

dtSearch Web

dtSearch Publish

Version 7

Manual

Copyright 1991-2011 dtSearch Corp. www.dtsearch.com

SALES

1-800-483-4637

(301) 263-0731

Fax (301) 263-0781

sales@dtsearch.com

TECHNICAL

(301) 263-0731

tech@dtsearch.com

UK DISTRIBUTOR

www.dtsearch.co.uk

MORE WORLDWIDE DISTRIBUTORS

See www.dtsearch.com

Table of Contents

1. dtSearch Web Quick Start	1
2. Customizing the Search Interface	4
Customizing the Search Form.....	4
Selecting Indexes to Include	6
Customizing the Search Results Format.....	7
Modifying a Search Results Item	9
Document Display Options.....	11
3. Technical Information	13
Generated Files.....	13
The Search Form	15
The Options File	17
Virtual Directories	22
Security	23
Windows Server Installation	23
4. Portable Media Publishing	25
Publishing (Overview)	25
Using the CD Wizard.....	25
CD Types	27
Standard CDs.....	28
Local HTTP Server CDs.....	32
Software Dependencies	34
cdrun.xml.....	35
5. Index	37

dtSearch Web Quick Start

dtSearch Web is a search engine that you can install on a web server to publish documents on your web site or to search remote web-based content. It can perform fast indexed searches using the same search features that dtSearch Desktop/Network supports -- fuzzy searching, phonic searching, natural language searching, boolean logic, proximity, etc. Indexed documents can be in any format that dtSearch supports.

After a search, dtSearch Web will convert documents as needed to HTML for display, with hits highlighted. If a retrieved document is already in HTML, dtSearch Web will display it with hits highlighted while preserving the HTML attributes (so links and images will work in your browser). PDF files will appear in Adobe Reader with hits highlighted.

dtSearch Web requires Microsoft Internet Information Server (IIS).

This Quick Start will describe how to get a basic search form running on your web site quickly. You can create any number of search forms, each with its own set of option settings and index selections. For information on setting up dtSearch Web to run from a CD, DVD, or other portable media, see CD Publishing (Overview).

1. Install the dtSearch and dtSearch Web program files on your web server.

If you have dtSearch Web on CD, run the setup program to install. If you downloaded dtSearch Web from the internet, follow the download instructions to open the dtSearch Web archive and install the files.

2. Build an index of your documents

dtSearch Web uses the same indexes as dtSearch Desktop/Network, so if you already have a dtSearch index of the documents, you can use this index with dtSearch Web. If not, follow the dtSearch Desktop/Network Quick Start to set up an index.

If your site includes dynamically-generated content rather than static content such as HTML and PDF files, or if you are searching content on other web sites, you can use the dtSearch Spider to index it.

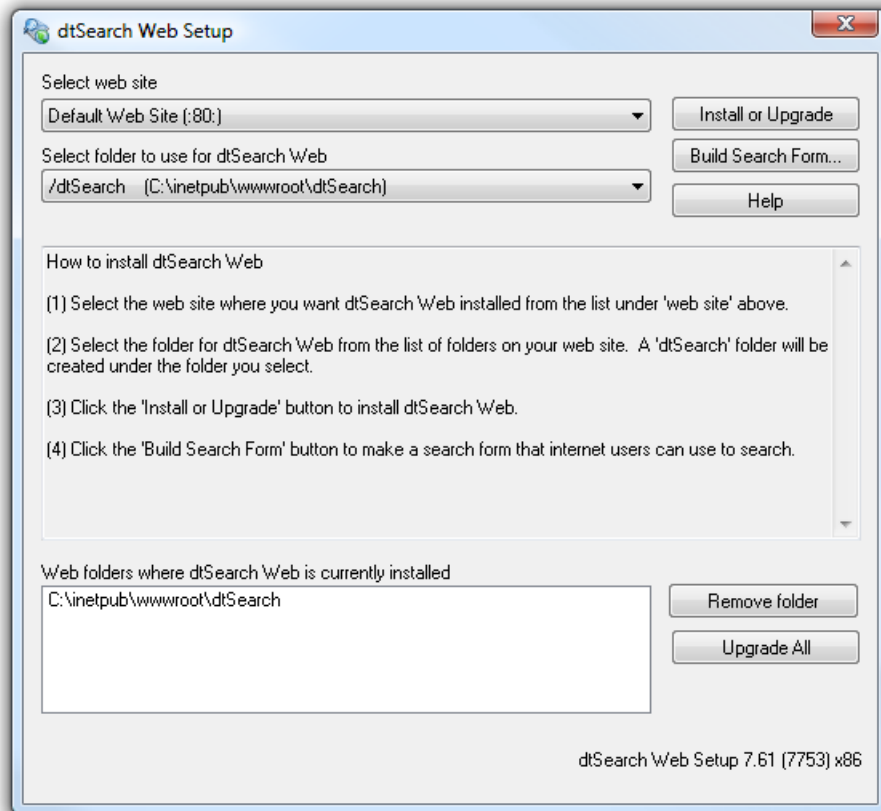
Note: When creating indexes to use with dtSearch Web, enable both the "Cache text" and "Cache original documents" options to ensure that search results and highlighted documents appear quickly.

3. Make a virtual directory for your documents

Designate each folder that contains documents to be published as a virtual directory. Open Internet Services Manager, select the web site, and click **Action > New > Virtual Directory**.

4. Run dtSearch Web Setup

In dtSearch, click **File > dtSearch Web Setup**. dtSearch Web Setup will display a drop-down list of the web sites on your server, and below it a list of the virtual directories defined for each site.



5. Select the web site to use with dtSearch Web

If you have more than one web site on this server, you can install dtSearch Web on each of them, or just one. (To install dtSearch Web on additional web sites, repeat the procedure described in this Quick Start for each site.)

6. Select the folder where dtSearch Web should be installed

dtSearch Web Setup will create a "dtSearch" folder under the folder you select where the dtSearch Web files will be installed. Click the **Install or Upgrade** button to install the dtSearch Web files in the folder you selected.

If the folder you selected does not already have 'Execute' permission enabled, dtSearch Web Setup will ask if it can enable 'Execute' permission in the folder so dtSearch Web can run from that folder.

7. Build a Search Form for your site

Click the **Build Search Form...** button to build a search form to use with your site. You can make as many search forms as you want for each site. After dtSearch Web Setup has generated a search form, you can use an HTML editor such as Microsoft FrontPage to edit it to fit into your web site.

In the **Form Builder** dialog box, click the **Indexes** tab to select indexes for the search form. Check the box next to each index to include it on your search form.

8. Click OK to build the search form.

After the search form is built, dtSearch Web will open it in your browser so you can try out a search. Once you have a basic search form working, you can run **Form Builder** again to customize the search form, the appearance of search results, and other options.

