



Xtreme Search Engine Studio Help

© 2007 Xtreme

Table of Contents

Part I Introduction	2
Part II Requirements	4
Part III Features	7
Part IV Quick Start Tutorials	9
1 Steps to create a search engine	9
2 Indexing	10
Indexing documents	10
Indexing online web site	12
Indexing local or network documents	13
Indexing documents from XML file	15
Mapping	16
Indexing document sections	18
3 Choosing appearance	19
Customizing the results page with a template	19
Modifying format of a single result	21
Customizing styles of predefined elements of the results page	22
Using style classes with custom HTML elements	24
Adding document section filter to the search box	26
Conditional displaying items in a result	28
4 Generating search engine	30
Generating CGI based search engine	31
Uploading search engine to an FTP server	32
Part V Tutorials	35
1 Creating a search engine for a website	35
2 Indexing & searching in document sections	38
3 How to create a more precise search	43
Part VI The User Interface	47
1 Main window	47
2 Main menu	48
3 Toolbar	50
4 Add Location dialog	52
5 Online Website (HTTP) Location dialog	53
6 Local/Network Folder Location dialog	54
7 FTP Location dialog	57
8 XML Location dialog	59
9 Assistant dialog	60
10 Indexing Options	60

11	Advanced Options dialog	62
12	Titles Transformation Rules dialog	62
13	Online Documents Mapping dialog	64
14	Documents Mapping dialog	65
15	Keywords dialog	66
16	Re-index URLs (from pattern) dialog	67
17	Re-index URLs (XML file) dialog	68
18	Remove URLs (from pattern) dialog	69
19	Remove URLs (XML file) dialog	70
20	Search Options dialog	71
21	Preview dialog	78
22	Template dialog	78
23	Log window	80
24	Search Box Builder dialog	81
25	Document section filter drop-down dialog	82
26	Document section filter query box (edit) dialog	84
27	Result Page Overview dialog	85
28	Result Formatting dialog	88
29	Password Manager dialog	91
30	Character Translation Table dialog	93
31	Exclude Document Parts Rules dialog	94
32	Proxy Settings dialog	96
33	Log Analyzer utility	98
34	Copy / Paste Output Code dialog	99
35	About dialog	100
Part VII Search syntax		102
Part VIII Logging search queries on server		105
Part IX Search engine types generated by the program		107
Part X Search Engine Studio Service		109
1	Introduction & requirements	109
2	Why should I use the service?	109
3	Installation of the service	110
4	Adding & editing tasks	111
5	Viewing the log	113
6	Tutorial: Automating search engine updates with Search Engine Studio Service	113
7	How to enter the registration key	115

Part XI Advanced topics	118
1 Search script parameters	118
2 Extensions - indexing non-HTML documents	119
3 Automatic updates of search engine databases	121
4 Indexing acceleration by updating only changed documents	122
5 Indexing document date	122
Part XII Customer and Technical Support	124
Index	0

Part



1 Introduction

Welcome to Search Engine Studio!

Search Engine Studio is a program that allows you to easily create a search engine either for your internet or intranet site or an offline search. A **search engine** is an information retrieval system designed to help find information stored on a web site or on a local or network disk. Using a search engine allows to minimize the time required to find the information that people visiting your web site or browsing your CD-ROM/DVD want to find.

Search Engine Studio software enables you to create a web-based search engine for your internet or intranet site (hosted on your server) or an offline search for a CD-ROM/DVD distributions. You can use one of four available methods to index an online web site via HTTP or FTP, or documents stored on your local or network disk. A variety of documents can be indexed, including static and dynamic HTML pages, PDF, MS Office (Word, Excel, PowerPoint), OpenOffice, CHM, Media & Flash files, archives and many other. The generated search engine doesn't require any installation on the server and can be used with most systems, like Windows (IIS Server), Unix (Apache Server on Sun SPARC, Linux, FreeBSD) or MacOS (WebSTAR or Apache).

How to get started

- The easiest way to get started with Search Engine Studio is using the [program assistant](#) that guides you through the process of creating a search engine and inserting it into either an existing frameset configuration or template (e.g. a DreamWeaver template). It is recommended that you use this assistant if you are not yet familiar with the program.
- You should also visit the [Quick Start Tutorials](#) section for some brief tutorials that will familiarize you with basic concepts of the program.
- Study [The User Interface](#) for an introduction to the program.
- For more samples go to the [Tutorials](#) section.

Part



2 Requirements

Below there is a list of what you need to install and use the Search Engine Studio application:

- One of the operating systems listed below:
 - Windows 98/ME
 - Windows NT 4.0 Service Pack 6 or later
 - Windows 2000/XP
 - Windows 2003
 - Windows Vista
- At least 100 MB of RAM memory available for the application. Note that the amount of required memory depends on the number and size of the documents you are indexing.
- Enough free disk space for storing temporary search data during indexing and saving/loading Search Engine Studio documents. Required disk space depends on a number and size of the documents that you are indexing. See the table below for sample evaluation of free disk space requirements:

Number and size of documents (HTML files)	Disk space required for indexing and saving/loading
2000 documents, 22 KB each	at least 350 MB
10,000 documents, 22 KB each	at least 2 GB
50,000 documents, 22 KB each	at least 8 GB
100,000 documents, 22 KB each	at least 16 GB

- Connection to the Internet for indexing online documents and automatic uploading of search engine files.
- Processor, memory, hard disk speed: the faster the computer the less time the indexing process takes. See the table below for sample indexing times. Note that these samples were made on locally stored documents (when indexing online web site downloading time also matters).

The test were made on Intel Core 2 Duo, 2.4 MHz machine, with 1 GB RAM, hard disk 7200 rpm.

Number and size of documents (HTML files)	Indexing time
2000 documents, 22 KB each	about 3 minutes
5000 documents, 22 KB each	7 minutes
10,000 documents, 22 KB each	15 minutes
50,000 documents, 22 KB each	1 hour 45 minutes
100,000 documents, 22 KB each	3 hours 50 minutes

Below is the list of requirements for running the search engine generated by Search Engine Studio:

- For PHP based search engine: a server supporting PHP version 4 or higher.
- For CGI based search engine: a server supporting BINARY CGI scripts.

-
- For Java servlet based search engine: a Java Web Server supporting servlets.
 - For an offline search: the end-user of the offline search engine needs a browser supporting either ActiveX (IE on Windows) or Java Environment installed on their machine.

Part



3 Features

Search Engine Studio Features

- works with any existing hosting solution
- generates search for a wide range of targets: highly optimized binary CGI scripts for Linux / FreeBSD / Sun SPARC / Windows / MacOS servers, PHP script, Java servlet, or an ActiveX/Java-based version for offline search (CD-ROM/DVD distributions)
- compatibility with all browsers on the client side
- automatic indexing of static & dynamic HTML pages (e.g. PHP, ASP, JSP, ColdFusion), Adobe Acrobat (PDF) documents, Microsoft Office and OpenOffice Documents, CHMs, media files (images, audio and video); also including indexing of password-protected documents
- 4 unique ways of scanning a web site: HTTP crawler / FTP / local+network disk / XML file
- search engine works with frame-based or frameless configurations
- highly customizable graphical appearance (e.g. fonts, colors)
- result listings show the exact location of searched words (marked in bold) within the document's contents
- wildcard search (e.g. the user can search for all words starting with, ending with, or containing a given letter sequence)
- advanced boolean search (using AND, OR, NOT keywords)
- expression search (embedding an expression in "" characters will only search for words grouped together)
- high performance
- you can create as many search engines, for as many web sites as you like, when you purchase a single license
- support for most languages: the search engine is based on the UTF-8 character encoding which means it will work well with non-English alphabets; additional tools are provided for complete support
- search filters: you can set a filter in order to limit the number of search results based on a wide range of criteria (including document date / site section / document type / document section / custom data field)
- FTP upload
- password manager for indexing protected areas

Part



4 Quick Start Tutorials

The tutorials in this section provide a quick introduction to using Search Engine Studio. They are intentionally kept brief so that you can actually start using the program as quickly as possible. The goal is not to teach you every single detail but to familiarize you with the basic principles of the program.

It is a good idea to start with the [Steps to create a search engine](#) topic to learn what you actually need to do to create your search engine.

You can also take a look at the [Indexing](#), [Choosing appearance](#) and [Generating search engine](#) sections to learn more about each phase.

4.1 Steps to create a search engine

In this topic you can get some quick information about what you actually need to do to create a search engine, either for your internet or intranet web site or CD-ROM/DVD.

Generally, to create a search engine you have to go through the three steps presented below:



Indexing documents

Before you let a customer using your search engine find information by running a query, first you have to index your documents. Generally speaking, indexing is the process when Search Engine Studio gathers information about words retrieved from documents and stores it in the database that next can be used to perform a search. There are four ways of indexing available in the program: you can choose to index an online web site via HTTP or FTP protocol, local or network files or use an XML file containing a list of document paths.

You can learn more about indexing in the tutorials available in the [Indexing](#) section.

Choosing appearance

Once you are finished indexing your documents you can work on your search layout. The easiest way to customize the appearance of the results page that lists documents matching the query is using a template file. To create the template you can use one of your web pages by removing its contents and leaving just the common header, menu, graphical frame etc. This way the results list will look like other pages of your site. Additionally, you can also customize the appearance and functionality of the search box. For example, you can add a filter in a form of a drop-down list that allows searching only in a title or a chosen meta tag of a document, or a filter that limits the search scope to documents of the

selected type, like PDF or MS Office.

You can learn more customizing appearance it in the [Choosing appearance](#) tutorials section.

Generating

When your search engine design is ready you can finally generate it. In this step you need to choose the type of your search engine: you can select PHP or CGI scripts that are usually used for internet or intranet web sites or an offline ActiveX/Java based search for CD-ROM/DVD. You also have to specify a few output options, like the location where the search files are generated. Once you are ready you just need to start generation by clicking the icon in the [program toolbar](#) or by choosing **Generate Search Engine** from the **Search Engine** menu.

You can learn more about it in the [Generating search engine](#) section.

4.2 Indexing

In this section you will find quick guides giving you a brief but still very useful information about the indexing process.

Expand the *Indexing* branch in the panel on the left and choose the topic you want to read.

- To get some general information about adding document locations go to the [Indexing documents](#) topic.
- To read more about selecting sections of a document that you want to index go to the [Indexing document sections](#) topic.

