](#) window you have changed the destination folder for the site. If you answer yes any stripped files in the old folder will be moved to the new folder. If you choose not to move the files you will not be able to update or resume stripping the site.

Select Directory

Select a directory in which to place the stripped site.

Browse the list of directories in the **Directories** list for a directory to place the site in. Double click on a directory to open it.

To create a new directory type its name into the **Directory Name** box. **You are strongly recommended to create a new directory for each site you strip.**

You can change to a new drive using the **Drives** list.

Proxy Port isn't a Number

The **Proxy Port** box must either be blank or contain a number. The usual value for proxy port is 80.

See also: [Options window](#).

Setting action to strip...

WebStripper sets the action for a site automatically depending on whether the site is paused (action: resume), waiting for an update(action: update) or new (action: resume). If you change the action to Strip any files already downloaded for the site will be deleted and the files will be downloaded over again. It is recommended that you leave the action at it's default.

One or more errors occurred...

An error occurred while the site was being parsed. The file in which the error occurred may not be browseable, but apart from that the site should be OK. This usually means there was an error writing to disk or a file wasn't found. You might like to try reparsing (Utilities, Re-Parse) the site and see if the error recurs.

A disk error occurred...

A disk error occurred whilst downloading. This usually means you have run out of disk space. The download operation has been aborted.

Free up some disk space, then go to the menu item **Download, View paused sites** and resume downloading for this site (or all sites).