Logging

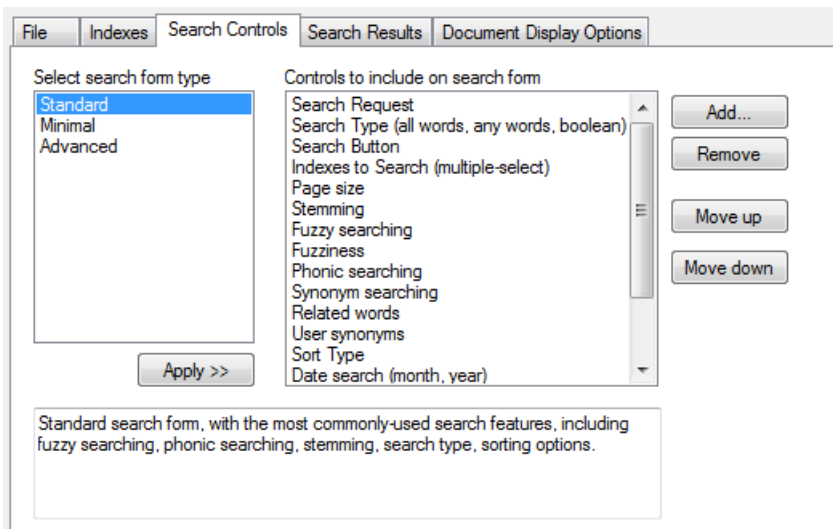
To enable logging in dtSearch Web, check the **Log document access** or **Log searches** checkboxes in the **File** tab of the **Form Builder** dialog box. For information on customizing logging, see the "Generated Files" chapter of this manual.

Customizing the Search Interface

Customizing the Search Form

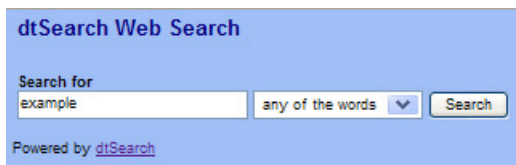
The **Search Controls** tab in the **Form Builder** lets you specify the contents of the search form that will be generated.

Once the form is generated, you can edit it using any HTML editor to customize the appearance to match the rest of your web site.



Click a search form type under **Select search form type**, and click **Apply>>**, to set up one of the standard sets of search controls. Once you have done this, you can click **Add...** or **Remove** to add or remove specific controls. Use **Move up** and **Move down** to change the order of the controls on the search form.

"Minimal" search form



"Standard" search form

dtSearch Web Search

Search for
example any of the words Search

Stemming
 Fuzzy searching
 Phonic searching

Sort type hits

Powered by [dtSearch](#)

"Advanced" search form

dtSearch Web Search

Search for
example any of the words Search

Items per page 10

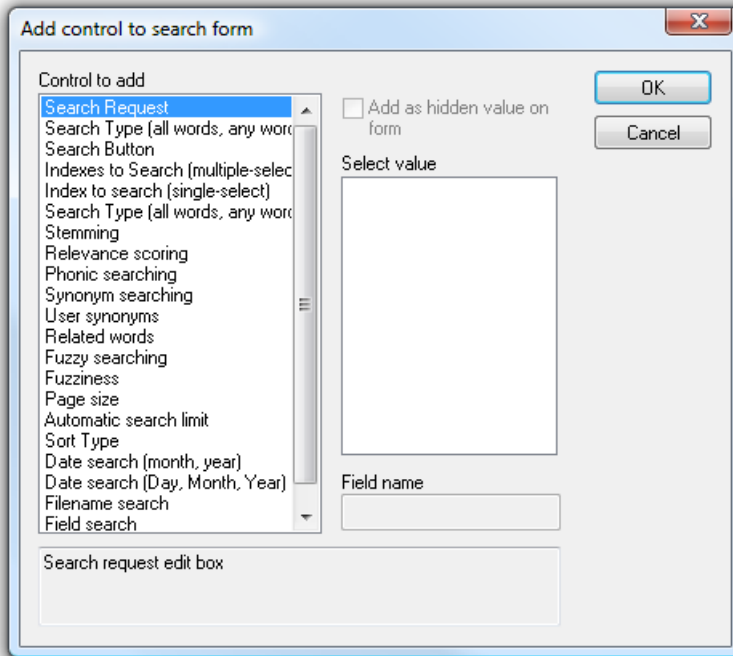
Stemming
 Fuzzy searching -- Fuzziness 0
 Phonic searching
 WordNet synonyms
 WordNet related words
 User-defined synonyms

Sort type hits

File date
Jan 1999 to Aug 2005 all

Powered by [dtSearch](#)

Adding a control



To add a control:

1. Select the type of control to add under **Control to add**.
2. To add a hidden control that specifies a default value, check the **Add as hidden value on form** box and select one of the listed values.

To add custom fields such as Subject or Author:

1. Select the **Field search** field type under **Control to add**.
2. Enter the name of the field under **Field name**.

Selecting Indexes to Include

In the **Form Builder** dialog box, click the **Indexes** tab to select the indexes that you want to put on the search form. Check the box next to each index to include it on your search form.

When you check the box next to an index name, dtSearch will include that index in the list of indexes on the search form.

Check the box to **Encrypt index paths in search form** to have the search form use numeric codes instead of folder names to identify indexes in the search form.

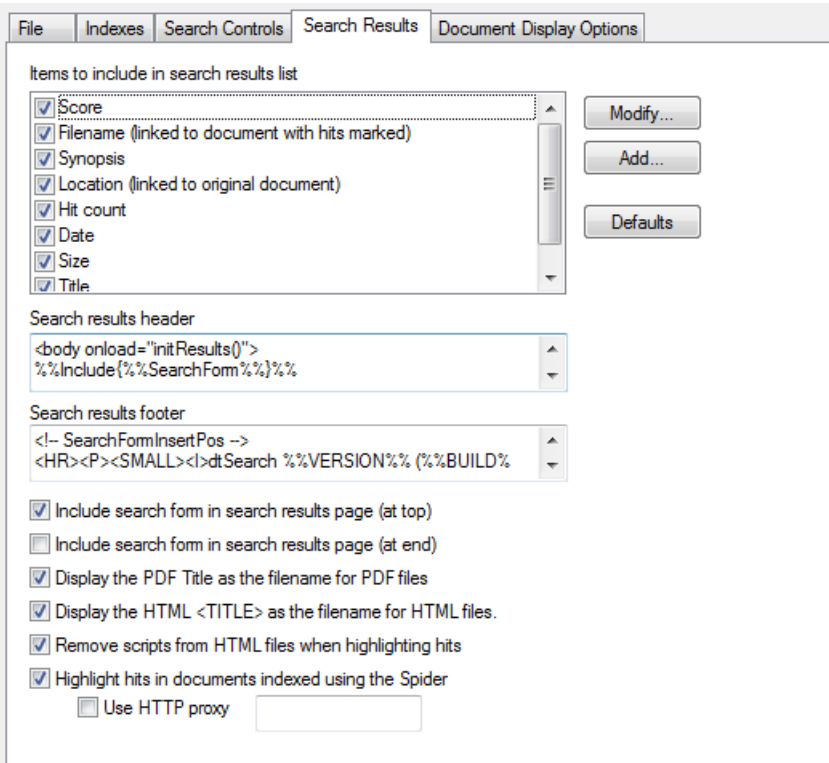
FindPlus® Integration

dtSearch FindPlus integration allows a dtSearch Desktop user to combine searches of local and network indexes with dtSearch Web searches through the dtSearch Desktop user interface. If some of your users will have dtSearch Desktop, check the box labelled **Make these indexes accessible to users through dtSearch Desktop** to allow these users to search directly from dtSearch Desktop. The search form generated will have a "Get index library" link added that these users can click on to integrate dtSearch Desktop with your site. Once a user has done this, the indexes listed on your search form will be added to the user's **Search** dialog box in dtSearch.

Please see the "Searching using dtSearch Web" topic in the dtSearch help file for more information.

Customizing the Search Results Format

The **Search Results** tab in the **Form Builder** lets you change the way dtSearch displays the results of a search.



