

WEBSTAR SERVER SUITE 4.2 ADDENDUM

UPGRADING FROM A PREVIOUS VERSION OF WEBSTAR

When upgrading from a previous version of WebSTAR it is recommended that you read this entire section as it includes important upgrade and change information. It is also highly recommended that you use one of the Installers Upgrade options rather than trying to upgrade to WebSTAR Server Suite 4 "by hand".

To upgrade from a previous version of WebSTAR, follow these steps:

- 1 Quit the WebSTAR server application in the folder you wish to upgrade.
- 2 If you have increased the servers memory allocation to support additional Plug-Ins or a larger number of connections you will need to assign more memory to the new WebSTAR 4.2 server application. Select the older WebSTAR application and choose Get Info from the File menu. Write down the difference between the Preferred Size and the Suggested Size.
- 3 Double-click the WebSTAR Server Suite 4.2 Installer application and select the appropriate upgrade option from the installers pop-up menu. If the folder you wish to upgrade contains a normal version of WebSTAR, select the Upgrade option. If the folder you wish to upgrade contains a copy of WebSTAR/SSL for U.S. and Canada (supporting RC4-40, DES-56 and RC4-128) select Upgrade with SSL - U.S. & Canada, users of WebSTAR/SSL for other countries (with support for RC4-40 only) will want to select Upgrade with SSL - International. Click Install.
- 4 The Installer will then present you with a dialog explaining that in order to upgrade WebSTAR Server Suite to support HTTPS (SSL) it is necessary to install or run this update off the WebSTAR CD or have an older copy of WebSTAR/SSL installed in the folder you wish to upgrade. If you have already installed a copy of WebSTAR Server Suite with HTTPS (SSL) support, click Continue to upgrade to the latest version with HTTPS (SSL) support.
- 5 Next the Installer will search your hard drives for a potential WebSTAR folders to upgrade. If necessary, the Installer will display a dialog prompting you to select the folder you wish to upgrade. Select the appropriate folder then click OK.
- 6 The Installer upgrades the selected folder by first copying and moving older files into a new "WebSTAR (old)" folder located at the root of the startup drive, then by installing copies of the latest WebSTAR Server Suite components into the selected folder. Third-party Plug-Ins, CGIs and Servlets, and most importantly, your sites contents are not disturbed during the upgrade process. Once completed, a dialog box will be displayed telling you that the upgrade was successful — click OK and select Quit to leave the Installer.

This chapter helps you make the transition to WebSTAR Server Suite 4.2. It also summarizes what's changed since 4.0 and 4.1.

If you're new to WebSTAR, start with the WebSTAR Technical Reference Guide. This is an addendum to that document.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

- 7 If necessary, select the new WebSTAR 4.2 server application and increase the memory allocation by choosing Get Info from the File menu. Increase the Preferred Size by the difference between Preferred Size and the Suggested Size in Step 2.
- 8 To automatically launch the WebSTAR server when you restart the machine, make an alias of the new WebSTAR 4.2 server application and move it to the Startup Items folder. Remove any aliases to older versions of WebSTAR.
- 9 Open your newly upgraded WebSTAR folder and double-click the WebSTAR 4.2 application. When launched the WebSTAR 4.2 server application is designed to update your server settings for use with WebSTAR Server Suite 4.2. While previous WebSTAR server application settings are maintained most WebSTAR/SSL settings (with the exception of the SSL certificate password stored in the WebSTAR/SSL Settings file) are not. Thus WebSTAR/SSL users will need to reenter Realms, Allow/Deny and other server settings manually.
- 10 You will next be prompted to enter your serial number. Enter your serial number and then click the OK button.
- 11 If not already entered, you will be prompted to enter an Admin Password. Enter a password then click OK.
- 12 With the WebSTAR server application running, double-click the WebSTAR Admin application.
- 13 Choose New Connection from the File menu and enter the primary host name or IP address of your WebSTAR server, the port number it is running on and the Admin Password.
- 14 Once connected verify your WebSTAR server settings, making any modifications that may be necessary.
- 15 Click Save.

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

This section summarizes the changes made to WebSTAR Server Suite since 4.1.

WebSTAR (Web) Server Application

- Corrected a problem where browsers could still connect with RC4-40 to the HTTPS/SSL server when all ciphers except 3DES and RC4-128 were turned off.
- Optimized for PowerPC G4 processors and accelerated for Velocity Engine.

WebSTAR FTP

- Optimized for PowerPC G4 processors.

WebSTAR Mail

- Fixed European character handling problems in To:, Subject:, Cc: and Body: fields of Web-based mail.
- Fixed Web-based email auto-recognized URLs to include & in the URLs.
- Fixed individual user "require APOP" flags to work.

WebSTAR Proxy

- Optimized for PowerPC G4 processors.

WebSTAR Admin (Plug-In)

- Optimized for PowerPC G4 processors.

WebSTAR Auto BinHex

- Optimized for PowerPC G4 processors.

WebSTAR Byte Server

- Optimized for PowerPC G4 processors.

WebSTAR Data Cache

- Optimized for PowerPC G4 processors and accelerated for Velocity Engine.

WebSTAR Directory Indexer

- Optimized for PowerPC G4 processors.

WebSTAR File Upload

- Optimized for PowerPC G4 processors.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

WebSTAR Image Map

- Optimized for PowerPC G4 processors.

