

Offline Browsers

Compiled by Grigory Naumovets <gri@aiha.kiev.ua>.

Offline Browsers.....	1
Why do you need Offline Browsers?	1
Internet Explorer v.5&6: Saving a "complete" Web page.....	1
Internet Explorer v.5&6: Viewing files from cache memory in offline mode	1
Offline Browsers and how to find them	2
Saving a Web site: Data required by an offline browser	2
Downloading files with WebStripper.....	2
Entering parameters for a new task	2
Automatic dial-up Internet connection.....	4
Settings for download through a proxy-server	4
Dial protection (links to missing files).....	4
Update	4
WebZip and Teleport Pro.....	4

Why do you need Offline Browsers?

If you make a copy of a Web-site on your local disk, then you would be able to browse it without connecting to the Internet. However, a Web-site may consist of many files located in a complex tree of directories. Just downloading all these files is not enough – the links would work only if each file is placed to the appropriate folder. Doing it "manually" with a Web-browser is a very laborious task which becomes virtually impossible for a large number of files. However, this job can be automated with the help of Offline Browsers.

Internet Explorer v.5&6: Saving a "complete" Web page

Internet Explorer (IE) v.5&6, as distinct from Internet Explorer v.4 and Netscape Navigator v.4, can save Web pages to a disk in different ways -- "complete" or "HTML only". If we open any HTML-document with IE and select in the *File -> Save As* menu the option *Save as type: Web page, HTML only*, then we would save to a disk only the HTML document. Selecting *Save as type: Web page, complete* will save both the HTML file and the pictures that should be shown on this page. These pictures will be saved to a new folder with a name based on the name of the HTML file (for example, if this name is "testdocument.html", then the additional folder will be named "testdocument_files"), and all references to these pictures in the mail HTML document will be rewritten so as to include the name of this folder. (For this reason, the file saved to the local disk will not be the exact copy of the file downloaded from the server). Note also that by default IE suggests for a "Web page, complete" a file name based on the document title instead of the original HTML file name. If we save other "complete" HTML documents from the same site, then every time we will have a new additional folder with all the pictures (if one and the same image is shown in several documents, a copy of it will be saved every time into each additional folder). Links from one document to another may not work. So, this function of IE is convenient for saving an individual Web-page with images, but inefficient for saving multiple files linked to each other.

If we select in IE menu *File -> Save As* the option *Save as type: Web archive, single file*, then a Web-page with all the images will be saved as one file with extension .mht. This format resembles an email message with the main HTML document in the body and images attached in MIME encoding. An *.mht file can be opened in IE or Netscape Navigator and viewed as a complete Web-page with images.

Internet Explorer v.5&6: Viewing files from cache memory in offline mode

Clicking the *History* button of IE will open a window showing the list of sites that have been recently viewed with this browser. If we set the offline mode through the menu *File -> Work Offline*, we can try to view some of these sites without connecting to the Internet (in this case, IE uses the temporary files stored in the "Temporary Internet Files" folder. This folder and the files can be viewed through the *Tools -> Internet Options -> General -> Temporary Internet Files ->Settings* menu). Those sites that cannot be viewed without connecting to the Internet are shown in a pale font.

Offline Browsers and how to find them

Quite a lot of Offline Browsers can be found on the Internet, for example:

WebStripper (www.webstripper.net)

WebZip (www.spidersoft.com)

Teleport Pro (www.tenmax.com/teleport/pro/home.htm)

Offline Explorer (www.metaproducts.com)

and others.

Also, these programs can be found in the "*Internet -> Web Browsers & Tools -> Offline Browsers*" section of the well-known Tucows site (www.tucows.com). For a fast download, you can choose a mirrors of Tucows in your country.

Saving a Web site: Data required by an offline browser

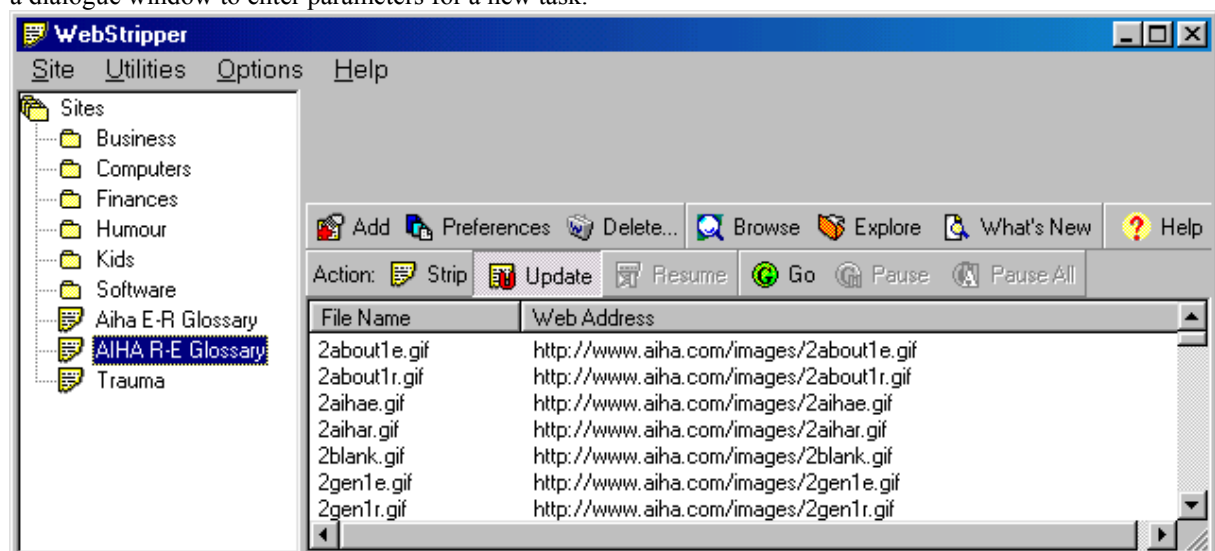
First of all, the offline browser should know the URL of the main Web-page to start the download from. The program analyzes this document and makes up the list of all links in it. These links can point to other HTML documents, graphic, sound, text and other files. These files can be located either on the same Web server or on a different one. After downloading the HTML documents linked to the start page, the program will find there links to other files (so-called "second-order links"), in those files – "third-order links," and so on. It can happen that this process would not stop until you run out of the disk space. This is why one should specify some conditions to define which files should be downloaded, and which should not. Usually, a user is asked to answer the following questions:

- Download only files from this server, or from other servers, too?
- Within this server, download files only from this directory and its subdirectories, or from other directories, too?
- How deep should the offline browser go to follow the "tree" of links? (E.g. up to third-order links etc.; this depth can be different for files located on the same server and on other servers).
- Download all types of files or impose some restrictions (for example, for sound files, pictures, for files exceeding certain size, etc.)?

Downloading files with WebStripper

Entering parameters for a new task

To download a site with WebStripper, open in the *Site -> Add site* menu (or by clicking the *Add* button) a dialogue window to enter parameters for a new task.



In this window, enter (or copy from a browser) the start URL (*Address of the site*), give some arbitrary name to this site (*Name of the site*) and specify the folder to store a copy of the site (*Folder to strip the site into*).

If access to this server is password protected, enter in the window "If logging in is required" the username and password (do not confuse these with those used for dial-up connection with your Internet provider!).

To specify which files should be downloaded, click *Range*.

In the Range window, we can define the range of files to be downloaded (Single page – current page only, Whole site, Sub-pages only – only files from this directory and its subdirectories), and types of files that should and should not be downloaded.

More options can be set with the help of the *Range* -> *Advanced* and *Type of files* -> *Advanced* buttons.

In this window (*Range* -> *Advanced*), one can set the depth to follow the "tree of links" on the primary server and other servers, and the depth to follow the "tree of servers" (if at all).

After setting all parameters for site downloading, click *Strip now* to start the download. The process will be shown in a separate window. It can be paused and resumed.

Upon completion of the download, the local copy of the Web site can be viewed in the WebStripper's window with the help of the built-in browser.

Note that sometimes one site can have several names (aliases), and some links may be made to different names of the same site. In this case, WebStripper would treat them as links to different sites.

Automatic dial-up Internet connection

Automatic dial-up Internet connection can be enabled and disabled through the menu *Options -> Options -> Connection -> Use a dial-up Internet connection* (this connection should be properly configured in the Dial-up Networking). With the option *Hang up when done* enabled, the connection will be interrupted upon completion of download.

Settings for download through a proxy-server

If your computer can download files from the Internet only through a proxy server, the proxy settings should be entered in the menu *Options -> Options -> Connection -> Use a proxy server*. The proxy settings of your Web browser can be viewed for Internet Explorer 5 via *Tools -> Options -> Connections -> LAN settings -> Proxy server*, and for Netscape 4.5-4.7: *Edit -> Preferences -> Advanced -> Proxies -> View*. If you have problems with proxy settings, ask your provider or network administrator.

Dial protection (links to missing files)

When browser's mode of automatic dial-up connection to the Internet is enabled, any attempt to view a document saved on a disk and containing links to files (e.g. images) that have not been saved on disk, or clicking such links, can activate the automatic dialing. To suppress it, WebStripper has a special *Dial Protection* mode controlled through the menu *Options -> Options -> Browsing -> Dial Protection*. If this mode is enabled (options *When page loads* and *When following a link*), links to the files that have not yet been downloaded are rewritten in such a way that to avoid the automatic dialing – e.g. replaced with a text saying that such file has not yet been downloaded. When this mode is disabled (option *None*), links to the missing files are left as is.

If we need to change the *Dial Protection* mode for a site that has been already downloaded, we don't need to download it again. Instead, we should change this mode in the program settings, then select the required site and run *Utilities -> Re-Parse*.

Update

When a site that has been already downloaded is updated, we don't need to download all the files again. Instead, we can use the *Site -> Action -> Update* menu.

WebZip and Teleport Pro

Unlike WebStripper, trial versions of WebZip and Teleport Pro have restrictions on the term of use and/or their functional capabilities (for example, the trial version of Teleport Pro restricts the number of downloaded files). However, these programs have some additional features for “filtration” of downloaded files as compared to WebStripper. For example, WebZip has special “excluding” and “including” filters which help to download only those files whose URLs contain or do not contain specific text strings (menu *Download Method -> URL Filters* in task properties).