Items to include in search results list

Add or remove items that will appear in search results. To remove an item on the list, un-check the box next to it. Click the **Add...** button to add a new item. For each item, you can use **Modify...** to customize the way it appears in search results.

Search results header

Search results footer

This is the HTML that appears at the top and bottom of the search results list. You can modify it to change the way search results appear. The header and footer can contain the following special codes:

Symbol	Meaning
%%Request%%	The search request that the user entered.
%%FileCount%%	The number of documents retrieved.
%%HitCount%%	In search results, the total number of hits in all files. For a document, the number of hits in the document.

Display the PDF Title as the filename for PDF files

PDF files have a "Title" attribute that is usually more readable than the filename. Check this box to make the Title appear in search results instead of the filename.

Display the HTML <TITLE> as the filename for HTML files

HTML files also have a title, included between <TITLE> and </TITLE> tags in the file, that is often more readable than the filename. Check this box to make the HTML title appear in search results instead of the filename for HTML files.

Remove scripts from HTML files when highlighting hits

HTML files often contain JavaScript code that may not work properly when the file is displayed outside of its usual context. For example, the JavaScript may refer to documents or objects that would appear in another frame. Check this option to have dtSearch remove any JavaScript it finds in an HTML file when highlighting hits. (The original HTML file will be unaffected. This option only affects what dtSearch Web will display when highlighting hits in an HTML file returned after a search.)

Highlight hits in documents indexed using the Spider

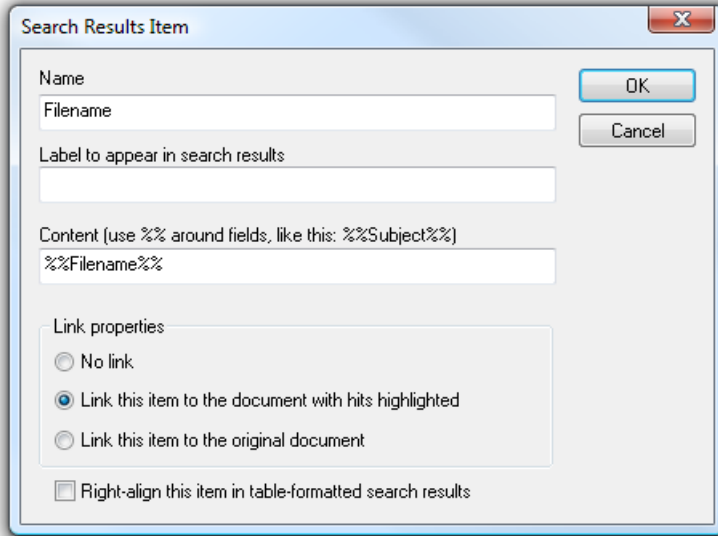
Check this box to enable highlighting of hits in documents that were indexed using the Spider. To make hit highlighting faster, create your index with the "Cache original documents" option enabled.

Use HTTP proxy

When highlighting documents indexed using the Spider, dtSearch Web must download a local copy of the file (unless the index was created with caching of original documents enabled). Check the "Use HTTP Proxy" box, and supply the URL of the proxy server to use, if dtSearch Web should use a proxy server when downloading web pages.

Modifying a Search Results Item

When you click **Add...** or **Modify...** in the **Search Results** tab of the **Form Builder**, you can specify the content of an item to appear in search results.



Name

The name of the item. This name is only used in the list of items in the **Search Results** tab and is not displayed in search results.

Label to appear in search results

This label is any HTML that you want to appear in front of the item. For example `Date:` would put **Date:** in front of the field.

Content

This is the information about the document that you want to appear in search results. The following pre-defined fields can be used:

String	Purpose
%%Hits%%	Hit count
%%Filename%%	Name of the retrieved document
%%Location%%	Path to the retrieved document
%%Date%%	Modification date of the document
%%Size%%	Size of the document
%%Synopsis%%	Brief segment of text showing the first few hits in context
%%Title%%	Text from the first few lines of the document, or the TITLE of an HTML document

Additionally, if you have stored user-defined fields in the index, you can insert these as well. For example, if you set up a stored field named "Subject", put %%Subject%% in the Content to insert the value of this field for each document.

Link properties

Select the type of link to insert for this search results item, if any.

Document Display Options

The **Document Display Options** tab in the **Form Builder** lets you specify how documents will appear in search results.

The screenshot shows the 'Document Display Options' tab in a software interface. It contains several text input fields and checkboxes. The 'Header' field contains HTML code: `<body onLoad="nextHit()">%%Filename%% (%%HitCount%%)</body>`. The 'Footer' field contains: `<HR><SMALL><l>dtSearch %%VERSION%% (%%BUILD%%)</l></SMALL>`. The 'Before hit' field contains: ``. The 'After hit' field contains: ``. The 'File types to display without conversion to HTML (list extensions)' field contains 'PDF'. There are two checkboxes: 'Allow display of documents that are not located in a virtual root folder.' and 'Display document properties'. A 'Defaults' button is located at the bottom right.

Header

Footer

This is the HTML that appears at the top and bottom of the search results list. You can modify it to change the way search results appear. The header and footer can contain the following special codes:

Symbol	Meaning
%%Filename%%	The name of the document
%%HitCount%%	The number of hits in the document

Before hit

After hit

HTML that is used to mark hits.

File types to display without conversion to HTML

List any file types that you want dtSearch Web to display without conversion to HTML. For example, if you want the Microsoft Word viewer plug-in to display .DOC files, add DOC to the list. PDF files are always displayed in Adobe Reader, without conversion to HTML.

Allow display of documents that are not located in a virtual root folder

This box should not be checked unless for some reason it is impossible to create a virtual directory for the documents on your site. Please see the **Security** topic before changing this setting.

Display document properties

Check this box to display document summary information (such as the Author or Subject field in a Word document) following the main text of each document.

Technical Information

Generated Files

For each search form generated in **Form Builder**, a set of HTML files is created. Assuming the form is named `dtsearch.html`, the files created are:

File	Purpose
<code>dtsearch.html</code>	Frameset for the search form and search results panels. The left panel will contain the search form, and the right panel will initially contain search request help. After a search, search results will appear in the left panel, and retrieved files will appear in the right panel.
<code>dtsearch_bar.html</code>	Button bar that appears at the top of search results.
<code>dtsearch_form.html</code>	Search form containing the list of indexes and the search options.
<code>dtsearch_minihelp.html</code> <code>dtsearch_help.html</code>	Help on search requests. This will appear in the right panel when the user initially opens the search form.
<code>dtsearch_options.html</code>	The <code>dtsearch_options.html</code> file contains settings that dtSearch Web uses to generate search results and to format retrieved documents for display.
<code>dtSearch_WebSearchForm.css</code> <code>dtSearch_SearchResults.css</code>	Style sheets that control the appearance of the search form and search results.
<code>dtSearch_WebSearchForm.js</code>	JavaScript that implements features of the

dtSearch_SearchResults.js search form and search results, such as hit
Utilities.js navigation.

After the search form is generated, you can edit it in an HTML editor to customize the appearance. You can also replace the text in dtsearch_help.html with other explanatory text, such as a detailed description of what is in each index. When editing the search form, be sure not to remove the META tag at the top of the form that looks like this:

```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html ;  
charset=utf-8" >
```

This tag ensures that non-English characters in your search form will be handled correctly. If you move the search form to another web page, copy the META tag into the <HEAD> area of the new page.