WebSTAR Lasso Publisher

- Error.lasso - Can create custom Lasso error pages.
- Added new tags: Include_Raw, Encode_StrictURL.
- The following substitution tag abbreviated syntax have been added: [Error_ColumnRestriction] (for [Error_FieldRestriction]), [Error_RequiredFieldMissing] (for [Error_ReqFieldMissing]), [Req_Column] (for [Required_Column]), [Req_Field] (for [Required_Field])
- An issue with named token values not being URL encoded within the [Shown_NextGroup], [Shown_PrevGroup] and [Link_Detail] tags has been addressed.
- A problem which prevented Lasso from recognizing the abbreviated syntax of the -required (-req) and -requiredFieldMissingError (-reqFieldMissingError) tags has been addressed.
- The [Include] tag no longer returns error messages unless the Debug module is active.
- [Link_Detail] now supports the NoLogicalOperator (NoLop and NoLogicOp) and NoTimeout keywords.
- [KeyField_Value] nows returns [RecordID_Value] for FileMaker databases. -keyValue can be used to specify the record ID. The ODBC module is not required.
- The [Loop] tag now validates the values of the LoopFrom and LoopTo parameters.
- An issue with decimal number truncation with the ODBC Datasource Module has been addressed.
- An issue with Lasso not being able to display empty repetitions for repeating fields in FileMaker Pro databases has been addressed.

WebSTAR Log Archiver

- Fixed bug where archiving didn't work correctly if WebSTAR was installed at the root of the volume (no WebSTAR folder).
- Fixes bug with incorrect status report display of 12:00 PM (said 12:00 AM).

WebSTAR QuickDNS Load Balancer

- Optimized for PowerPC G4 processors.

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

WebSTAR Search

- Fixed a crashing which could occur while indexing. The crash would occur shortly after a redirect was encountered and handled.
- Fixes another crash while indexing, a side-effect of encountering a malformed URL.
- Fixes bug where denial of one URL by a 'robots.txt' file stopped indexing of all other URLs in the index.
- Fixes problem with filtering of JavaScript from summaries.
- Corrects a serious limitation in the customizable results code. Previously, each custom tag was only searched-for and replaced once. This version replaces all instances of a tag within that result "body part".
- Further enhances results customizability by affording access to the `###PREVIOUS_SEARCH_QUERY` attribute from the `<RESULTS_HEAD>` body part.
- Further enhances results customizability by affording access to the `###SEARCH_INDEX` attribute from the `<RESULTS_FOOT>` body part.
- Improves error-handling of multiple URL indexes. Previously, a failure to connect to a site (i.e. server busy errors) that prevented even a single file from being downloaded would cause the loss of all previous info indexed from that site. Now, if an index already exists for a set of URLs (i.e., it is not being created for the first time), and complete failure to connect to one of the sites is experienced, the index operation is aborted and the original index is preserved untouched.
- Image map `<AREA>` tag HREF URLs are now crawled correctly.
- Permits "minscore" (relevance) setting of 0 (means "apply no minimum relevance to matching results, causing all to be displayed").
- Sets "minscore" (relevance) default to 0 if unset or if found to be negative.
- If terms within bracketed search queries are only separated by whitespace, insert an OR boolean ("|") between.
- Error pages returned on 404 errors that contain links are not crawled.
- Double quotes in search strings are ignored.

WebSTAR SSI

- Addressed a problem introduced in WebSTAR Server Suite 4.0 where no data was returned from a HEAD request.
- Optimized for PowerPC G4 processors.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

WebSTAR SSI-WebInclude

- Optimized for PowerPC G4 processors.

WebSTAR Virtual Hosts

- Optimized for PowerPC G4 processors.

WebSTAR WebObjects Adaptors

- Optimized for PowerPC G4 processors.

WebSTAR Developer Toolkit

- The WebSTAR Developer Toolkit now includes a Template for creating Carbon API-based Plug-Ins and/or Plug-Ins accelerated for the PowerPC G4 processor with Velocity Engine.

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.1

This section summarizes the changes made to WebSTAR Server Suite since 4.0.

Changes to the WebSTAR (Web) Server Application

- Corrected memory reporting inaccuracies with the memory "thermometer" bar.
- Fixed a problem with incorrect port numbers being passed to CGIs from HTTPS connections.
- Now handles full URL requests in the form of:

```
GET http://www.some.dom:80/somefile.html HTTP/1.1
```
- Removed command keys for enabling/disabling of Verbose Messages and Suspend Logging.
- Updated to OpenSSL 0.9.4 for improved SSL compatibility and to correct connection related problems.
- Now appends a !! to the "Bad URL reference" message.
- Fixed errors where the server would get confused when it received a HTTP/1.0 request with a "Connection: close" field in the header and turn persistence on. It would also get confused if it received an HTTP/1.0 request with a "Keep-Alive" field in the header, and turn persistence off.
- Will now request a (SSL) client certificate and optionally pass it to a Plug-In, when configured and requested to do so.
- Added new parameters to StarNine.h for handling of (SSL) client certificates and requests.
- Fixed bug with referer keyword sent to CGIs.
- Modified the server to allow the use of NULL parameter blocks during WSAPI_xStream() calls.

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

- For increased security WebSTAR now supports private keys of various lengths, from 256 (minimum) to 2048 (maximum).
- For supporting private keys of various length, WebSTAR now uses a new key format. Users who generate Private Key files with the new 4.1 version of the CSR Utility will not be able to downgrade to WebSTAR 4.0 or earlier, however WebSTAR 4.1 will continue to support older Private Key files without modification.
- Correctly resets the persistent connection timeout value when connections are handled by Plug-Ins. This fix corrects the persistent connection "broken/partial images" bug.
- The WebSTAR.Messages log will now be appended to instead of being overwritten at each launch.
- Data is now written to the WebSTAR.Messages file when WebSTAR quits normally, receives a Quit AppleEvent, shuts down due to a memory allocation error, or experiences a severe Open Transport problem.

Changes to WebSTAR FTP Server

- Improved disconnection messages to be more descriptive.
- Fixed problems which would cause the server to report memory allocation problems.
- Fixed a bug that could corrupt the system heap on server startup.
- In response to a SIZE command, now computes and returns the correct MacBinary size.
- When reporting UNIX as the server type, now returns a more UNIX-like file listing. This should correct display listing problems with various types of FTP clients.