4.2.1 Indexing documents

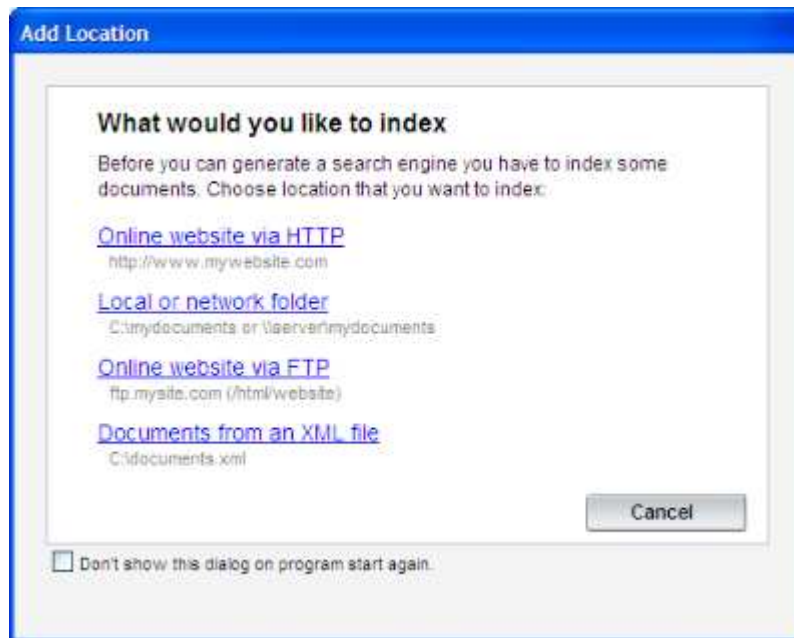
The first step of creating a search engine is indexing documents. Once a document is indexed a visitor to your search engine will be able to find it by typing the appropriate keywords in the query box. In other words, **indexing is the process that retrieves from documents information about their contents and stores it in the search database so it can be later used for searching.**

Before you can run the indexing process you have to specify a location of the documents that you want to include into your search. Search Engine Studio offers four ways of indexing that you can use:

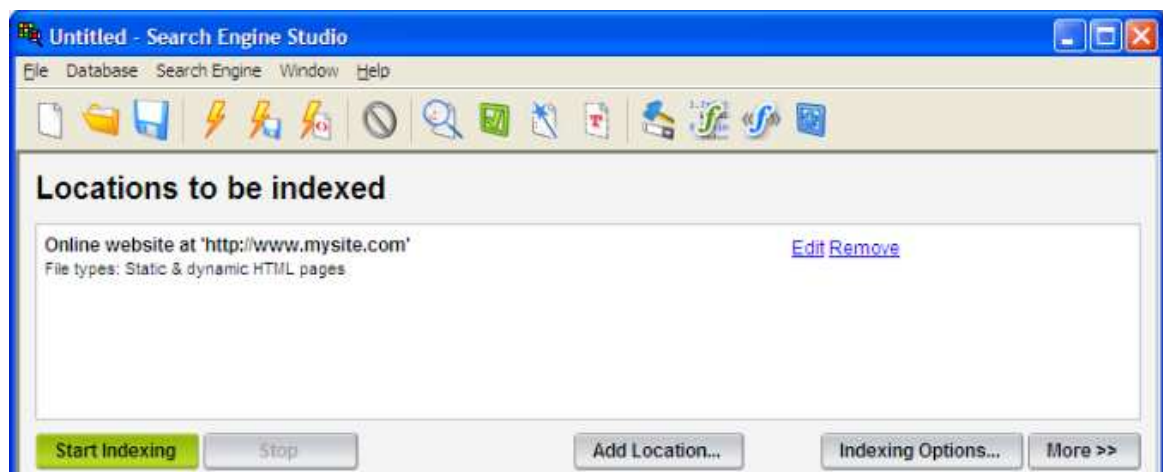
- [Online web site via HTTP](#) that can be used to index an online web site. You just need to type a start URL, for example <http://www.mysite.com>, and the program automatically spiders through the whole site and indexes its pages.
- [Local or network folder](#) that allows you to index documents on your local or network disk.
- [Online web site via FTP](#) that uses FTP protocol to index documents stored on your server.
- [Documents from an XML file](#) that takes a list of documents for indexing from an XML.

To add a location and start indexing follow the steps below:

1. Click the **Add Location** button in the [main program window](#).
2. Select one of the available location types in the [Add Location dialog](#).



3. In the window that opens for a chosen option you can type the URL or folder path that you want to index and set more options that influence the indexing process. See the [Indexing online web site](#), [Indexing local or network documents](#) or [Indexing documents from XML file](#) for more information.
4. Once you are ready click **OK**.
5. The location entry appears in the **Locations to be indexed** panel of the main window (see the picture below).



6. When you are done you can run indexing by clicking the **Start Indexing** button that you can see in the picture above. You can follow the progress of indexing documents in the [Log Window dialog](#) that opens automatically.

Note that Search Engine Studio also allows you to choose which sections of a document will be

actually indexed. For example, you can choose to index document's content, title or a custom meta tag. You can learn more about it in [Indexing document sections](#).

See also:

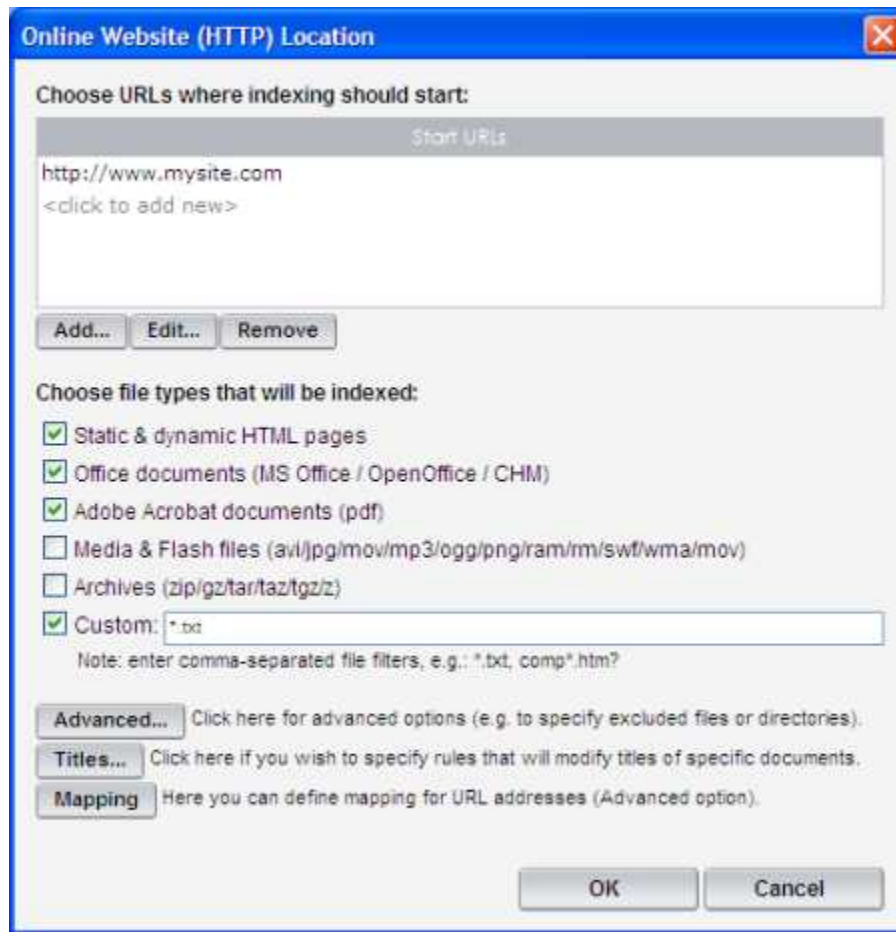
[Indexing online web site](#)
[Indexing local or network documents](#)
[Indexing document sections](#)

4.2.2 Indexing online web site

If you want to index a web site that is already running on your server you should use the **Online website via HTTP** location.

To index an online web site follow the steps below:

1. Click the **Add Location** button in the [main program window](#).
2. Choose the **Online website via HTTP** option in the [Add Location dialog](#) that opens. The [Online Website \(HTTP\) Location](#) window appears (see below).



3. In the **Start URLs** panel at the top of the window add one or more entry URLs to your web site where crawling should start: click the **Add...** button, and type URL, e.g. <http://www.mysite.com>, and click **OK**. The URL appears in the panel.
4. In the **Choose file types...** section in the middle of the window specify the document types that you want to index.
5. (Optionally) If you want to customize the indexing process even more you can click the **Advanced** button to go to the [Advanced Options dialog](#). In this dialog you can set up many indexing options, like choose to index files only up to the chosen level in the hierarchy, exclude from indexing documents based on a URL or title or turn on HTML form submission which is especially useful when indexing password-protected pages.
6. (Optionally) If you want to set up title transformation rules, click the **Titles...** button that opens the [Titles Transformation Rules dialog](#). Modifying titles can be useful when your documents' titles always contain a common part, e.g. company name, which you don't want to be displayed in the search results.
7. (Advanced) You can also use the **Mapping** button to add mapping rules that will be applied to URLs being indexed. Note, that in most cases setting up mapping is not needed for an online web site. You can read more about this in the [Mapping](#) topic.
8. Once you are ready click the **OK** button.
9. The newly added location entry appears in the **Locations to be indexed** panel of the [main window](#).
10. Now, you can run indexing by clicking the **Start Indexing** button in the main window. You can follow the progress of indexing documents in the [Log Window dialog](#) that should open automatically.