Moving the files

When dtSearch Web receives a search request, it uses the name of the search form to find the options file, which it uses to format search results. For example, if the search form is dtsearch_form.html, it will look for the options file in dtsearch_options.html. Therefore, if you move or rename the search form, it is important to move or rename the options file so that dtSearch Web can find it. If the search form name ends with "_form.html", the options file should end with "_options.html". If the search form name does not end in "_form.html", the options file should be the same as the search form name, but with "_options" added before the filename extension.

Examples:

Search Form Name	Options File Name
example_form.html	example_options.html
example.html	example_options.html

Another way to ensure that the options file remains linked to the search form is to add a hidden form variable with the full path and filename of the search form, like this:

```
<input type="hidden" name="OrigSearchForm"  
value="/dtsearch_form.html" >
```

dtSearch Web will check this form variable only if the filename matching method described above does not work.

The Search Form

The search form generated by dtSearch Web Setup is a standard HTML form that you can edit in an HTML editor or cut and paste into other pages.

Form Variables

dtSearch Web recognizes the following form elements:

Form Element	Meaning
cmd	A hidden form element that must have the value "search"
request	Search request
fileConditions	Optional additional search criteria based on a file's name, modification date, or size
booleanConditions	Optional additional boolean search criteria
searchType	Specifies the type of query syntax used in the request. If present, must be one of: allwords, anywords, phrase (for exact phrases), bool.
fuzziness	Level of fuzziness in a fuzzy search (0-9)
fuzzy	Enables fuzzy searching
index	Path to the index to search on the server
autoStopLimit	Search will automatically halt after this many documents have been found
maxFiles	Maximum number of files to retrieve (selects the best-matching files)
pageSize	Number of items to return per page of search results
phonic	Enables phonic searching
sort	Sorting method (name, hits, size, or date)
stemming	Enables stemming
synonyms	Enable synonym searching
userSynonyms	In synonym searches, use the user thesaurus
wordNetRelated	In synonym searches, use WordNet related words (antonyms, subcategories, etc.)
wordNetSynonyms	In synonym searches, use the WordNet synonyms
searchFlags	A numerical value with any combination of the search flags in the dtSearch developer API. See dtSearchApiRef.chm for more information on search flags
returnXml	Return search results as XML rather than HTML

If `fileConditions` and/or `booleanConditions` are included on the form, and are not blank, then they are combined with the user's search request. All conditions included in a search must be satisfied by each document retrieved.

Frames

dtSearch Web uses a frame-based layout for the search results and retrieved documents. The frame is set up before the search, with the search form on the left and search help on the right. After a search, the search results list appears in the left frame, and the documents appear in the right frame. The frame layout is created in `dtsearch.html`, which is a small HTML file that defines the frame boundaries and contents. The search form is in `dtsearch_form.htm`, which `dtsearch.html` references. To rearrange the positions of the frames, you can edit `dtsearch.html` in a text or HTML editor.

To eliminate the frames, link to `dtsearch_form.html` instead of `dtsearch.html`. After a search, the search results will fill the browser window, and retrieved documents will go in a separate full browser window.

Selecting Multiple Indexes

Using the standard dtSearch Web search form, your users can select multiple indexes to search by holding down the CTRL key and clicking on the index names.

You can also edit the search form to add a single option that would cover multiple indexes. In an HTML "Select" control, the options look like this:

```
<option value="something"> visible text
```

The "visible text" appears in the list of choices, and if it is selected, the *value* is what gets sent to the server. To add an option that includes more than one index, make the *value* a list of index paths, with commas separating them. The visible text can be anything you want. Example:

```
<option value="C:\indexes\first, c:\indexes\second,  
c:\indexes\third"> Search all indexes
```

Sorting

If sort is not "size", "name", "date", or "hits", then dtSearch Web will assume that the sort key is a stored field and will use the `dtsSortByField` search flag. Sort can be followed by a colon and a numerical value that will be combined with the sort type. Example: "subject:0x210002". See `dtSearchApiRef.chm` for more information on sort flags.

hits

In a search that is sorted by hits, dtSearch will return up to maxFiles of the most relevant documents, organized into pages each with pageSize documents. If pageSize is not specified in the search form, the maxFiles value will be used as the page size.

date

Sorting by date works like sorting by hits, except that the most recent documents are returned instead of the most relevant.

size, name, and custom fields

When sorting by criteria other than hits or date, dtSearch will return up to maxFiles of the *most relevant* files, organized into pages each with pageSize documents, with the entire results list sorted by the specified criteria. For example, if the sort criterion is "size", pageSize is 10, and maxFiles is 100, dtSearch will find the 100 most relevant files (not the 100 largest), and will display them in pages of 10 documents, sorted by size.

The Options File

The option settings in dtsearch_options.html control the appearance of search results and retrieved files, generation of the synopsis, logging, and other options.

Each setting is bracketed with HTML comments, like this:

```
<!-- $Begin DocHeader -->
%%Filename%% (%%HitCount%% hits)
<!-- $End -->
```

The text that appears between the \$Begin and \$End comments has to be valid HTML. Text that is not between \$Begin and \$End comments is ignored, and can be used to insert explanatory comments.

Template Settings

Because dtsearch_options.html is a valid HTML file, you can edit it directly in an HTML editor to change the appearance of retrieved documents or search results. When editing the HTML, be careful to keep the \$Begin and \$End comments around each option setting.

Setting	Purpose
DocHeader	Text displayed above each retrieved document
DocFooter	Text displayed below each retrieved document
DocScript	JavaScript inserted in each retrieved document to

	enable hit navigation
ResultsHeader	Text displayed above each search results list
ResultsFooter	Text displayed below each search results list
ResultsScript	JavaScript inserted in each search results list to enable navigation between documents.
BeforeHit	Text displayed before each hit in a document
AfterHit	Text displayed after each hit in a document
ResultsTableHeader	Top row of search results table (for column labels)
ResultsTableItem	Format of each item in the search results table
ResultsTableFooter	End of search results table (generally </table>)

In **ResultsTableItem**, the following symbols identify where document-related information is displayed:

Symbol	Purpose
%%Hits%%	Hit count
%%PhraseCount%%	Hit count, counting a phrase as a single hit
%%HitsByWord%%	List each word or phrase found in the search and the number of hits for each
%%Filename%%	Name of the retrieved document
%%Synopsis%%	Brief snippet of text showing the first hits in the document with a few words of context around each hit. See Synopsis Settings, below.
%%Location%%	Path to the retrieved document
%%Date%%	Modification date of the document
%%SizeK%%	Size of the document in kilobytes
%%Size%%	Size of the document in bytes
%%Title%%	Text from the first few lines of the document, or the TITLE of an HTML document
%%DirectLink%%	The string to be used for the HREF for an http: link to the document without hits highlighted (to directly open a file over a web connection)
%%HighlightLink%%	The string to be used for the HREF for a link to the document with hits highlighted
%%LocalLink%%	The string to be used for the HREF for a link to the local path to the document (for use with dtSearch Publish only)
%%ThisDocOrdinal%%	Sequential number of this item in the search results list (1, 2, 3, ...)

The search results format can also include a string that tells dtSearch Web to include the search form at the end of the search results list. This string is:

```
%%Include{%%SearchForm%%} %%
```

Link Types

ResultsTableItem usually will contain one or more links to open the retrieved document. In these links, use %%DirectLink%%, %%HighlightLink%%, or %%LocalLink%% to specify the type of link you want. For example:

```
<a href="%%HighlightLink%%">%%Filename%%</a>
```

This would produce a link with the name of the document in the link. When clicked, the document would open with hits highlighted.