Changes to WebSTAR Mail Server

- Resolved mail delivery issues when sending SMTP mail messages to servers that require the Message-Id to have <text@domain> syntax.
- Changed message date to come from the message Date: field, instead of the date/time actually received.
- Corrected periodic attachment corruption problems.
- Fixed intermittent halt due to bad pointers from the OT callback to our UDP drivers.
- Now resets the MX lookup cache and UDP session every 15 minutes to address mail delivery issues.
- Fixed intermittent halts when performing DNS look-ups.
- Changed default disk cache (for new installations) to 10240K

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

- Fixed Web-based email interface to show the message date when reading mail. The Web-based email interface now shows the date from the message header.
- Changed Web-based email listing page to show the date created, but to sort by date received, with most recently arrived items at the bottom of the list.
- Added explicit connection close of WebSTAR API streams used to accept a listen. This should address the majority of connection problems seen by most sites.
- Fixed bug in incoming dates with a GMT timezone.
- Fixed crashing bug with memory reclamation while adding entries to the DNS and MX caches.
- Fixed POP3 to log APOP-style logins..
- Added ETRN to EHLO extensions list, and respond correctly to ETRN command.
- Added a log entry when forwarding expansion results in an attempt to deliver to an unknown local user:
"Warning--forwarding to an unknown local user". The original RCPT and the missing names are also written to the log.
- Fixed "Cc" field in Web-interface to be labeled correctly.
- Fixed "Cc" field in Web-interface to convert European characters correctly.
- Fixed problem where not all local recipients would receive a copy of the mail.
- Now sends ETRN commands to the SMTP server connected to (one ETRN command for each Host domain), when the routing for the connection is marked as "ETRN".
- Added logging of the client IP address for all Incoming SMTP mail.
- Changed Web-based list of messages to just show the sender's name.
- Added check-boxes to Web-based list of messages to allow multiple items to be deleted with one command.

Changes to WebSTAR Admin Application

- Fixed a problem with the ETRN setting not getting saved.
- Added support for entering of period (.) and other standard email characters in the Email field of the Mail Settings -> Users & Accounts screen.
- Fixed problem which prevented the Alternate DNS field from being left blank.
- Corrected problem which prevented the HTTPS port number from being changed..

WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

- Added screens for configuration of WebSTAR Auto BinHex, WebSTAR Directory Indexer, WebSTAR File Upload, WebSTAR Form Mail, WebSTAR Search and NetCraft registration settings.
- Changed the name of the Account ID field to Friendly Name. This field is **not** used as the mail client login name.
- WebSTAR Admin now requires ControlLib (Apperance Manager 1.0.3 or later) before it would run on Mac OS 8.0 or 8.1 systems.
- Added user database "Backup Now" option to Mail Settings -> Defaults screen.
- The 'BYTES_SENT' token is now displayed in the Log Format List after a default install.

Changes to WebSTAR Admin Plug-In

- Includes more secure password handling.
- Configuring (turning on/off) of HTTPS/SSL interfaces will now cause them to actually be turned off, as appropriate.

Changes to WebSTAR Auto BinHex Plug-In

- Added support for configuration of settings via the WebSTAR Admin application.

Changes to WebSTAR Data Cache Plug-In

- Now responds properly to POST requests.
- Now properly handles the NoAccess role, returning the currently defined No Access file.
- Before serving a file that is not cacheable, WebSTAR Data Cache gets the latest file information. This fixes problems that occur as a result of files being replaced on disk. Symptoms included retrieving the incorrect file, broken sockets, and truncated files. Flushing the cache would typically fix the problem.

Changes to WebSTAR Debug Plug-In

- Added Plug-In name to each entry in the WebSTAR API callback log (when available).
- Reduced number of WebSTAR API callbacks in the log to 100.
- Correctly identify valid addresses when virtual memory is turned on.

Changes to WebSTAR Directory Indexer Plug-In

- Added support for configuration of settings via the WebSTAR Admin application.

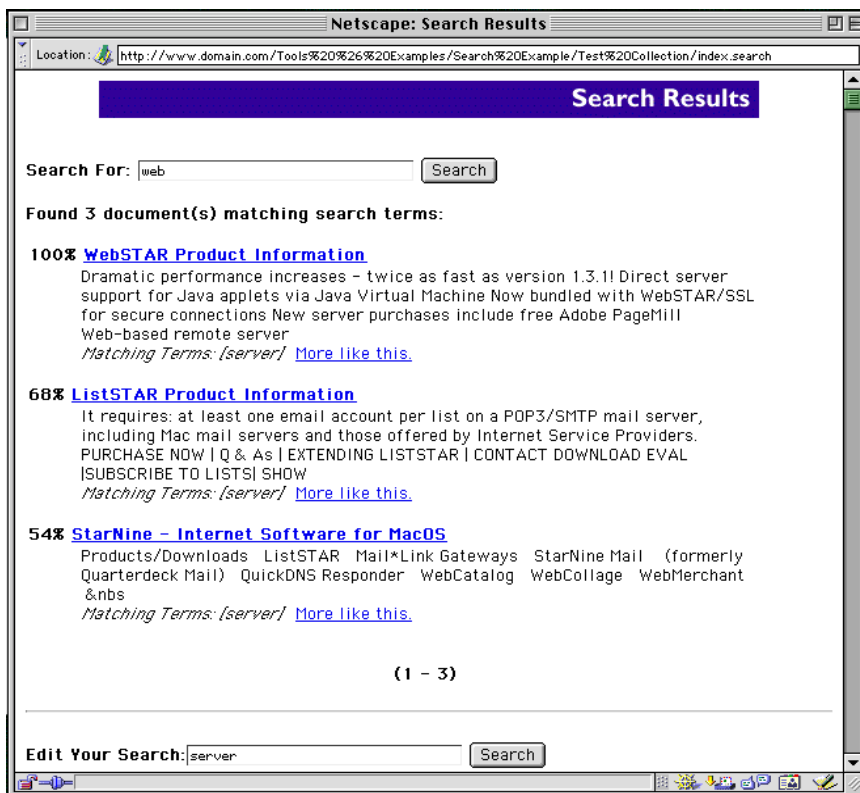
CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