Useful information:

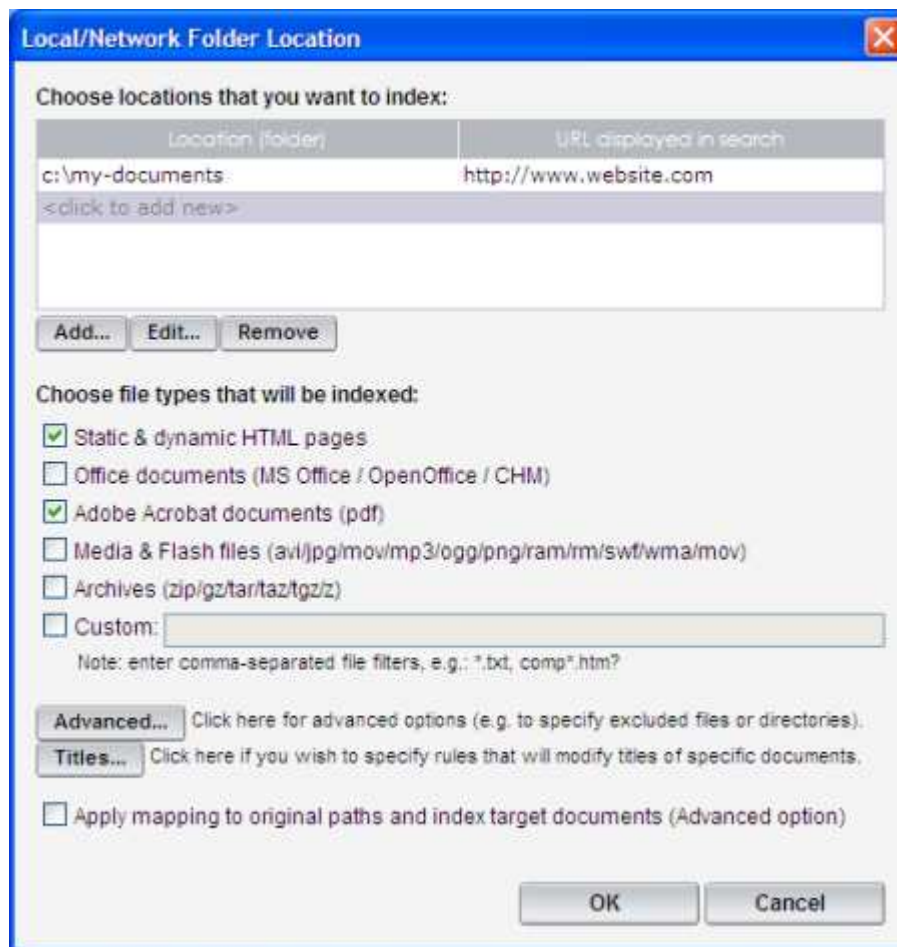
- When you are using the **Online website via HTTP** indexing method, Search Engine Studio crawls your site using the HTTP protocol. It opens the entry URL, indexes it and retrieves all ULRs that point to other documents of your web site. Next the program does the same for those documents and repeats that until the last page of your web site is indexed. Note that only those pages that match the specified indexing rules, like document type or URL filter, will be indexed.
- This indexing method uses HTTP protocol which allows indexing dynamic web pages, like PHP, ASP, JSP or ColdFusion, so they can be processed by a hosting server first.

4.2.3 Indexing local or network documents

If you want to index documents located on your local or network disk you should use the **Local or network folder** location.

To index local or network files follow the steps below:

1. Click the **Add Location** button in the [main program window](#).
2. Choose the **Local or network folder** option in the [Add Location dialog](#) that opens. The [Local/Network Folder Location](#) window appears (see below).



3. In the panel at the top of the window you can add folders that you want to index. Click the **Add...** button and in the dialog that opens type a path to a local or network folder. In most cases you should also type a URL that will be shown in search results instead of the local path. For example, instead of a `c:\my-documents\products.html` local path the <http://www.website.com/products.html> URL will be shown. You can read more about it in [Mapping](#).
4. In the **Choose file types...** section in the middle of the window specify the document types that you want to index.
5. (Optionally) If you want to customize the indexing process even more you can use the **Advanced** button to go to the [Advanced Options dialog](#). In this dialog you can set up many indexing options, like choose to index files only up to the chosen level in the hierarchy or exclude from indexing documents based on a path or title.
6. (Optionally) If you want to set up title transformation rules, click the **Titles...** button that opens the [Titles Transformation Rules dialog](#). Modifying titles can be useful when your documents' titles always contain a common part, e.g. company name, which you don't want to be displayed in all search results.
7. (Advanced) Usually you leave the **Apply mapping to original paths...** option off. This option allows you to index target/online documents instead of their local copies (see [Useful information](#) below).
8. Once you are ready click **OK**.

- The newly added location entry appears in the **Locations to be indexed** panel of the [main window](#).
- You can start the indexing process by clicking the **Start Indexing** button in the main window. You can follow the progress of indexing documents in the [Log Window dialog](#) that should open automatically.

Useful information:

- When you use **Local or network folder** Search Engine Studio indexes each document found in the specified folder or folders that match the indexing rules, like document type or path / title filter.
- The **Apply mapping to original paths...** option is an advanced option you can use when you want to index target/online documents instead of their local copies. For example, when you are indexing a 'c:\documents' folder and define '<http://www.website.com>' as the mapping then once a 'c:\documents\index.php' file is found in this folder mapping is immediately applied and the target '<http://website.com/index.php>' file is indexed instead. This option is useful when your documents contain dynamic web pages that should be indexed rather through the HTTP protocol than locally.

4.2.4 Indexing documents from XML file

Sometimes, instead of indexing the whole web site or a local folder, you want to index only some selected document. In this case you can use the **Documents from an XML file** location. You just need to create an XML file listing paths of documents that you want to index (see **Useful Information** below to learn about the format of this XML).

To index documents from an XML file follow the steps below:

- Click the **Add Location** button in the [main program window](#).
- Choose the **Documents from an XML file** option in the [Add Location dialog](#) that opens. The [XML Location](#) window appears (see below).



- In the **Choose XML file with path listing** field type a path to the XML file that contains paths of

the documents that you want to index. You can also use the **Browse...** button to pick a file from a disk. You can learn more about the format of this XML in **Useful information** below.

4. If you have a different format of XML data, you can easily convert it to the desired format by specifying an XSLT transformation file. In this case you just need to turn on the **Convert from a different XML...** option and choose a path to a file that contains the XSLT transformation.
5. (Optionally) If you want to set up title transformation rules, click the **Titles...** button that opens the [Titles Transformation Rules dialog](#). Modifying titles can be useful when your documents' titles always contain a common part, e.g. company name, which you don't want to be displayed in all search results.
6. (Optionally) If you need to map a ULRs or a paths of documents from the XML file you can do this by clicking the **Mapping** button. This is especially useful when you are indexing local files and want to display online URL instead of the local one in the search results. You can read more about this in the [Mapping](#) topic.
7. Once you are ready click **OK**.
8. The newly added location entry appears in the **Locations to be indexed** panel of the [main window](#).
9. You can start the indexing process by clicking the **Start Indexing** button in the main window. You can follow the progress of indexing documents in the [Log Window dialog](#) that should open automatically

Useful information:

The XML file that stores paths of documents form indexing should conform to the following DTD (document type definition) format:

```
<!DOCTYPE sitemap [  
  <!ELEMENT sitemap (node+)>  
  <!ELEMENT node (node*|text)*>  
  <!ATTLIST node href CDATA #IMPLIED>  

```

Here's a sample XML file that can be used:

```
<sitemap>  
<node href="http://www.xtreeme.com/">  
<node href="http://www.xtreeme.com/sitexpert/">  
<node href="http://www.xtreeme.com/search-engine-studio/">  
<node href="http://www.xtreeme.com/dhtml/">  

```

4.2.5 Mapping

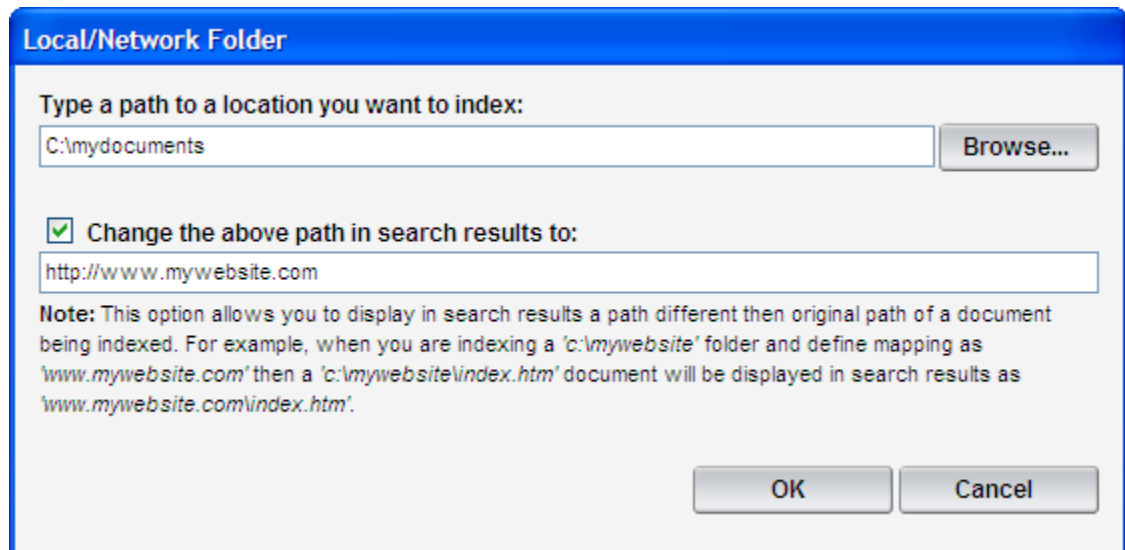
When a user performs a search and a list of matching documents is presented, one of the most important elements of the list is the URL or path to the target document of each matching result. When you directly index an online web site this is simply a URL of the document that was indexed, for example <http://www.mysite.com/index.html> or <http://www.mysite.com/products/features.html>. Once a user clicks this link in the results the target document opens.

On the other hand, when you are indexing local documents or files on your FTP server, then the local

or FTP path of a document will not work in the same way. For example, when a file at `c:\mysite\index.html` is indexed showing this path in the results usually doesn't make much sense, especially when the search engine is published online. Instead of that, you would probably want to display an online equivalent of a local or FTP path, which is pointing to the online copy of the document so it can be opened by a customer using your search. In this case a mechanism called **mapping** can be used. Mapping allows you to replace an original path of a document with another one, like online URL or UNC path. For example, you can choose to replace `c:\mysite` with <http://mysite.com/>. As the result the local document mentioned above will be displayed in the search results as <http://mysite.com/index.html>.

Setting up mapping for **Local or network folder** and **Online website via FTP** locations.

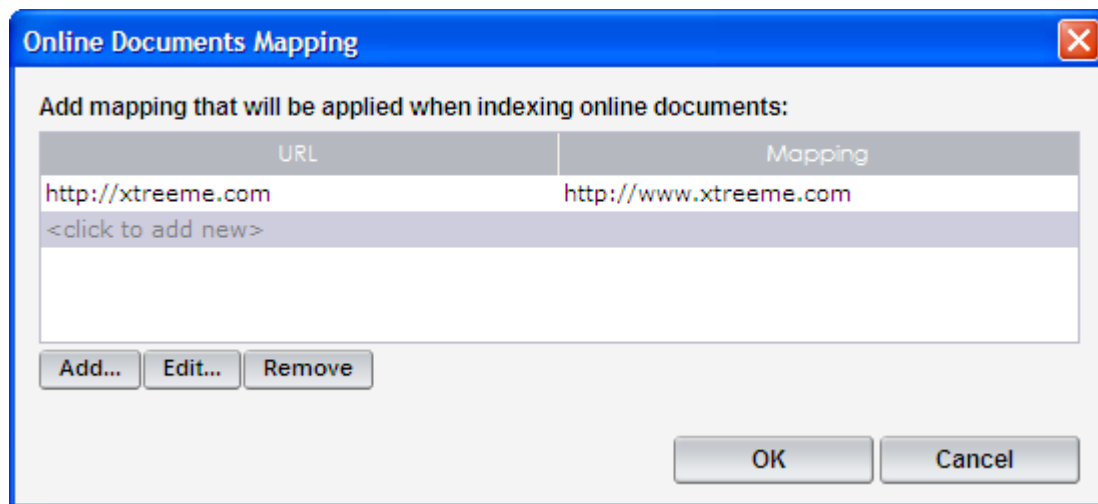
When you are using **Local or network folder** or **Online website via FTP** location you can set up mapping when you add a new local or remote folder (see the picture below).