If you want to open the file without highlighting, to allow the native viewer for the file to open it, use either %%DirectLink%% or %%LocalLink%%.

%%DirectLink%% produces a link to the file using http, and is appropriate for use over a web connection (example:

http://www.example.com/documents/sample.doc). %%LocalLink%% produces a link to the location of the file on disk (example: c:\documents\sample.doc) and is appropriate for use in dtSearch Publish or to give users on a LAN direct access to files.

Option Settings

The options file also contains settings that control searching behavior and the way links and file information appear in search results.

Setting	Purpose
FieldWeights	Specify the weight for hits that occur in fields. Example: subject:10, HtmlTitle:15
HtmlRemoveScripts	Disable JavaScript in retrieved HTML files
HtmlUseTitleAsName	Use the Title of HTML files as the filename
PdfUseTitleAsName	Use the Title of PDF files as the filename
MaxUrlSize	Maximum size of a URL to generate
MaxWordsToRetrieve	Maximum number of words to match in a single search
MaxWordsMessage	Message to display when too many words matched
UnconvertedTypes	File types to display without conversion to HTML
NoFilesMessage	Message to display when no files are retrieved
HighlightHttpDocs	Highlight hits in documents indexed via HTTP (using the dtSearch Spider)

HttpProxy	Proxy server to use to access web resources
SERVER_NAME	Server address for the dtSearch Web server in search results. (Specify only if it is necessary to override the automatically-detected server name.)

Synopsis Settings

The %%Synopsis%% symbol in search results represents a brief snippet of text including the first hits in each document, with a few words of context around each hit. The settings below provide options to customize how the synopsis is generated.

Performance

Generating a synopsis requires that dtSearch Web open the original document and scan through it to extract the text around each hit, which can be a time-consuming operation. To make generation of a synopsis faster, enabling caching of text when you create the index of the documents. For more information on this option, see "Caching Text" in the dtSearch Desktop help file.

Formatting

dtSearch Web will format the synopsis so it can be inserted into a search results table. Line breaks, paragraph formatting, colors, and extra spacing will all be removed to produce a simple snippet of text, with hits marked in bold.

Setting	Purpose
SynopsisMaxContextBlocks	Number of blocks of context to include in the synopsis.
SynopsisContextHeader	Text to include in front of each block of context.
SynopsisWordsOfContext	Number of words to include around each hit in the synopsis.
SynopsisMaxWordsToRead	Number of words in each document to scan looking for blocks of context to include in the synopsis.

Log Settings

To enable logging in dtSearch Web, check the **Log document access** or **Log searches** checkboxes in the **File** tab of the **Form Builder** dialog box. This will set up the search form for default logging of search requests or document access. The options, like the options for document display, are controlled by a list of templates that you can customize by editing the generated options file.

Setting	Purpose
LogSearches	Set to 1 to enable logging of all search requests
LogDocumentAccess	Set to 1 to enable logging of document access
DocumentLogNameTemplate	Template used to generate the filename for the document access log
DocumentLogItemTemplate	Template used to generate a single entry in the document access log
SearchLogNameTemplate	Template used to generate the filename for the search log
SearchLogItemTemplate	Template used to generate a single entry in the search log

The two filename templates, `DocumentLogNameTemplate` and `SearchLogNameTemplate`, are used to generate log filenames. By building date symbols into the log name, you can have a new log file start every day, month, or year. Example:

`c: \logs\SearchLog%%Year%%-%%Month%%.log`

The two item templates, `SearchLogItemTemplate` and `DocumentLogItemTemplate`, are used to generate the lines added to the log file. The following symbols can be used in the templates to customize the content of the logs:

Symbol	Meaning
%%DateTime%%	The date and time of the search
%%Result%%	"OK" if the request succeeded, "DENIED" if access was denied, "FAILED" on other errors
%%REMOTE_USER%%	The value of the REMOTE_USER HTTP variable (unless your site requires a login, it will be blank)
%%REMOTE_ADDR%%	The value of the REMOTE_ADDR HTTP variable (the IP address of the user accessing the site)
%%DocName%%	The name of the document accessed (document log only)
%%SearchRequest%%	The user's search request (search log only)

%%FileConditions%%	The value of the fileConditions form variable (search log only)
%%BooleanConditions%%	The value of the booleanConditions form variable (search log only)
%%SearchIndex%%	The index (or indexes) searched (search log only)
%%DocCount%%	The number of documents retrieved (search log only)
%%Month%%	The month of the search (01-12)
%%Day%%	The day of the search (01-31)
%%Year%%	The year of the search (4-digit)

The log files are generated in CSV format, which you can import into Excel or most databases.

Virtual Directories

When users access your web site through a web server, they do not see the same folders and files that you would see in Explorer. Instead, they see "virtual" folders with names like "/Scripts" or "/Docs" that you map to folders on your hard disk when you set up your web server.

Virtual folders (also called virtual roots, virtual paths, or virtual directories) are the names of folders that you decide to publish on your site. For example, if your site is www.sample.com and you decide to publish `c:\website\docs` on your hard disk as "/Docs", then users accessing that folder would go to <http://www.sample.com/Docs>. The local path is `c:\website\docs`, and the virtual path is /Docs.

To set up a virtual folder using Internet Information Server, open Internet Service Manager and click the right mouse button on the web site entry. (It will usually be called "Default Web Site" unless you have renamed it.) Select **New > Virtual Directory** and enter the name and location of the virtual directory that you want to create.

Note: After you have created a new virtual directory, run the dtSearch Web Setup program again so that dtSearch Web will know about the new directories. (You do not have to do anything in dtSearch Web Setup -- just open the program and then close it again.)

Security

dtSearch Web does not alter Windows security settings and only provides access to documents when the user seeking access has the necessary permissions. To secure a site, or to make a site open to the public, use Explorer and Internet Service Manager to set the permissions you want and dtSearch Web will recognize those permissions automatically. There is no need to rebuild your indexes after changing security settings.

Documents on an internet site are usually placed in virtual directories. These are folders that have been designated as part of your site and that have been given an "alias" such as /Docs. dtSearch Web will only display documents that are located in a virtual directory, and will display an error message if a user tries to access documents located in other folders. The purpose of this is to provide an additional layer of protection against unauthorized access to documents.

There is an option to override this setting in the Document Display Options tab, but this option is **not** recommended, because of the risk that documents could be inadvertently made available on the web through dtSearch Web.

Windows Server Installation

Windows Server has a "Web Service Extension" manager that prevents a web application from running unless it is specifically enabled. When an application is invoked and it has not been enabled, a 404 ("Not Found") error results.

dtSearch Web Setup can automatically register dtSearch Web with Windows Server with the Web Service Extensions manager. To have dtSearch Web Setup do this, run the dtSearch Web Setup wizard. When the setup program asks if it should register dtSearch Web with Windows Server, click the button, **Yes, register dtSearch Web with Windows Server**.

If you prefer to register dtSearch web with Windows Server yourself, follow the steps below.

Before you install

- (1) Make sure that the Internet Information Service (IIS) is installed. By default, IIS is not installed in Windows, so you must either use Add/Remove Programs or Configure Your Server to set it up.
- (2) Log in as the Administrator or a user with equivalent rights. Other user accounts will not have sufficient permissions to enable a web application.