Changes to WebSTAR File Upload Plug-In

- Changed return value behavior: If the file specified in the form ACTION tag exists, that file is returned to the user as text/html. (RAW! files are returned without any header.) If the form ACTION file doesn't exist, the client is redirected to the current folder with a full URL (eg: <http://www.my.dom:80/somefolder/>).
- Fixed a problem with extra characters being added to the uploaded files name.
- Added support for configuration of settings via the WebSTAR Admin application.
- Now supports uploads into the WebSTAR root folder.

Changes to WebSTAR Form Mail Plug-In

- Added support for configuration of settings via the WebSTAR Admin application.



WHAT'S NEW IN WEBSTAR SERVER SUITE 4.2

Changes to WebSTAR Log Archiver Plug-In

- After archiving, now creates a new, empty log file.
- Now checks for and prevents the archiving of zero-length log files.
- After log archiving options are changed in admin, a new "LOG ARCHIVER STATUS REPORT" will appear in the log in a short while (less than 1 or 2 minutes), showing the new configuration.

Changes to WebSTAR QuickDNS Load Balancer Plug-In

- Corrects an Open Transport "deafness" issue which can result during high periods of activity.

Changes to WebSTAR Search Plug-In

- Now supports customizable search results pages, and web crawler capabilities.

Changes to WebSTAR Virtual Hosts Plug-In

- Increased performance of host-matching process.
- No longer responsible for configuring (turning on/off) of HTTPS/SSL interfaces. This capability has been moved to the WebSTAR Admin Plug-in to better support the use of HTTPS/SSL in conjunction with third-party Router Plug-Ins.
- If the Directory Indexer plug-in is present, look for a '.message' file before searching for a valid index file. If the '.message' file is present, don't change the request.
- When WebSTAR Data Cache is not installed, WebSTAR Virtual Hosts does not attempt to resolve multiple index file names, although it will still allow the multiple index file list to be manipulated.

Changes to WebSTAR WebObjects Adaptor Plug-Ins

- Added 'x-client-ip', 'x-webobjects-remote-addr' and 'https' headers to WebObjects requests.

Changes to CSR Utility Application

- Key generation is now supported by the CSR Utility.
- Now supports the generation of keys with various lengths.
- Certificate Extraction, which was no longer needed with WebSTAR 4 servers, has been removed from the CSR Utility.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

WEBSTAR SEARCH

WebSTAR Search is a site search engine, allowing visitors to search the content of local and remote web sites. WebSTAR Search can be used to index local content, or as a web crawler (sometimes called robot or spider) to download and index web sites. WebSTAR Search supports the download and indexing of text, web pages (HTML), PDF, and supports customizable search interfaces, Boolean search operators, relevance ranking, results formatting and more..

Searching Mechanics

Before you can search, you must create a search **index**: a file that stores data about the files being queried. Search index files can be created with the WebSTAR Search Indexer application or by using the WebSTAR Admin applications Search screen to specify a web site to crawl.

The web crawler capabilities of the WebSTAR Admin applications Search screen are best use to index static and dynamic content from local, remote and multiple web sites.

The WebSTAR Search Indexer application is best used to index static content, stored on the local hard drive, that is not referenced via a web page.

Once you have an index, you can perform a query by using a **search form**: an HTML form with an Action link to the index file. When a user enters search terms into the form and clicks the "Submit" button, the browser sends a form action request to the server. Because the form URL suffix is `.search`, the request is passed to the WebSTAR Search Plug-In, which searches the index for the requested terms, and returns the results in a formatted page.

Installation

When you choose WebSTAR Search from the Custom Install options, following items are installed in the WebSTAR Plug-Ins folder:

- *WebSTAR Search (Plug-In)*
- *WebSTAR Search Indexer (application)*
- *WebSTAR Search Data (folder) containing the files described in "Search Dictionaries" on page 269 of the WebSTAR Server Suite 4 manual.*



WebSTAR does **not** install the WebSTAR Search Plug-In by default. Use the WebSTAR Server Suite Installer to custom install it. It does not require any additional RAM.

The WebSTAR Search Plug-In is always enabled, and will always be called by URLs ending in `.search`.

WebSTAR Search automatically registers the `.search` suffix. All index files created by the WebSTAR Search Plug-In or the WebSTAR Search Indexer application should always be named with a suffix of `.search`.

WebSTAR Search also requires that the WebSTAR SSI Plug-In be installed.

WebSTAR Search Memory Issues

WebSTAR Search can be very memory intensive, especially when indexing large amounts of information or searching large index files. To ensure responsive search queries and efficient content download and indexing, it is good to keep the following memory related issues in mind:

- When searching large index files, it may be necessary to allocate additional RAM to the WebSTAR server application. To track memory use, check the Status window memory fields (described in "Memory" on page 79 of the WebSTAR Server Suite 4 manual). Most of the memory required to search an index will be retained by the WebSTAR Search Plug-In following the first search query - it will not be freed once the search is completed.
- To index thousands of individual files or large PDF files, it may be necessary to increase the RAM allocation to the WebSTAR server (when using the WebSTAR Search Plug-in to crawl and index web sites) or the WebSTAR Search Indexer application (when using the WebSTAR Search Indexer application to index local content).
- While crawling web sites, WebSTAR Search temporarily downloads their content to a temporary folder within the WebSTAR Search Data folder. Make sure that enough hard drive space is available for these temporary files.