Simply turn on the **Change the above path in search results to** option and type a path or URL that will be shown in the results instead of the original one. In the sample above a `c:\mydocuments` path will be replaced with a <http://www.mywebsite.com> URL.

Setting up mapping for **Online website via HTTP** location.

In most cases there is no need to set up mapping when indexing an online web site. On the other hand, setting up mapping for online documents can be sometimes useful when you are using two kinds of links on your pages, for example <http://xtreeme.com> and <http://www.xtreeme.com>. Because Search Engine Studio treats these two URLs as completely different it may occur that the same document will be indexed twice. For example, if you have a `products.htm` page and you link to it from other pages as <http://xtreeme.com/products.htm> and <http://www.xtreeme.com/products.htm> then this page will appear twice in the search database, even if both the URLs are pointing to the same file. You can change this behaviour by defining an online mapping so one of the URLs above is mapped into the other one (see the picture below). This way the <http://xtreeme.com/products.htm> and <http://www.xtreeme.com/products.htm> URLs will be treated as exactly the same, so the resulting page is indexed only once.



Setting up mapping for **Documents from an XML file** location.

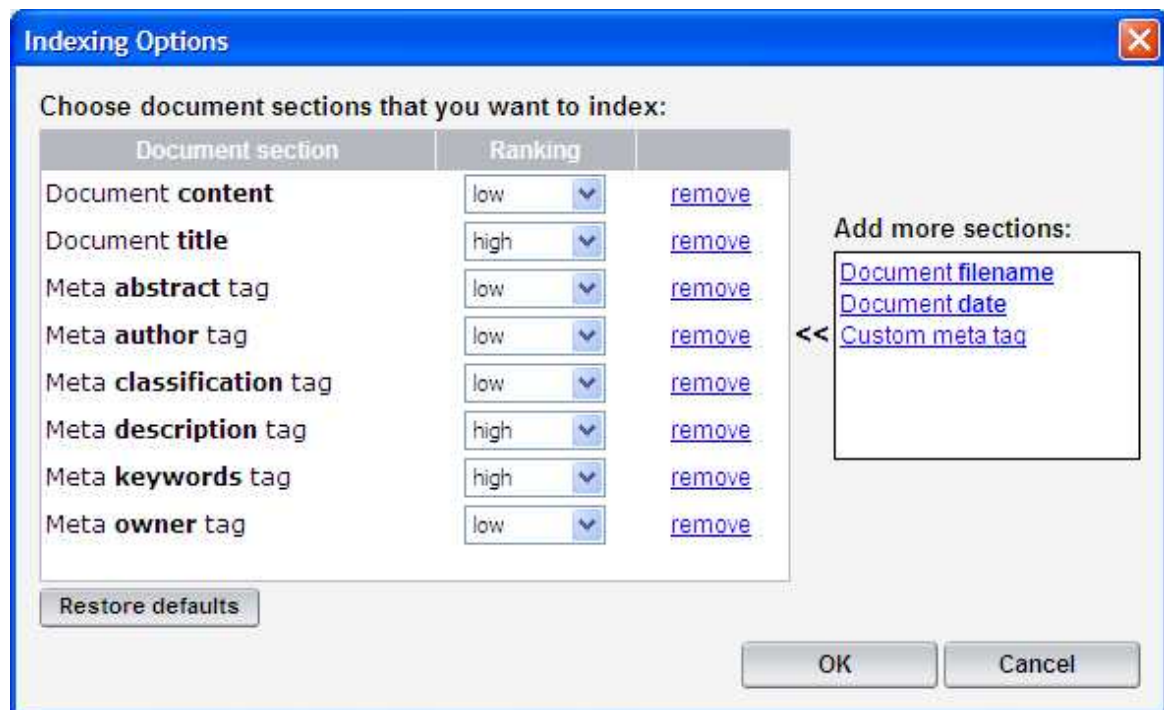
In the **XML Location** dialog you can set up mapping by clicking the **Mapping** button at the bottom of the window. In the dialog that opens you can add a base path or URL that should be replaced and type a path that you want to display instead.

4.2.6 Indexing document sections

Generally, an HTML document consists of a number of data fields or sections. A document's body tag stores the main content while a number of header tags contain information like title, keywords, description, author etc. Usually, most header data is stored in a form of <meta> tags (see the code sample below):

```
<html>
<head>
  <title>Title of page</title>
  <meta name="description" content="This is a sample HTML page.">
  <meta name="keywords" content="sample, HTML, document, meta, body">
  <meta name="author" content="John">
</head>
<body>
  This is the main content of the document.
</body>
</html>
```

In Search Engine Studio you can control which sections of a document are actually indexed and included into the search database. You can choose this in the [Indexing Options](#) dialog that you open by clicking the **Indexing Options...** button in the [main window](#) of the program. In most cases you want to index the main document content, title and some standard meta tags, like author or description. Additionally, you can also choose to indexing custom meta tags that your documents contain.



The document sections you have chosen for indexing can be later used to allow the end-user to limit the scope of the query. For example, you can let a user choose to search only in a document's title or an author meta tag. This way the search will be more precise and return better matching results. You can learn more about this in the [Indexing & searching in document sections](#), [Adding document section filter to the search box](#) and [How to make a search more precise](#) topics.

4.3 Choosing appearance

In this section you will find quick guides giving you a brief but still very useful tips about choosing appearance of your search box and the search results page.

Expand the *Choosing appearance* branch in the panel on the left and choose the topic you want to read more about.

4.3.1 Customizing the results page with a template

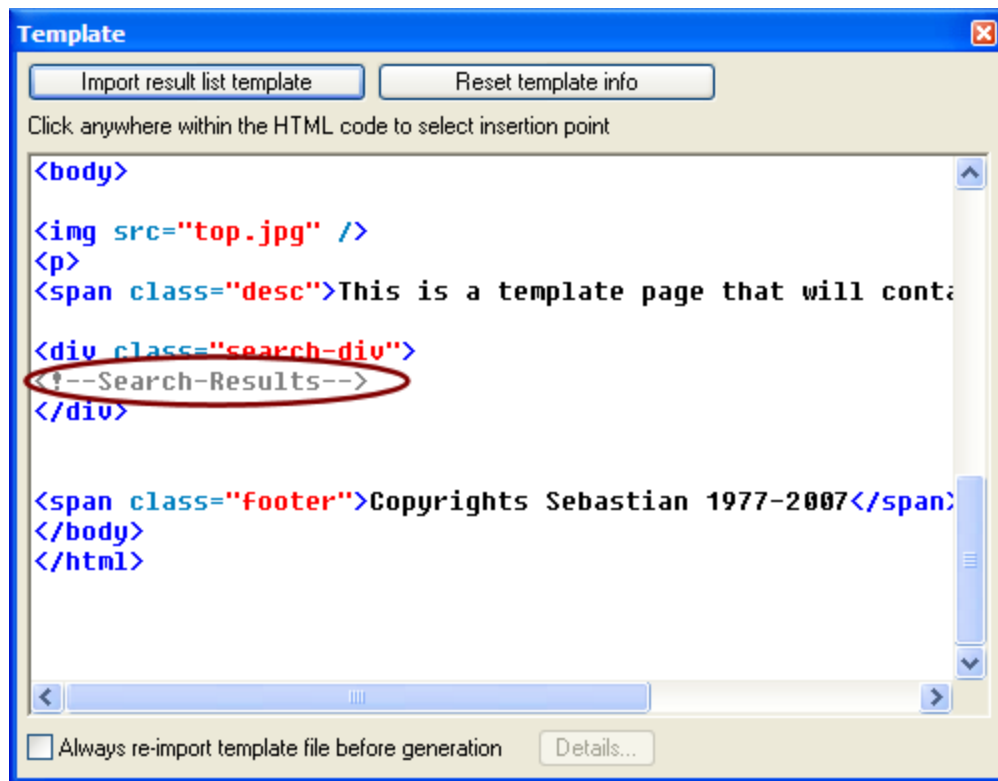
In this tutorial you will learn how to customize the layout of your search results page with a template file.

Note: The template is usually an HTML page that contains a graphical elements that you want to appear on the results page, like graphical header, footer, company logo, navigation menu etc.

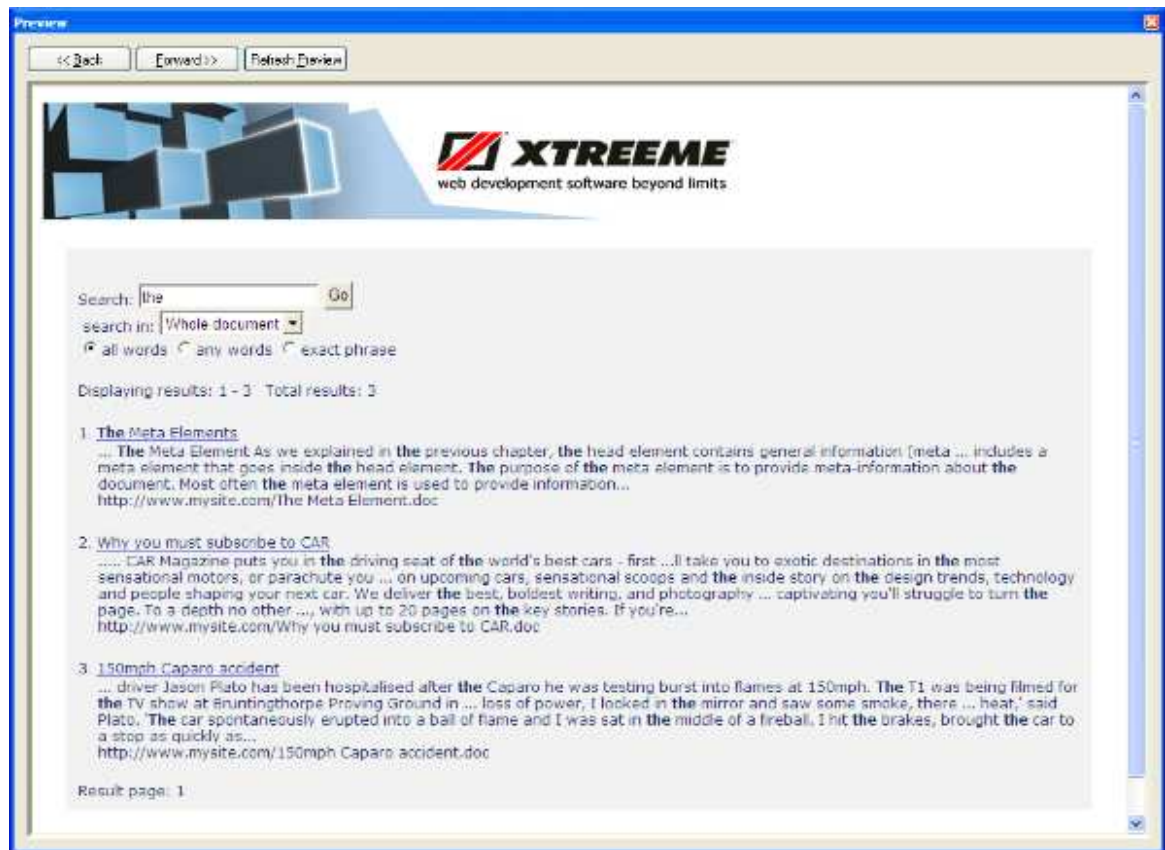
Follow the steps below to change the default appearance of the results page so it looks like other pages of your web site:

1. Open the **Template** dialog by clicking the  icon in the program toolbar or choosing the *Template* option in the *Windows* menu.

2. Click the **Import result list template** button and browse for the template file that you want to use to customize your results page. HTML code of the selected file appears in the main window of the **Template** dialog.
3. Now, scroll to the part of the code where you want to insert search results and click there with the left button of your mouse. The place is marked with the `<!--Search-Results-->` tag (see the picture below).



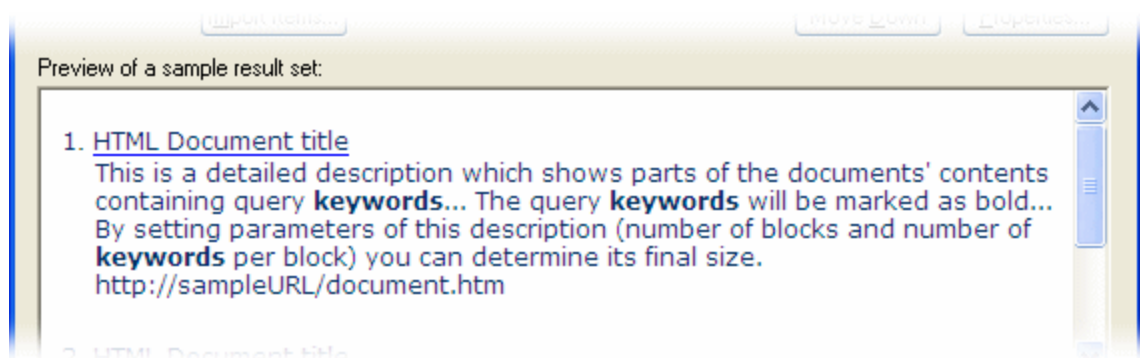
4. Close the **Template** dialog, open the preview window and run a sample query so display the results page (see the picture below). Note that if you are using in your template any external resources, like CSS Stylesheets file or images, they may be not displayed in this preview. On the other hand, once you generate your search engine and upload it to the final location where these external resource files are available, they will be displayed properly.



4.3.2 Modifying format of a single result


In this tutorial you will learn how to modify a format of a single result that is shown on the results page when a customer using your search engine runs a query.

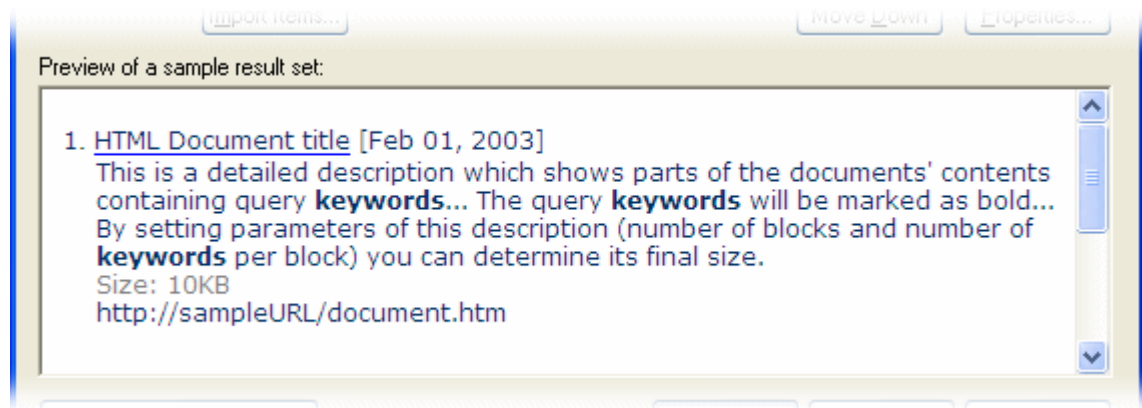
By default each matching result displays an order number, title, document's content preview and document URL (see the picture below).



Let's say that you want to change this format and show a document date after the title and a document size before the URL.

Follow the steps below:

1. Open the **Result Formatting** dialog by clicking the  icon in the [toolbar](#) or choosing the option from the **Search Engine** menu.
2. Select the **Document title** item in the **Used items** panel on the right. Add a **Custom HTML code** element after the title by double-clicking it in the **Available items** panel on the left and type ` ` in the HTML code section.
3. Next, add a **Document date** item and another **Custom HTML code** by this time type the `]` character in the code section.
4. This way you have displayed a document date after the title. At any time you can see the changes you make in the **Preview** panel at the bottom of the dialog (see the picture below).
5. Now, after the **Line break** item following **Document's content preview** add **Custom HTML code** and type `Size: `
6. Next, add a **Document size** element and one more **Line break**.
7. Now, you can take a look at the preview panel to see the result (see the picture below). When you generate your search engine and run a query the displayed results will have the same format as presented on this preview.



If you want to learn how to customize a style of the newly added elements see the [Customizing styles of predefined elements of the results page](#) and [Using style classes with custom HTML elements](#) topic.

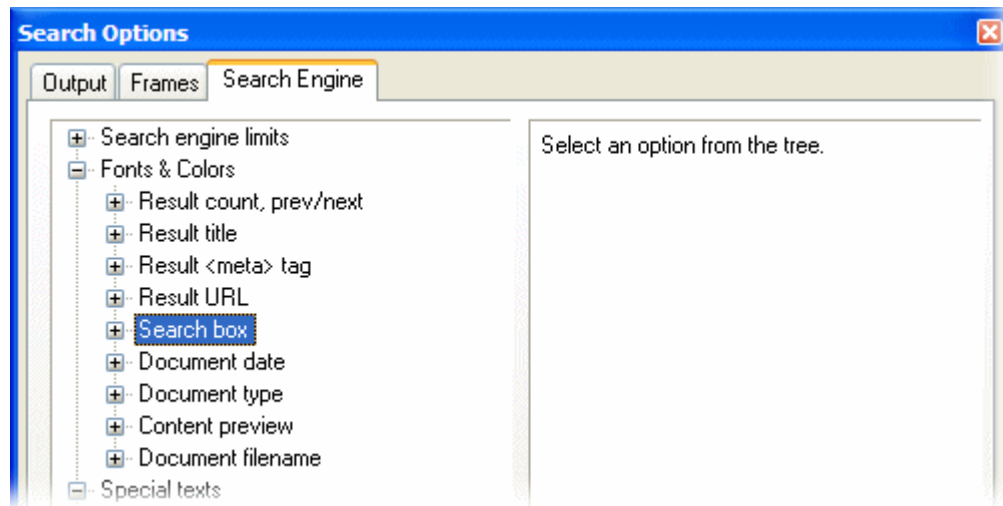
4.3.3 Customizing styles of predefined elements of the results page

In this tutorial you will learn how to customize appearance of the predefined elements of the search results page.

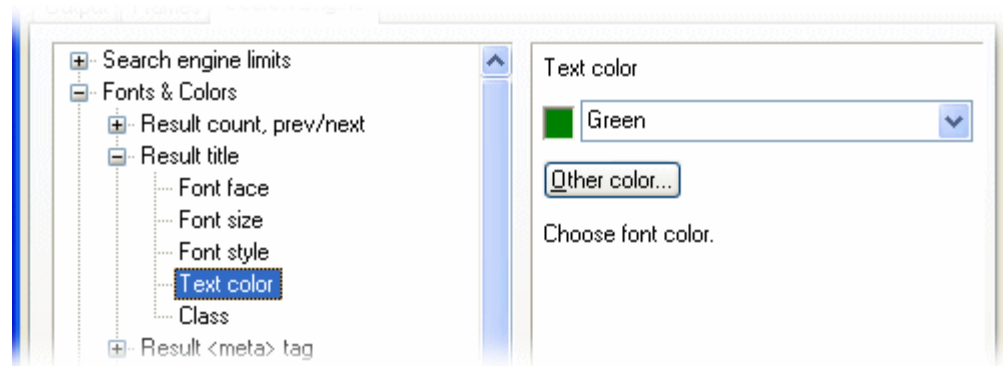
Let's that you want to modify style of the **Document title** and **Document's content preview** elements on the results page.

Follow the steps below:

1. Open the [Search Options](#) dialog, choose the **Search Engine** tab and expand the **Fonts & Colors** branch. As you can see in the picture below this is the place where you can find style options for the predefined result formatting items like *Document title*, *Result URL* or *Document's content preview*.



2. For the sake of the example, let's make a result title green and bold. Expand the **Result title** item and next set the **Font style** options to **Bold** and in **Text color** select **Green** from the drop-down list.



3. Let's also make the preview of a result's content gray. Expand the **Content preview** item and select **Gray** as the text color.
4. Now, open the preview dialog and run a sample search to get some results displayed (this will work assuming that you have some documents indexed). In the picture below you can see the style changes you have made.



See also:

To learn how you can change style of a custom HTML text you can also use for result formatting see [Using style classes with custom HTML elements](#).


Visit the [Modifying format of a single result](#) topic to learn how you can customize a format of a single result on the results page.

4.3.4 Using style classes with custom HTML elements

In this tutorial you will learn how to modify style of custom HTML text that you can use in the result formatting.