Installing dtSearch Web

- (1) Run dtSearch Web Setup as usual. When you install dtSearch Web, note the location where dtisapi6.dll was copied (you will need it in step 7 below).
- (2) Open Internet Services Manager.
- (3) In Internet Services Manager, open "Web Service Extensions".
- (4) Click "Add a new Web service extension".
- (5) Under "Extension name" enter "dtSearch Web".
- (6) Click "Add..." to add a file under "Required Files".
- (7) Locate the dtisapi6.dll file under the c:\inetpub folder, and select it.
- (8) Click OK.

Portable Media Publishing

Publishing (Overview)

dtSearch Publish provides an easy way to publish documents on a CD, DVD, or other portable media, using a browser-based user interface so users can access the content as they would access a web site. dtSearch Publish uses the same search components and templates as dtSearch Web, so the search forms and customization options look exactly the same as for dtSearch Web running on a web server.

Note: a dtSearch Publish distribution license is required before media containing dtSearch components can be distributed to users.

Some advantages of a browser-based user interface are: it provides high-quality display of HTML files, so web sites will appear just as they do on the original site; it can be customized simply by changing some HTML files; it is easy for customers to use; and no software has to be installed on the user's hard disk to access the content.

As with dtSearch Web, PDF and HTML files are displayed exactly as they would appear in a web browser, but with hits highlighted. Other file types are converted to HTML with hits highlighted for display in the browser. To use a CD or other media created with the CD Wizard, the user needs to have Windows XP or newer and, for viewing PDF files, Adobe Reader. To enable hit highlighting in PDF files, the user will need a compatible PDF viewer and may need a plug-in. For more information, please see "PDF viewers that support highlighting hits".

Note: While this manual refers to publishing to "CD", the content can be located on any type of media that is accessible through the file system, including DVDs, flash drives, USB drives, internal drives, etc.

Using the CD Wizard

The CD Wizard helps you to create one or more CD master folders. Each folder contains a set of documents, software, and dtSearch indexes that is ready to be transferred to a CD. You can create any number of CD master folders, and each folder can contain any number of document folders and indexes.

Creating a CD

To create a CD master folder,

1. Install the dtSearch Publish program files on your computer.

If you have dtSearch Web on CD, run the setup program to install. If you downloaded dtSearch Web from the internet, follow the download instructions to open the dtSearch Web archive and install the files.

2. Start the dtSearch CD Wizard

In dtSearch, click **dtSearch CD Wizard...** in the **File** menu.

3. Make a new CD master folder

Click **New CD...** to make a new CD master folder, enter the name of the folder for the CD contents, and click **OK**.

4. Add documents to the CD

Click **Add Documents...** to add documents to the folder. The **Add Documents** dialog box will appear. When the CD master folder is set up, a root\data folder will be created where the documents should be stored. To add documents to the CD, you can copy them into this folder using Windows Explorer, or you can use the **Add Folder...** button to have the CD Wizard do this automatically. When you click OK after selecting a folder to add, all of the documents in the folder will be copied into the CD master folder.

5. Create an index for the documents

Click **Create Index...** to create an index for the documents. You can create any number of indexes on each CD (just click **Create Index...** for each one). The process of creating and updating an index works exactly as it does in dtSearch Desktop.

6. Build a search form

Click the **Build Search Form...** button to build a search form to use with your site. You can make as many search forms as you want for each site.

After the search form is built, dtSearch Web will open it in your browser so you can try out a search. Once you have a basic search form working, you click Build Search Form again to customize the search form, the appearance of search results, and other options, and to create additional search forms.

7. Make a "home" page for the CD

The home page is the first page that users will see when they insert the CD. The home page is named index.html and is located in the root\data subfolder of the CD master folder.

Once the CD is done, transfer the contents of the CD master folder to a CD. Copy the contents of the CD master folder, but not the CD master folder itself. For example, if the CD master folder is C:\CDMaster, there should not be a folder named CDMaster on the CD, but everything in C:\CDMaster should be copied to the CD. (This way the autorun.inf file that the CD Wizard creates will be in the root folder of the CD.)

Modifying a CD

To add more documents, click **Add Documents...** and copy additional folders to the CD. After you have added documents click **Update Index...** to add the new documents to your indexes.

For information on changing the CD type, see **CD Types**.

Deleting a CD

To delete a CD master folder, just delete the folder in Windows Explorer. The CD Wizard will detect that the folder is gone the next time it runs and will remove it from the drop-down list of CD master folders.

"Recognizing" a CD

To access a CD master folder that the CD Wizard does not list (for example, if the CD master folder was copied from another computer), click **Recognize CD** and browse to the CD master folder. The CD master folder will be added to the list in the CD wizard.

CD Types

dtSearch Publish can generate two types of CDs:

- (1) "Standard" CDs that use the dtSearch Publish viewer, lbview.exe, to view content on the CD.
- (2) "Local HTTP Server" CDs that use Internet Explorer to view content on the CD, and use a localhost-only HTTP server to enable Internet Explorer to access the CD.

For most applications, the "Standard" CD type is recommended.

To change the type of a CD that you have already set up, click **Change CD Settings...** and select one of the CD types.

Standard CDs

A Standard CD uses a viewer program, `lbview.exe`, to view content on the CD. Web pages and search forms appear exactly as they would in Internet Explorer, and all search functions work as they would on a web site, including any JavaScript embedded in search forms.

Because no HTTP server is needed, the CD is not affected by firewall software such as Zone Alarm, Norton Internet Security, and the Windows Firewall.

CD Layout

The top-level folder of the CD will contain these files:

<code>autorun.inf</code>	Specifies the program to start when the CD is inserted
<code>cdrun.exe</code>	The program launched when the CD is inserted, as specified in <code>autorun.inf</code>
<code>cdrun.xml</code>	Options specifying what <code>cdrun.exe</code> should do when it starts

Below this folder will be a **root** folder with two sub-folders:

- **data**, which has documents, indexes, and the search forms, and
- **cgi-bin**, which has the search program and any other executable content that you add to the CD.

The `root\data` folder will be equivalent to the `/` folder on a web site, so `/something.html` would be found in `root\data\something.html`

The `root\cgi-bin` folder will be equivalent to the `/cgi-bin` folder on a web site, and can hold any CGI programs for the web site.

Startup

The startup sequence when a CD is inserted into a user's CD drive is as follows.

1. Windows opens the `autorun.inf` file to get the program to launch, which is `cdrun.exe`.

2. cdrun.exe starts and checks that the components listed in the dependencies section of cdrun.xml are present. If any components are missing, cdrun.exe will display a warning message (as specified in cdrun.xml) and exit.
3. cdrun.exe launches lbview.exe, the browser that accesses the CD.
4. lbview.exe starts and checks the lbview.ini file for the recommended Internet Explorer and Adobe Reader versions. If either product is older than the recommended version, the user will be prompted to download the latest version from the Microsoft or Adobe web site.
5. lbview.exe opens the home page for the CD, root\data\index.html
6. If the user clicks the search icon in the browser, the search page for the CD will open, root\data\dtsearch.html

System Requirements

Windows version: Windows 2000 SP 4 or later.

Internet Explorer version: Internet Explorer 6 or later is required to enable hit navigation and hit highlighting to work.