Search Examples

The Search Examples folder in the Tools & Examples folder includes a number of search examples. For more information on these files please see the "Customizing Search Results" text file.

Some setup is required to use the included Search Examples. To setup the Search Examples, follow these instructions:

- 1 Launch the WebSTAR Admin application and connect to the WebSTAR server. From the Edit menu select Server Settings. Select Web Settings -> Search.
- 2 Click the "New Index" button and specify a Search index path of:
`:Tools & Examples:Search Example:index.search`
- 3 Next, enter in the URL field the URL you would like to download and index. For example:

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

`http://www.starnine.com/`

For testing purposes it maybe wise to limit the URL's "Levels Deep" option to 1 or 2.

- 4 Click Save.
- 5 Select the Search Index and click the "Update Now" button. This will cause WebSTAR Search to immediately begin downloading and indexing the content for the defined URL. Once completed a file named "index.search" will be created in the Search Example folder found within the Tools & Examples folder.

It may take several minutes to download the content of the specified site. Please wait until the index file has been created before continuing on to the next step.

- 6 Using a browser, enter your host name and the path:

`/Tools%20%26%20Examples/Search%20Example/search.html`

When you do a search, you'll see a result list that looks like this:

Indexing and Index Files

Search index files can be created, updated and maintained via one of two methods: the WebSTAR Admin applications Search screen (recommended for most uses) or the WebSTAR Search Indexer application. Search indexes defined in the WebSTAR Admin application are *web crawler indexes*, while those created and defined in the WebSTAR Search Indexer application are *local file indexes*.

Web Crawling

As a web crawler (sometimes called robot or spider) WebSTAR Search acts as an HTTP-client, locating files by following URL links. In the WebSTAR Admin applications Search screen you specify which URLs you would like to assign to a given index. WebSTAR Search then uses these URLs as a starting point from which to follow links.

To setup a web crawler index, following these steps:

- 1 Open the WebSTAR Admin applications Search screen.
- 2 Click the "New Index" button to create a new search index listing.
- 3 In the "Search Index" field, enter the Mac OS file path to the search index file. If no search index file exists at this location, WebSTAR Search will auto-create one once it has finished downloading and indexing the specified content.

File paths can be absolute or relative, and should include a ":" as a delimiter. For example:

`Macintosh HD:WebSTAR:mysite:index.search`

:mysite:index.search

- 4 Click on the URL assigned to the search index and enter the URL(s) to crawl. WebSTAR Search uses this URL as the starting point which to follow links from. WebSTAR Search will not crawl URLs referenced outside this site, neither will it crawl outside assigned URLs containing subdirectories. URLs which require username/passwords to be accessed can be entered in the form of:

`http://username:password@website.com/`

- 5 In the Levels Deep pop-up, specify the number of URL directory levels deep for it to crawl. It is a good idea to limit the number of levels deep to crawl, especially for large sites that may take many minutes if not hours to download and index.
- 6 To enable WebSTAR Search to auto-update the specified index, check the "Update search index" checkbox and define the number of days and the time at which to perform search index updates.
- 7 Click Save to record changes.
- 8 Click the "Update Now" button to have WebSTAR Search begin crawling the specified URLs and update or create the specified index file. Once the web crawling process is started a dialog box will be displayed in the WebSTAR Admin application stating Search Index update started successfully."

Following Links and Downloading Files

As a web crawler, WebSTAR Search will follow and download any link that includes a correctly formatted URL, usually in the form of

```
<A HREF=" thepage.html ">
```

or URL which is specified in a frame set, client-side image map or META refresh tag.

To follow and download a given link, the URL must be an index file or end in a suffix of: .HTM, HTML, .SHTML, .SHT, .ASP, .PDF, .SSI, .JSP, .TXT, .TEXT, .LASSO, .LAS, .TMPL, .TAF, .CGI, .PL, .CFM, .XML or .SGML.

NOTE: While WebSTAR Search 4.2 supports the downloading of Adobe Acrobat PDF files, it does not support the indexing and summarization of their content.

Temporary Files

While crawling a site, WebSTAR Search downloads and stores the site's content within a folder called "WebSTAR Search Temp*f*" within the WebSTAR Search Data folder. The contents of this folder are deleted once the whole site has been crawled and the content has been indexed.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

Robot Exclusion Standard

WebSTAR Search fully supports the robot exclusion standard as defined at:

```
http://info.webcrawler.com/mak/projects/robots/norobots.html
```

Before downloading the content of a web site, WebSTAR Search first looks to see if a file named "robots.txt" is present in the root directory of the specified site. If it is, WebSTAR Search downloads the file, looks for a user-agent of "WebSTAR" or "*" and adheres to the Disallow statements assigned to that user-agent.

HTTP User-Agents

Depending on the type of method being used to crawl a web site, WebSTAR Search uses the following HTTP user-agent names:

```
WebSTAR Search/4.2 (http://www.starnine.com/webstar/; URLBypass)
WebSTAR Search/4.2 (http://www.starnine.com/webstar/; URLAccess)
```

URLBypass is an HTTP client built into the WebSTAR Search Plug-In which, when used, results in much faster downloading than URLAccess. URLAccess is an Internet client built into Mac OS 8.6 and later that is slower but offers more features than URLBypass.


WebSTAR Search first uses URLBypass to access and download a given link. If it encounters an error, WebSTAR Search retries the request using URLAccess.


Local File Indexing

To create a local file index, decide which folder will be the search root folder. The WebSTAR Search Indexer application will open and index all text and PDF files in this folder and its subfolders.