Let's say that you want to display for a single result a custom text, for example *Author:*, that will be followed by a content of the *Author* meta tag. You also want to customize the style of this text by specifying some font, size and color options.

Follow the steps below:

1. Open the [Result Formatting](#) dialog by clicking the  icon in the [toolbar](#) or choosing the option from the **Search Engine** menu.
2. Add a **Custom HTML code** item after the **Custom HTML code [<dd>]** element in the **Used items** panel on the right and type in the code panel of the window that opens the following HTML code:

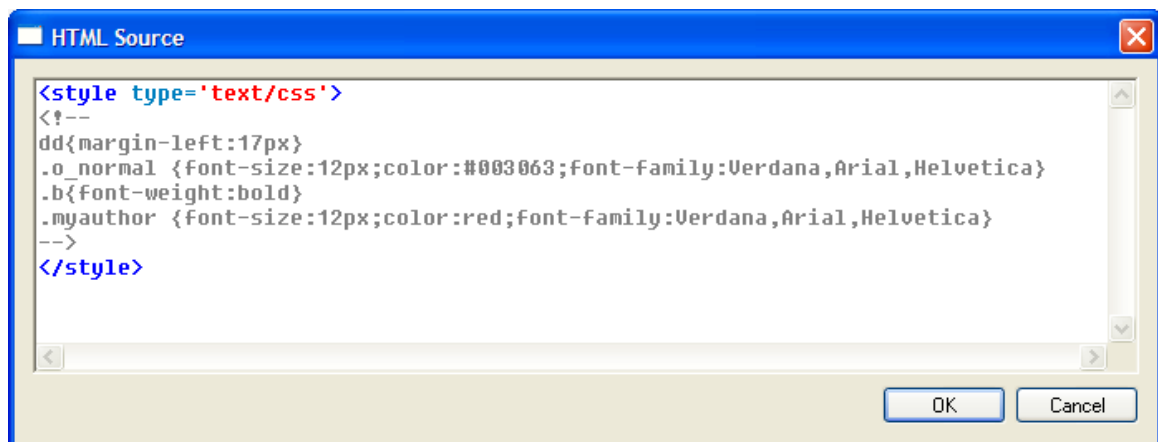
```
<span class="myauthor">Author:</span>
```

This way you have added a text formatted with the *myauthor* class.

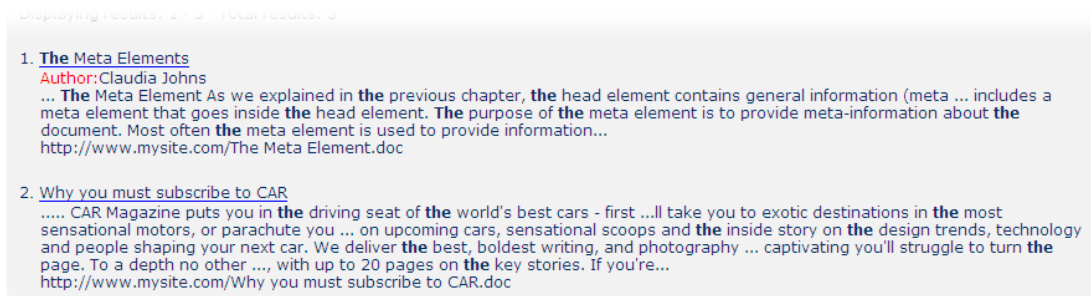
3. Next, add a **Custom meta tag** element, select *Meta author tag* from the available options in the window that opens and click **OK**. Additionally, add a **Line break** item.
4. As you can in the preview panel at the bottom of the dialog the *Author* text following by a content of an author meta tag is displayed after the document title.
5. Click **OK** to close the **Result Formatting** dialog.
6. Open the [Result Page Overview](#) window.
7. Double-click the first **Custom HTML code** element in the **Used items** panel on the right to open it. Assuming you have left the default configuration of the items this element should contain style definitions.
8. After the last style inside the tag add the following definition of the *myauthor* class that was used above. For the sake of the example, we have used a red font color:

```
.myauthor {font-size:12px;color:red;font-family:Verdana,Arial,Helvetica}
```

The code of the style tag should look more less like in the picture below.



9. Click **OK** to close the properties and again to close the **Results Page Overview** window.
10. Go to the preview dialog and run a sample search to display some results and see the newly defined style.



See also:

To learn about customizing predefined elements of the Result Formatting dialog see [Customizing styles of predefined elements of the results page](#).

Visit the [Modifying format of a single result](#) topic to learn how you can customize a format of a single result on the results page.


4.3.5 Adding document section filter to the search box

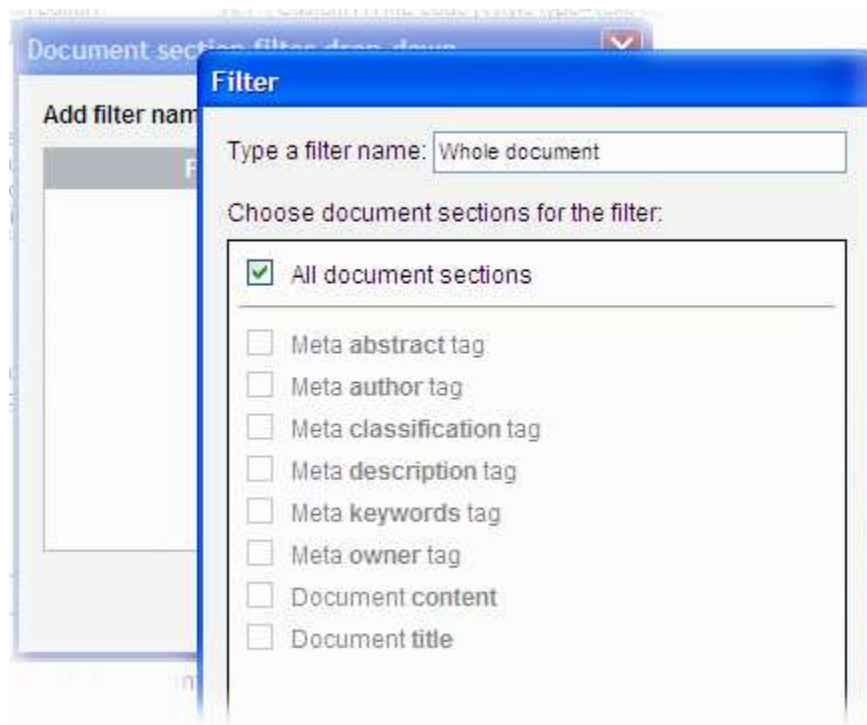
In this tutorial you will learn how to customize the search box by adding a filter that will allow customer using your search engine to limit the search scope to a part of a document, for example document main content, a title or a custom meta tag.

Let's say the HTML documents that you have indexed, apart from the the obvious body tag, also contain a number of header tags that are used to better describe a content of each web page. Note that in Search Engine Studio you can choose which sections of a document will be indexed (see [Indexing document sections](#)). Let's assume that the format of your document header looks like in the following sample:

```
...
<head>
  <title>Title of page.</title>
  <meta name="description" content="This is a sample HTML page.">
  <meta name="keywords" content="sample, HTML, document, meta, body">
  <meta name="author" content="...author of a document...">
</head>
...
```

Now, we will add a drop-down list that will allow customers to choose the scope of their search:

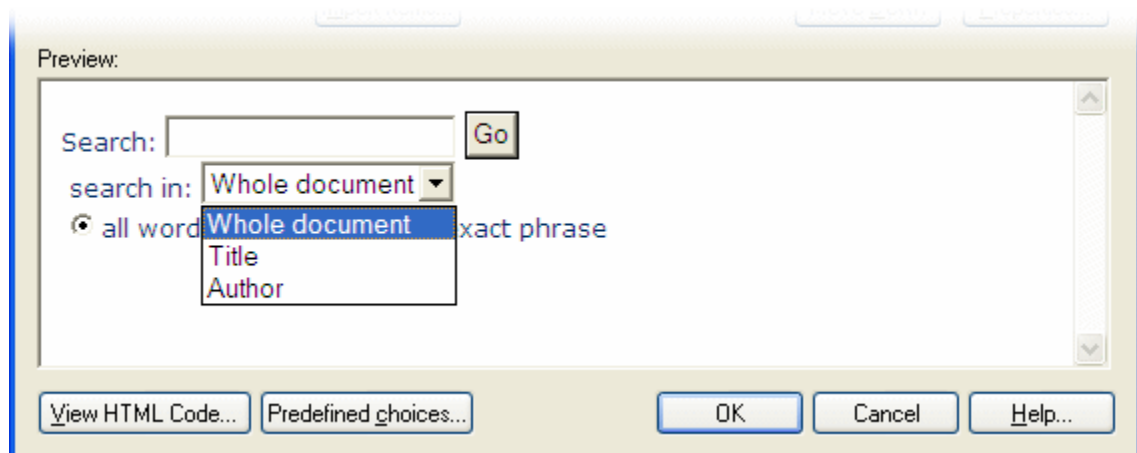
1. Open the **Search Box Builder** dialog by clicking the  icon in the toolbar or the option from the Search Engine menu.
2. After the **Line break** element following **Text-based search button** in the **Search Box Items** section on the right add a new **Custom HTML code** item by double-clicking it in the **Available items** panel on the left.
3. In the **HTML Source** dialog that opens type ` search in: `
4. Next, add a **Document section filter drop-down** item. The **Document section filter drop-down** dialog opens.
5. In the dialog click **Add** to insert a new drop-down entry option. In the **Filter** window that appears type *Whole document* as the filter name, select the *All document sections* option (see the picture below) and click **OK**.



- Repeat the operation to add two more entries. For the first one type a name *Description* and select the *Meta description tag* option and for the second one type *Title* and select *Document title*.

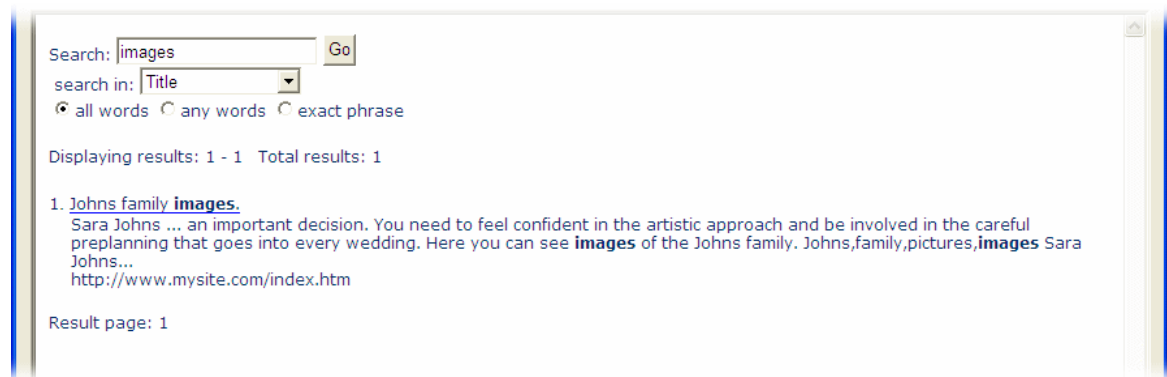
Note: You can select more than one section for a single drop-down option. See the note at the bottom of this page for more info.