Adobe Reader version: A PDF viewer that supports highlighting hits is required or PDF files will appear without hit highlighting. For information on PDF viewers that support hit highlighting, please see PDF viewers that support highlighting hits.

Error Pages

The root\data\builtin folder on the CD contains these pages that are displayed in lbview.exe when an error occurs.

GetNewIE.html	Appears at startup when the Internet Explorer version is older than what is specified in lbview.ini as the minimum recommended Internet Explorer version. This page contains a link that the user can click to suppress the page from appearing the next time the CD starts.
GetAdobe.html	Appears at startup when the Adobe Reader version is older than what is specified in lbview.ini as the minimum recommended version. This page

	contains a link that the user can click to suppress the page from appearing the next time the CD starts.
GetPlugin.html	Appears at startup when an Adobe Reader plug-in is needed to highlight hits in PDF files. This page contains a link that the user can click to suppress the page from appearing the next time the CD starts.
Error.html	Appears when any other browsing error occurs, such as a broken link leading to a "page not found" error.

lbview.ini Settings

The following settings in lbview.ini can be used to control the behavior of lbview.exe

HomePage=/index.html

Specifies the first page that opens when the CD starts.

SearchPage=/dtSearch.html

Specifies the page that opens when the user clicks the search button.

WebLinksInBrowser=1

Specifies whether external links should open in the user's web browser or in the lbview.exe program. For example, suppose a page on your CD contains a link to <http://www.microsoft.com>. If WebLinksInBrowser is set to 1, when this link is clicked the user's web browser will open over the lbview.exe program. If WebLinksInBrowser is set to 0, the link will open in the lbview.exe program.

ExternalLaunchExtensions=xls;doc;ppt;wpd;docx;pptx;xlsx;rtf

List extensions to launch outside of the lbview.exe program. This enables you to create links in the CD to Word or other Office documents and have them open in their application when the user clicks a link.

CopyFileExtensions=zip

List extensions to copy instead of opening when clicked. These extensions can be used to create links to automatically copy certain files from the CD.

MinimumAdobeReaderVersion=6

Specifies the minimum version of Adobe Reader (or Adobe Acrobat) recommended to use with this CD. If an older version is present, or if Adobe Reader is not installed, the user will be prompted to get Adobe Reader from the Adobe web site. If your CD will not contain PDF files, you can set this to 0 (zero).

GetAdobeReaderPage=/builtin/GetAdobe.html

The page to display when a newer version of Adobe Reader is needed.

MinimumIEVersion=5

The minimum version of Internet Explorer recommended to use with this CD. If an older version is present, the user will be prompted to upgrade.

GetNewIEPage=/builtin/GetNewIE.html

The page to display when a newer version of Internet Explorer is needed.

MinimumPdfPluginVersion=1

The minimum version the dtSearch PDF Search Highlighter Plug-in for Adobe Reader that is needed to highlight hits. Set to zero to suppress prompts to download the plug-in.

GetPdfPluginPage=/builtin/GetPlugin.html

The page to display when the dtSearch PDF Search Highlighter Plug-in for Adobe Reader is not installed.

EnablePdfHighlightMessage=Would you like to enable hit highlighting for PDF files (currently disabled in Adobe Reader or Adobe Acrobat)?

SilentEnablePdfHighlighting=1

If PDF hit highlighting is disabled in Adobe Reader 9 or Adobe Acrobat 9, the lbview.exe program can change this setting to enable hit highlighting before opening a PDF file. You can use EnablePdfHighlightMessage to ask permission to make this change, or set SilentEnablePdfHighlighting to have lbview.exe make the change without asking permission.

The option to disable hit highlighting is present starting in Adobe Reader and Adobe Acrobat 9.0, in the Preferences > Search settings ("Enable search highlights from external highlight server").

Launching Programs from the Viewer

Standard CDs support a special URL format that lets you launch a program from a link in a page on the CD. To create a link that launches a program,

(1) Put the program in the root/cgi-bin folder under the CD master folder.

(2) Create a link on a page as follows:

```
<a href="cmd://launch/YourProgramName.exe">Click here  
to launch the program</a>
```

The lbview.exe viewer will execute the program YourProgramName.exe in root/cgi-bin when this linked is clicked.

Using Tabs

You can make a link open in a different tab using the *target* attribute in the link tag, like this:

```
<a href="othertab.html" target="_new">click here</a>
```

A user can close a tab by clicking the X icon, or you can create a link to close a tab by using javascript:window.close(), like this:

```
<a href="javascript:window.close();">click here to  
close this tab</a>
```

Internet Explorer normally pops up a warning message when a script attempts to close a window. To prevent this, lbview.exe watches for the exact sequence "javascript:window.close();" and closes the current tab silently when this occurs in a link.

Local HTTP Server CDs

The default and recommended type of CD to use with dtSearch Publish is the Standard type.

A Local Server CD uses an HTTP server to provide the browser interface. When the CD is inserted, the local HTTP server starts, and Internet Explorer is launched to open the home page for the CD.

This type of CD has a significant disadvantage: the HTTP server may trigger warning messages, or may be blocked, by firewall software such as Zone Alarm, Norton Internet Security, or the Windows Firewall. For the CD to work, the firewall software has to be configured to allow the HTTP server to "access the internet". The local HTTP server does not really access anything outside of the local machine, but because it communicates with the web browser using HTTP, some firewall software treats it as if it were accessing the internet. The Windows Firewall may pop up a warning message when the local HTTP server starts, but does not block it.

For Local HTTP Server CDs, you can use either the HTTP server that comes with dtSearch Publish, dts_svr.exe, or another product, Microweb.

Startup

When the CD is inserted, the autorun.inf file on the CD will start the cdrun.exe program on the CD. The cdrun.exe program will then do the following:

- (1) Check the cdrun.xml file to get configuration settings
- (2) Check that the computer has any required dependencies. These are listed in the cdrun.xml file. If a dependency is not present, the file in browserError.html is launched to prompt the user to get a newer browser version.
- (3) Start the HTTP server and then launch the home page for the CD (index.html) in the user's web browser. When the browser is closed, the HTTP server will stop automatically.

CD Layout

Assuming that a CD master folder is created in c:\sample, the following sub-folders will be created when the CD Wizard sets up the CD:

Folder	Contents
c:\sample\root\cgi-bin	dtSearch Web program files
c:\sample\root\data	Documents and indexes

System Requirements

Windows version: Windows 2000 SP 4 or later.

Internet Explorer version: Internet Explorer 6 or later is required to enable hit navigation and hit highlighting to work.

Adobe Reader version: A PDF viewer that supports highlighting hits is required or PDF files will appear without hit highlighting. For information on PDF viewers that support hit highlighting, please see PDF viewers that support highlighting hits.

Firewalls: If the system has a firewall installed, the program `dts_svr.exe` on the CD may have to be given permission to access the internet.

HTTP Server

The HTTP server is `dts_svr.exe`, a localhost-only HTTP server that runs on a unique port number, so it will not interfere with any other HTTP servers that may be running on the computer when the CD is inserted. A configuration file, `dts_svr.ini`, has option settings to control the behavior of the HTTP server (for example, to specify the browser to open when the CD starts). Documentation on these options is included in the `dts_svr.ini` file.

dtSearch Web interfaces with the HTTP server using the `dtcgi2is.exe` utility, which the CD Wizard sets up along with dtSearch Web. `dtcgi2is.exe` translates dtSearch Web's ISAPI interface into a CGI interface, so dtSearch Web can be run as a CGI program. `dtcgi2is.exe` maps virtual directories to local directories for dtSearch Web using the root folder provided in an XML configuration file, `dtcgi2is.xml`.