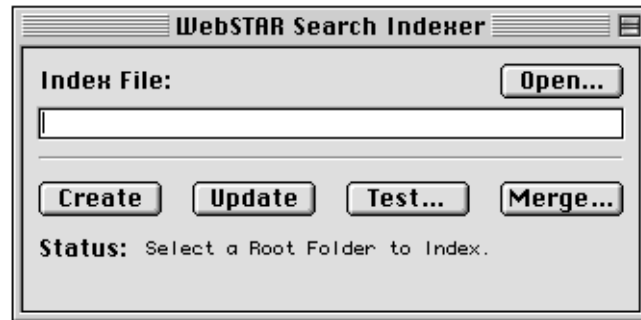
For security, you should not make your WebSTAR root folder the search root. You should also index your secure realms separately from the public data. Limit your indexes to data folders, and merge them later if you want cross-folder searching.

To create a local file index, following these steps:

 If you expect to use this application often, make an alias of it and put it in a convenient location.

 See also “Web File and Folder Name Rules” on page 189.

- 1 Open the *Plug-Ins* folder and launch the Search Indexer application.



- 2 Click the **Create** button and set the name and folder for your search index. Note that all index names must end in the suffix `.search`, but you can customize the rest of the name. You can also choose **New** from the File menu to make a new index.
- 3 Once you have named the file, click **Save**.
- 4 Next, select the **search index root folder** for this index. Only files within this folder or a sub folder will be included in the index. In most cases, this will be a subfolder of the WebSTAR folder, especially if you're serving Virtual Hosts.

Be sure that there are no encrypted PDF documents in this folder hierarchy
- 5 The Search Indexer will open, analyze and index each text and PDF file within the root folder. It will not follow aliases to other files or folders.

This may take several minutes, depending on the amount of data and the speed of the machine. It can operate, more slowly, in the background if you need to use the machine for other purposes.

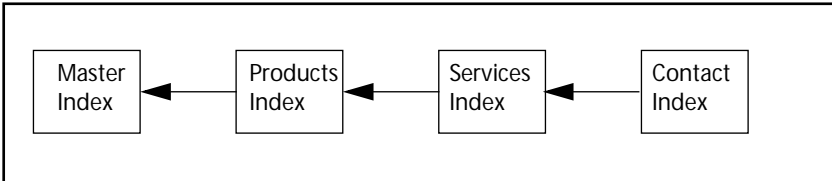
If the indexer runs out of memory, just quit the indexer, increase its memory partition, launch it again, and **update** the index that was being built.
- 6 Once done, you can choose the **Test** button to try out your index. Enter a query in the Test field and you'll see the same result as if you entered the data in a search form.

Once you have created your index, you can design search forms and links for your visitors: see “Web Tools For Searching” on page 27 of the WebSTAR Server Suite 4 manual.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

Merging Local File Indexes

You can also use the Search Indexer application to merge index files together. You can use this to index only some subfolders of a site and search them as one:



To merge indexes:

- 1 Launch the Search Indexer
- 2 Choose **Open** to select and open the first index.
- 3 Click **Merge**, open the next index file, and wait while the application merges the second index file into the first.

Note that each index can only have **one** other index merged into it. Therefore, you must design an index chain of parent and child indexes, as shown above.

Updating Local File Indexes

To keep your indexes synchronized with your files, you must update them periodically. Updating is much faster than indexing, because it just checks the root folder and subfolders, and adds new files, removes deleted file data, and re-indexes the data in files that have changed. Updating can run in the background without slowing down your server.

Otherwise, you'll have to update the indexes using the WebSTAR Search Indexer application and an AppleScript.

This example shows that you must include the exact path to the index file (you can copy it from the Indexer and paste it into your script):

```
tell application "WebSTAR Search Indexer"
    update "server HD:WebSTAR Server Suite:search:index.search"
end tell
```

Search Index Security


Files created by WebSTAR, WebSTAR Admin, and WebSTAR Search Indexer (with Creator Codes of "WWW", "WWWx", and "WSlx") will not be indexed.

To avoid having a file indexed, move it out of the search root folder hierarchy, rearrange your folder hierarchy, or use a file resource editing application to change its Creator Code to "noIX", which will indicate to the WebSTAR Search Indexer that this file should be skipped.

If you have security realms or confidential information in one of your indexes, name it carefully, and consider protecting it by using a security realm entry. That way, no one can search through it to find private data.

Search Dictionaries

The Dictionaries installed in the *WebSTAR Search Data* folder, *Stopword Dictionary* and *Substitution Dictionary* provide lists used by the search engine in comparing queries with documents.

 Dictionaries are optional, although very useful.

The Stopword Dictionary is just a list of words (one per line) that are never indexed or searched. These include single letters, HTML tags, and extremely common words such as "allow", "is", and "somewhere". You can change this list and re-create your index to improve responsiveness and reduce inappropriate results.

The Substitution Dictionary helps the search engine match words and their various formats. For example, the past tense of "go" is "went", but this would never be found in a direct comparison. This dictionary allows the index and search tools to make that kind of match.

The format of this dictionary is a list of words, one per line, where a substitution consists of a bullet, followed by a root word, followed by any number of words that should be reduced to that root.

•fall fell fallen

You can edit these files to change or remove entries, add more entries, and include text for languages other than English. Be sure to synchronize the index and the searching by re-creating your indexes if you change the dictionaries.

About Search Queries

With WebSTAR Search, you can enter the words in your *query* in any order. The search engine will compare them to the data in the index: it

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

will find all files with at least one of the search terms. When it returns the results, WebSTAR Search will automatically rank them from most relevant to least relevant.

WebSTAR Search performs a **vector search**, where every indexed document is stored in the index as a point in a multi-dimensional space (a **vector**), and the dimensions correspond to terms found in the body of all documents. Normal queries are then interpreted as very short documents, and are also converted into a vector. The search engine then looks at the index for all document vectors within a certain distance of the query vector. Thus, the more that your query resembles the documents you are looking for, the better your results will be. That's why natural language queries, such as questions or descriptions, work so well with WebSTAR Search. For example, you could enter the query:

```
improve performance of web server
```

and get back results ranked according to how closely they resemble that query.