- Click **OK** to close the **Document section filter drop-down** dialog. As you can see in the preview panel at the bottom of the **Search Box Builder** window a new drop-down list was added with the three options you have just inserted (see the picture below).



- Close **Search Box Builder** and go to the preview window.

Now you can test the functionality of the drop-down. If you want to find documents that contain the keywords you are looking for wherever in their contents you need to choose the *Whole document* option from the drop-down. On the other hand, if you want to list only those documents that contain these keywords only in their title tag, you should choose *Title* and run the search again (see the picture below).




This way customer using your search engine can search for the information they are looking for in a more precise way.

Note: When you are adding a new drop-down filter option, in the **Filter** dialog you can select more than one section. For example, you can select meta *description* and *comments* tag and type *Page descriptions* as the name of this option. When a user selects it and performs a search then the search scope will be limited to these two meta fields.

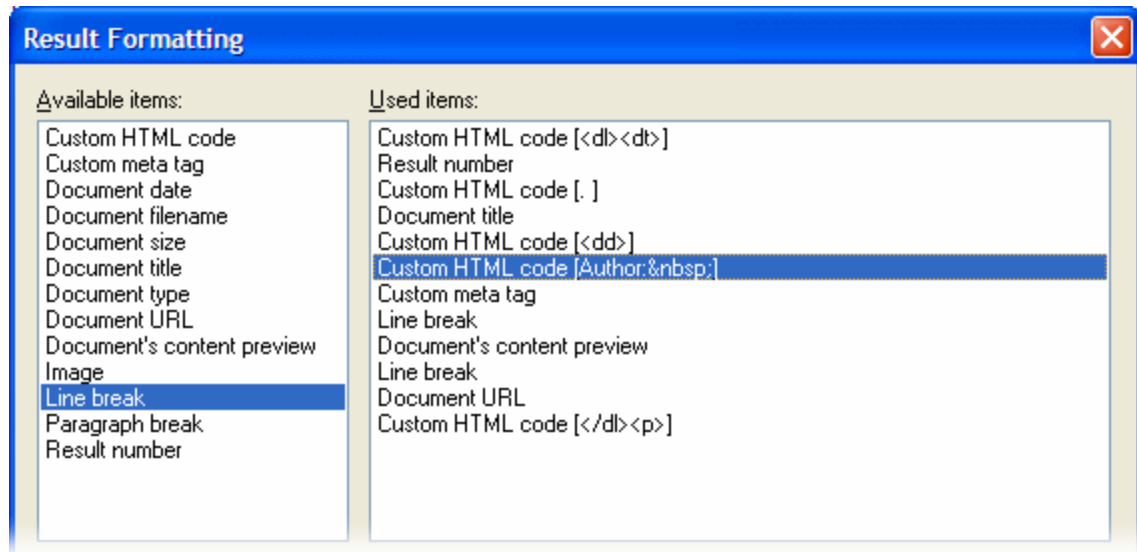
4.3.6 Conditional displaying items in a result

In this tutorial you will learn how to use conditions to customize a format of a single result that is shown of the results page.

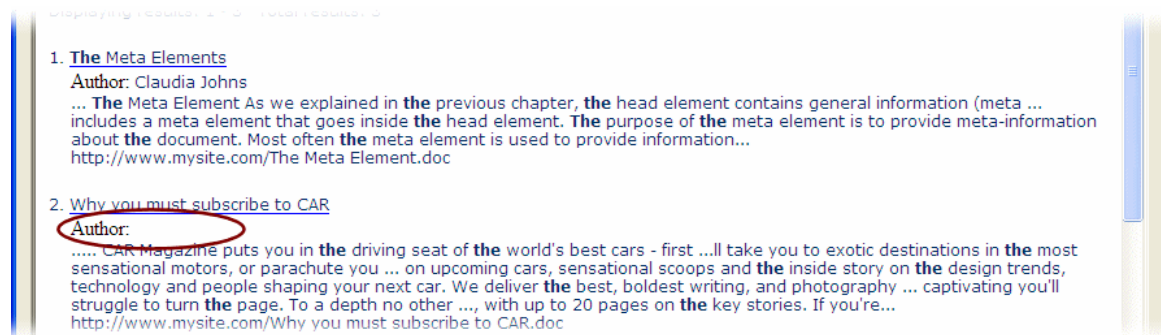
Let's say you have indexed some MS Word documents, including their author tags (you can read more about indexing chosen parts of a document in the [Indexing document sections](#) topic). Now we will customize a format of the result so that the author of a document is also displayed:

- Open the [Result Formatting](#) dialog by clicking the  icon in the [toolbar](#) or choosing the option from the **Search Engine** menu.
- Select a **Custom HTML code [<dd>]** element in the **Used items** panel that should be right after the **Document title** item (this is assuming that you have left the default items configuration).
- Double-click a **Custom HTML code** item from **Available items** on the left to add a new one. In the **HTML code** section of the **Result Formatting Item Properties** dialog that opens type: `Author: ` and click **OK**.
- Now, add a **Custom meta tag** element. In the windows that opens choose *Meta author tag* from the available options and click **OK**. The element should appear in **Used items** right after the previously added **Custom HTML code[Author:]** item.
- Additionally, add the **Line break** item to move the rest of the content on line down.

6. The items in the **Result Formatting** dialog should look more less like on the picture below. At any time you can view the current appearance of the result in the preview panel at the bottom of the window.



7. Now we can see how it works in a sample search. Click **OK** to close the dialog, open the preview window and run a sample query, for example *the*, so a few documents are displayed in the results list (see the picture below as a sample).



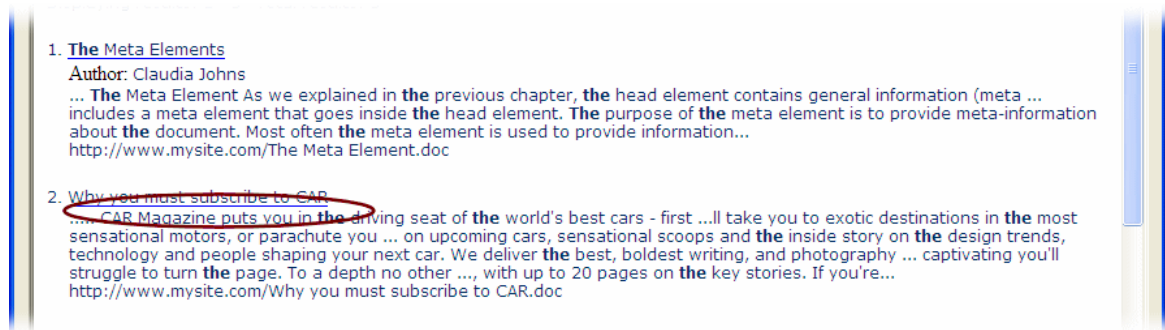
As you can see in the picture above it may happen, that some documents don't have the author property set. In this case only the *Author:* text is shown without any real author name following it. Let's say that we want to fix this in a way that the *Author:* text is not shown in this case at all.

8. Open the **Result Formatting dialog** again.
9. Open the properties window of the **Custom HTML code** element displaying the *Author:* text by double-clicking it in the **Used items** list.
10. In the **Choose when the item should be shown** section at the top of the dialog that opens select *All Office documents* in the **Display if document type matches the filter** drop-down and in the **Display if document has** list choose *Meta author tag* (see the picture below).



This way the *Author:* text will be displayed only for Office documents and only when a result contains an author tag. In other case the text will not be shown.

11. Click **OK** to close the properties dialog.
12. Now, repeat setting the same condition options for the **Line break** element following the **Custom meta tag** item so the line break is also displayed only when the author.
13. Click **OK** to close the **Result Formatting** dialog. Next, go to the preview and run the query again.



As you can see in the picture above no *Author:* text is shown for the second document that doesn't contain information about its author. Still, for the first result item the information is displayed as previously.

See also:

To learn how to change the style of the *Author* text go to the [Using style classes with custom HTML elements](#) topic.

Visit the [Modifying format of a single result](#) topic to learn how you can customize a format of a single result on the results page.

4.4 Generating search engine

In this section you will find quick guides giving you a brief but still very useful information about generating a search engine.

Expand the *Generating search engine* branch in the panel on the left and choose the topic you want to read more about.

4.4.1 Generating CGI based search engine

In this tutorial you can learn how to generate a search engine that uses CGI scripts.

Note: Before you can generate your search engine first you should index some documents. You can learn more about this in [Steps to create a search engine](#) and [Indexing](#).


Follow the steps below to set up necessary output options and generate a search engine:

1. Open the **Search Options** dialog and choose the **Output** tab.
2. Go to the **Target Type** option and select one of the *Binary CGI script* options that is appropriate for your system. For example, if the search engine will be running on Windows IIS you should select *Binary CGI script for MS Windows / IIS*.

Now, you need to specify paths where the search box page, the search database files and script will be generated.

3. Expand the **Physical location of the generated (output) files** option and in the **Output directory for search box page files** field type a path where you want a page with the search box code to be saved, for example `c:\inetpub\wwwroot\cgi-bin\search`.
4. If you want to generate the database and script files to a different location you can specify the path in the **Output directory for CGI/search scripts** field. On the other hand, if you want to save the files to the same folder as the search box page you can leave this option empty so the path from the **Output directory for search box page files** field will be used.
5. Now, expand the **HTTP paths used in output HTML code** option and select **CGI/search scripts directory (HTTP path)**. Here you should type an HTTP path that points to the folder containing the script and database files which you have specified in the previous step. Note, that this shouldn't be a local or FTP path but an HTTP equivalent pointing to the folder with the search files. Since in the sample above we have chosen `c:\inetpub\wwwroot\cgi-bin\search` as the output location for these files, in our case the HTTP path pointing to that folder would be `/cgi-bin/search`.

Note: If you want to automatically upload the files to an online server see the [Uploading search engine to and FTP server](#) topic.