Dependencies

If your content requires a current browser version, plug-ins, or other software that must be installed, you can use the Dependencies entries in the `cdrun.xml` file to have your CD install these as needed.

Software Dependencies

If your content requires a specific browser version, plug-ins, or other software that must be installed, you can use the Dependencies entries in the `cdrun.xml` file to have your CD install these as needed. Each dependency is numbered from 0 to 9 and contains the following information:

Component	A DLL that must be loaded successfully for the dependency to be satisfied. The wizard-generated <code>cdrun.xml</code> checks for the Winsock 2 DLL, <code>WS2_32.DLL</code> .
Message	A prompt that will be displayed if the component is missing. If <code>FileToLaunch</code> is not blank, the message

must be a question asking the user for permission to install the software. If FileToLaunch is blank, the message should be an error message.

FileToLaunch	Program to execute if the component is missing (if blank, nothing will be launched). If the required software is on the CD, this should be the path to the executable, relative to the top folder on the CD.
ErrorPage	An HTML file to open if the dependency cannot be satisfied. This can contain links to a web site to download the required software.

Example 1. Displays a warning message if the Winsock 2 DLL is not installed.

```
<Dependency0>
  <Component>ws2_32.dll</Component>
  <Message>A newer version of Internet Explorer or
  Netscape is needed.</Message>
  <FileToLaunch></FileToLaunch>
  <ErrorPage>browserError.html</ErrorPage>
</Dependency0>
```

Example 2. Offers to install Internet Explorer if the Winsock 2 DLL is not installed. This example assumes that the CD will contain the Internet Explorer setup files in a folder named msie on the CD.

```
<Dependency0>
  <Component>ws2_32.dll</Component>
  <Message>A newer version of Internet Explorer is
  needed. Would you like to install Internet
  Explorer now?</Message>
  <FileToLaunch>msie\i setup.exe</FileToLaunch>
  <ErrorPage>browserError.html</ErrorPage>
</Dependency0>
```

cdrun.xml

The cdrun.xml configuration file controls the behavior of cdrun.exe, which is the program launched to view a CD created with dtSearch Publish. Each of the option settings in cdrun.xml is explained below.

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<dtSearchCdSettings>
  <UseLocalBrowser>1</UseLocalBrowser>
  <ServerToLaunch>root\data\lbview.exe</ServerToLaunch>
  <Dependency0>
    <Component>ws2_32.dll</Component>
    <Message>A newer version of Internet Explorer
or Netscape is needed.</Message>
    <FileToLaunch></FileToLaunch>
    <ErrorPage>apache\browserError.html</ErrorPage>
  >
</Dependency0>
</dtSearchCdSettings>
```

ServerToLaunch

To start the HTTP web server from the CD, cdrun executes the dts_svr.exe program. dts_svr.exe then launches the browser with the start page for the CD (default.html or index.html). dts_svr.exe will automatically close when the user closes the browser window. If UseMicroweb is true (see below) then ServerToLaunch should be set to root\data\microweb.exe

UseDefaultServer

UseMicroweb

UseLocalBrowser

Specifies the interface type for the CD.

Dependency0

Up to 10 dependencies, numbered from 0 to 9, can be included in a cdrun.xml file, specifying components or programs that must be installed for the CD to work. See "Software Dependencies" for information on the contents of this section.

Index

%		
%%Date%%	9, 17	
%%Filename%%	9, 11, 17	
%%HitCount%%	11, 17	
%%Hits%%	9, 17	
%%Location%%	9, 17	
%%Size%%	9, 17	
%%Title%%	9, 17	
A		
apache.exe	35	
apache_cd.exe	35	
autoStopLimit (dtSearch Web form variable)	15	
B		
booleanConditions (dtSearch Web form variable)	15	
C		
CD Publishing (Overview)	25	
CD Title	35	
CD Wizard	25	
cdrun.xml	35	
CdTitle	35	
CloseWarningSeconds	35	
CloseWarningText	35	
cmd (dtSearch Web form variable).....	15	
ConfFileTemplate	35	
CopyrightNotice	35	
Customizing Search Form	4	
Search results format.....	7	
Customizing the Search Form	4	
D		
Dependencies	34, 35	
Dependency0.....	35	
DialogCaption	35	
DialogText	35	
DocCount.....	17	
DocFooter.....	17	
DocHeader	17	
DocName.....	17	
Document Display Format.....	11	
DocumentLogItemTemplate	17	
DocumentLogNameTemplate	17	
dtSearch Publish	25	
dtSearch Web Quick Start.....	1	
dtSearch Web search form.....	13, 15	
dtsearch.html.....	13	
dtsearch_bar.html.....	13	
dtsearch_form.html.....	13	
dtsearch_help.html	13	
dtsearch_options.html	13, 17	
dtSearchCdSettings	35	
E		
ErrorPage	35	
F		
fileConditions (dtSearch Web form variable)	15	
FileToLaunch.....	34, 35	
FindPlus.....	6	
Form Builder	4, 7	
G		
Generated files	13	
H		
HighlightHttpDocs.....	17	
HighlightLink.....	17	
HtmlRemoveScripts.....	17	
HtmlUseTitleAsName	17	
httpd.conf	35	
I		

<i>index (dtSearch Web form variable) ...</i>	15	<i>Search Results Format.....</i>	7
<i>indexes</i>	6	<i>searchFlags (dtSearch Web form variable)</i>	15
<i>Internet Information Server.....</i>	1	<i>SearchForm.....</i>	17
L		<i>SearchIndex</i>	17
<i>Log Settings.....</i>	17	<i>SearchLogItemTemplate</i>	17
<i>LogDocumentAccess.....</i>	17	<i>SearchLogNameTemplate</i>	17
<i>LogSearches.....</i>	17	<i>SearchRequest.....</i>	17
M		<i>searchType (dtSearch Web form variable)</i>	15
<i>maxFiles (dtSearch Web form variable)</i>	15	<i>Security.....</i>	22
<i>MaxUrlSize</i>	17	<i>Selecting Indexes to Include</i>	6
<i>MaxWordsMessage</i>	17	<i>ServerName</i>	35
<i>MaxWordsToRetrieve</i>	17	<i>ServerToLaunch.....</i>	35
<i>MinimizeDialog</i>	35	<i>ShowDialog</i>	35
<i>Modifying a search results item</i>	9	<i>Software Dependencies</i>	34
N		<i>StartPage.....</i>	35
<i>NoFilesMessage</i>	17	T	
O		<i>TechSupportContact</i>	35
<i>OrigSearchForm</i>	13	U	
P		<i>UnconvertedTypes</i>	17
<i>PdfUseTitleAsName</i>	17	<i>UseMicroweb.....</i>	35
R		<i>Using the CD Wizard.....</i>	25
<i>ResultsFooter</i>	17	V	
<i>ResultsHeader</i>	17	<i>Virtual Directories</i>	22
<i>ResultsScript.....</i>	17	<i>virtual directory</i>	22
<i>ResultsTableFooter</i>	17	W	
<i>ResultsTableHeader.....</i>	17	<i>w95ws2setup.exe</i>	34
<i>ResultsTableItem.....</i>	17	<i>Winsock.....</i>	34
<i>returnXml (dtSearch Web form variable)</i>	15	<i>WS2_32.DLL</i>	34
S			