Adding Boolean Operators to Searches

In addition to the vector searching, you can add Boolean operators to your searches, if you prefer.

- **&** (ampersand) used between search terms is the Boolean AND operator. It indicates that both terms must be in a document. This is very useful for getting a more relevant set of results. For example, to search for documents that mention both "Sherlock" and "Apple", enter `sherlock & apple`
- **|** (vertical bar or pipe character) used between search terms is the Boolean OR operator. It indicates that either term can be in a document, for example, to search for documents that either "AIAT" or "Sherlock", enter `aiat | sherlock`
- **!** (exclamation point) used between search terms is the Boolean NOT operator. It indicates that only documents matching the terms before the it and not matching the term after it should be matched. Again, this removes irrelevant results. For example, to search for documents that mention "Sherlock" but not "Holmes", enter `sherlock ! holmes`

Grouping

Another way to control the query is to explicitly **group** the terms—control the order of processing

For example, you may want to link the terms "Sherlock" and "Apple" and ignore "Holmes". So you can use the grouping brackets "[" and "]" to specify how the search should be grouped:



The search engine uses special functions to improve word matching: see "Search Dictionaries" on page 27.

```
aiat | [[sherlock & apple] ! holmes]
```

Search Forms

To search an index, you must have an interface. The simplest interface is an HTML form. The easiest way to start making such a form is to copy the simple search form file from the *Tools & Examples: Search Examples* folder. For testing, just change the form action so that it refers to your index, for example:

```
ACTION="/widgets/widgets.search"
```

Once you have a form, you can open that page on your server and search your index.

The results of the search will be returned in the pre-defined page format. As you can see with the *advanced.html* form, WebSTAR Search will display 10 results per page, with page forward and back links, by default.

Search Form Error

If a visitor sees the message "The search parameters were not received, probably because you pressed Return instead of clicking on the submit button. Go back, and try clicking the button.", it may also mean that there is no such index. To allow visitors to press the "Return" key in your form, move the "Submit" button out of the table. That way, all the browsers will do the right thing.

Search Links

You can set up URLs and links in pages that passes the query parameters in the search arguments section of the URL. For example, this URL:

```
http://www.domain.com/Tools%20%26%20Examples/Search%20Example/Test%20Collection/index.search?query=webstar
```

will search the example index for the word "webstar", assuming you have created the index and you changed the host name from "www.domain.com" to your host name.

Likewise, the following link, in a file in the *Search Examples* folder, will search the Test Collections index for the query "web sever":

```
<A HREF="http://Test%20Collection/index.search?query=web&20server">
```

Each link can also include parameters, as described below.



Note that the space between "web" and "servers" is encoded (%20). You will have to make sure that all characters are encoded as part of the URL.


CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

Search Tags Using WebSTAR SSI

WebSTAR Search works with WebSTAR SSI to give you more control over the search parameters and the location of the search result. You can set the query in an SSI variable and send that to WebSTAR Search, rather than having the user enter it.

You customize the search results page returned to the browser by creating your own SSI page that defines the desired format. Within this SSI page, a `<SEARCH. . .>` tag should be inserted at the point where the search results should be inserted. Note that the results are still formatted in the pre-defined layout.

For an example, see the *custom.html* and *custom.ssi* files in the Search Examples folder.

 To be sure that your file will be processed by WebSTAR SSI, save it with the `.ssi` suffix.

Search Parameters

The user may specify the several search parameters to the WebSTAR Search Plug-In. These parameters may be specified either via `<INPUT>` statements in an HTML form, as search arguments to a URL, or as tag parameters to the `<SEARCH>` tag handler.

Search Parameters List

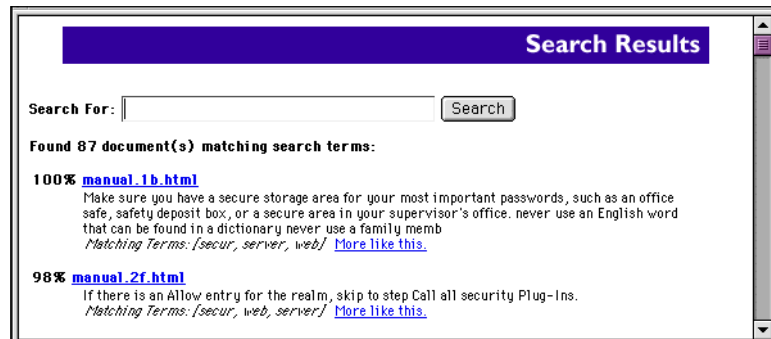
PARAMETER	STATUS	USE
query	required	The question, description, or list of keywords that are to be matched against the documents in the index.
numdocs	optional default is 10	The maximum number of matching documents to return.
minscore	optional default is 50%	The minimum match score for documents that are to be displayed to the client. Scores range from 100% down to 0%, with 100% representing the best match.
To specify pagination of results		
firsthit	optional default is 1	Out of <code><numdocs></code> matching documents found by the plug-in, this parameter defines the first that is to be displayed to the client.

Search Parameters List

PARAMETER	STATUS	USE
lasthit	optional default is 10	Out of <numdocs> matching documents found by the plug-in, this parameter defines the last that is to be displayed to the client.
For the <SEARCH> Tag only		
index	required	The path name to the index file to be used in the search. This path may be either a virtual path (e.g., /folder/index.search) or an absolute Mac pathname (e.g., Disk:WebSTAR Folder/folder/index.search).

Search Results

WebSTAR search will automatically display formatted results of your search:



Note that the links for the results will take you to the beginning of the document containing the matched text. For an exact location, you must use the Find function of your browser.

Search Results Elements

- **Number Found:** displays the number of results which match. Note that if you do not specify a maximum number of results, WebSTAR Search will return no more than 10 matches.
- **Relevance Ranking:** the percentage number in the left column (100%, 58%, etc.) shows the how relevant the document is to your query. This number is calculated by the search engine.
- **File Link** the name of the matching document, with a link to the file itself. Acrobat files include the document title if there is one.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

- **Summary:** For PDF documents which include Subject and Keyword fields, they are displayed here. For other files, the search engine computes the most important words and phrases from the document and displays them.
- **Matching Terms:** the query words which were matched in the document. If the word was converted by the Substitution Dictionary (for example "fallen" changed into "fall"), you'll see the converted word here.
- **More like this:** WebSTAR Search and the search engine allow you to refine your search by using the document itself as a query string, to find similar documents. This works best for large collections of single-concept files.



See "Search Dictionaries" on page 27.

Customizing Results

The default settings are to return search results based upon an internal results format template. The default settings are also set to find and return only the 10 most relevant documents which have a ranking of 50% or greater.

In the included Search Example files, you'll see that the forms include options to specify the relevance, along with the total number of results. In this case, WebSTAR Search will, by default, display 10 results per page and provide navigation links to go forward and backward, if there are more than 10 matches.

You can use the Search Parameters in your forms and URLs to specify the relevance ranking, and how many results to display per page.

To customize the format of the search results, you can use a search template. (An example search template can be found in the Search Example folder.) A search template is divided into six sections, some of which may include variables (described below):

- `<RESULTS_PREVIOUS>` and `</RESULTS_PREVIOUS>`
Used for defining of text to use in "Previous Results" link.
- `<RESULTS_NEXT>` and `</RESULTS_NEXT>`
Used for defining of text to use in "Next Results" link.
- `<RESULTS_HEAD>` and `</RESULTS_HEAD>`
Used for defining header and form results to return. Supports the following variables (described below): `###SEARCH_INDEX`, `###NUMBER_DOCS`, `###MINIMUM_SCORE`, `###RESULTS_HIT`

- **<RESULTS_HIT>** and **</RESULTS_HIT>**

Used for defining the format of the actual query results. Supports the following variables (described below): ###RELEVANCE, ###URL, ###TITLE, ###SUMMARY, ###MATCHING_TERMS, ###MORE_LIKE_THIS

- **<RESULTS_FOOT>** and **</RESULTS_FOOT>**

Used for defining footer to return at bottom of results. Supports the following variables (described below): ###RESULTS_PREVIOUS, ###RESULTS_CURRENT, ###RESULTS_NEXT, ###PREVIOUS_SEARCH_QUERY

- **<RESULTS_FAILURE>** and **</RESULTS_FAILURE>**

Used for defining the page to return when no matching items were found. Supports the following variables (described below): ###SEARCH_INDEX, ###NUMBER_DOCS, ###MINIMUM_SCORE, ###PREVIOUS_SEARCH_QUERY

Search Variables

To display results specific to a particular search query, you can insert variables into the search template. Search variables include:

- **###SEARCH_INDEX**

Defines the path name to the index file to be used in the search, as for use in the Form Action.

- **###NUMBER_DOCS**

Defines the maximum number of matching documents to return.

- **###MINIMUM_SCORE**

Defines the minimum match score for documents that are to be returned in the results.

- **###RESULTS_HIT**

Defines the number of documents returned for the results page.

- **###RELEVANCE**

Defines the relevance ranking, showing how relevant the document is to the query.

- **###URL**

Defines the URL of the found document.

- **###TITLE**

Defines the title of the found document.

- **###SUMMARY**

Defines the summary of the found document.

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM

■ ###MATCHING_TERMS

Defines the matching terms of the found document..

■ ###MORE_LIKE_THIS

Defines a link for refining the search by using the document itself as a query string.

■ ###RESULTS_PREVIOUS

Defines a link to the previous results displayed. This variable is may not always be displayed.

■ ###RESULTS_CURRENT

Defines the number range of results currently displayed.

■ ###RESULTS_NEXT

Defines a link to display the next results. This variable is may not always be displayed.

■ ###PREVIOUS_SEARCH_QUERY

Defines the terms used in the previous search query.

Search Troubleshooting

Search Form Error

If a visitor sees the message "The search parameters were not received, probably because you pressed Return instead of clicking on the submit button. Go back, and try clicking the button.", it may also mean that there is no such index. To allow visitors to press the "Return" key in your form, move the "Submit" button out of the table. That way, all the browsers will do the right thing.

WEBSTAR SERVER SUITE 4 TECHNICAL REFERENCE ERRATA

Page 201

- Some of the log tokens include the host name of the client or the server. These will only appear if you have selected the Use DNS for Server and Client Lookups option (described on page 136).

Page 384

- Limit cached DNS entries to: limits DNS caching to certain number of records. The range is from 1-999 DNS entries, and there is no default.

Page 395

WEBSTAR SERVER SUITE 4 TECHNICAL REFERENCE ERRATA

- WebSTAR Mail has a simple tagged format for importing and exporting mail user settings and messages. It uses tags and attributes, which are based upon SGML and XML.

Page 401

- Should not include:[tbd:need new picture]

Page 402

- Send Time - time the server will attempt to send the message

Page 455

- The correct URL to the WebSTAR API SDK is:

`ftp://ftp.starnine.com/pub/webstar_dev/wsapisdk.sea.hqx`

TROUBLESHOOTING

Problem: One or both users cannot log in if two User Accounts have similar email names *and* identical passwords.

Solution: The problem is that POP clients do not usually pass the hostname when they login, so WebSTAR Mail has to use the password to differentiate between accounts. If you need to use similar email accounts on the server (e.g. luke@totalystupid.com, luke@notsostupid.com) with the same passwords, then set up the POP user names in the client(s) to include the hostnames, using a percent sign instead of the at sign (e.g. luke%totalystupid.com, luke%notsostupid.com).

CHAPTER : WEBSTAR SERVER SUITE 4.2 ADDENDUM