6. Now you can generate the search engine by clicking the  (generate) icon in the [toolbar](#).
7. Now, start your browser and open the generated search box page on your server, for example `http://localhost/cgi-bin/search/search.htm`, and run a search to test the search engine. Note, that this will work only when you have a Windows IIS server set up on your machine. If you want to use another server you have to upload there the generated files first.

Note: Usually you don't use the search box page directly on your web site. In most cases, you only copy its search box HTML code and paste it onto your pages or a template file to embed the search box on your web site. You can also use the other available method of generation, which is a copy&paste method, that saves the database and script files as previously but instead of creating a search box page a dialog with search box code is presented so you can copy the code and use on your web site.


See also:

[Search engine types generated by the program](#)
[Uploading search engine to and FTP server](#)
[Steps to create a search engine](#)

4.4.2 Uploading search engine to an FTP server

In this tutorial you will learn how to automatically upload a generated search engine to an FTP server.

Let's say that you already have a search engine ready to be generated and you want to set up uploading it to your server instead of generating it locally:

1. Open the **Search Options** dialog and choose the **Output** tab.
2. Select the **Upload output to an FTP server** option and choose **Yes**.
3. Expand **Upload output to an FTP server** and in the **Server address** field specify an address of an FTP server where you want to upload the generated files, for example <ftp://myserver.com>.
4. In the **Login** field type a user name that should be used for authentication. If you want the program to remember the password so you don't have to type it every time you generate your search engine you can go to the **Remember password** option, select **Yes** and in the **Password** field type the password for the specified user name.
5. Now it is time to specify a directory where you want the files to be saved on your server. Expand the **Physical location of the generated (output) files** option and in the **Output directory for search box page files** field type a remote path on your server where the search box and the database and script files will be saved. This can be an absolute path, for example `/usr/www/users/<user-name>/search`, or a path relative to the default folder on your FTP server, usually your home directory, for example `public_html/search`.
6. If you want to save the database and script files to a different folder than the search box page you can type it in the **Output directory for CGI/search scripts** field. On the other hand, if you leave this field empty the database and script files will be uploaded to the folder specified in the **Output directory for search box page files** option.
7. Now, expand the **HTTP paths used in output HTML code** option and select **CGI/search scripts directory (HTTP path)**. Here you should type an HTTP equivalent of the path where the database and script files are saved. In other words, this is an HTTP path that points to the folder containing the search script and database files. For example, if you have chosen to upload the script files to `public_html/search` the HTTP equivalent would be `/search` or `http://www.mysite.com/search`.
8. Now you can generate the search engine and upload it to the FTP server by clicking the  (generate) icon in the [toolbar](#).
9. Once the uploading is finished you can start your browser, open the search box page on your server, for example `http://www.mysite.com/search/search.htm`, and test your search engine.

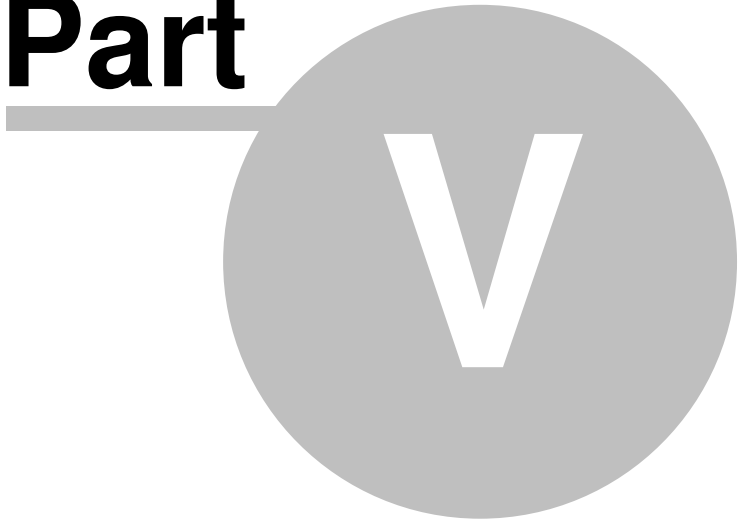
Note: Usually you don't use the search box page directly on your web site. In most cases, you only copy its search box HTML code and paste it onto your pages or a template file to embed the search box on your web site. You can also use the other available method of generation, which is a copy&paste method, that saves the database and script files as previously but instead of creating a search box page a dialog with search box code is presented so you can copy the code and use on

your web site.

See also:

[Search engine types generated by the program](#)
[Steps to create a search engine](#)

Part



5 Tutorials

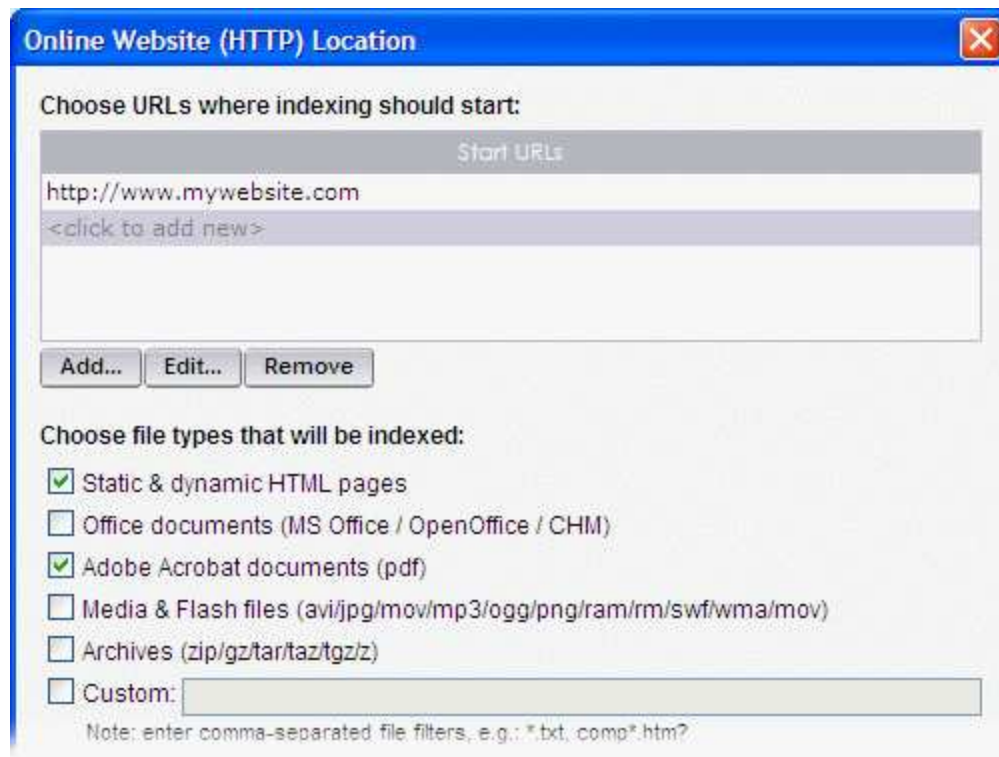
Expand the *Tutorials* branch in the panel on the left and choose the tutorial that you want to go through.

5.1 Creating a search engine for a website

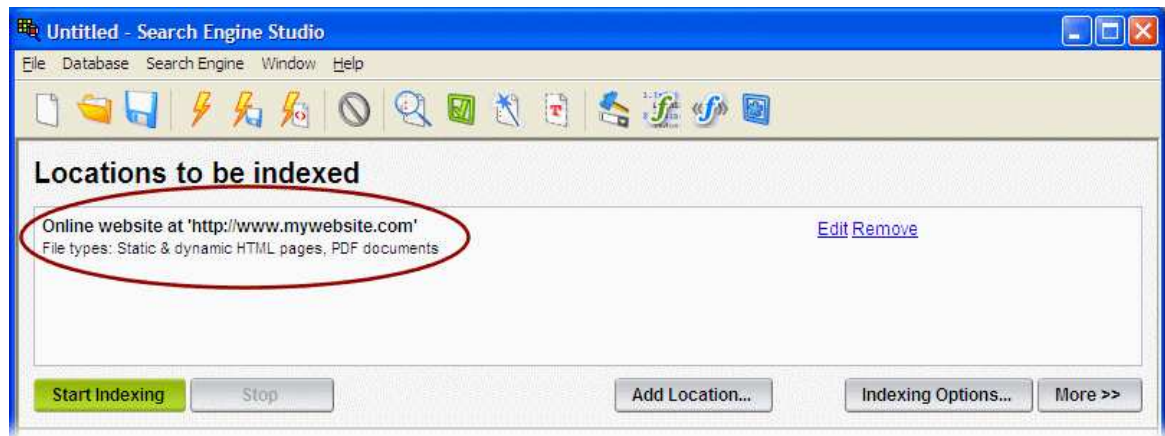
In this tutorial you will learn how to index and generate a search engine for an online web site.

Let's say that you have a personal web site at *http://www.mywebsite.com*. Follow the steps below to create a search engine that allows visitors to your web site to find information they are looking for in a fast and easy way:

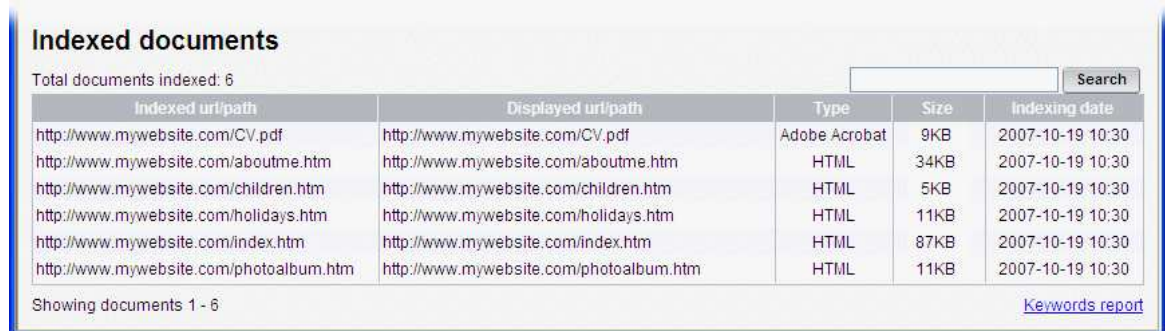
1. In the main program window click the **Add Location** button and choose **Online website via HTTP**. The **Online Website (HTTP) Location** window opens.
2. In the window click the **Add...** button below the **Start URLs** panel, type *http://www.mywebsite.com* and click **OK**. The address appears in the **Start URLs** list (see the picture below). This URL is the place where crawling your web site will start.
3. Now we can choose which document types will be indexed and included into the search database. Let's assume that your web site contains HTML pages and PDF documents. In the **Choose file types that will be indexed** panel select **Static & dynamic HTML pages** and **Adobe Acrobat documents (pdf)** (see the picture below).




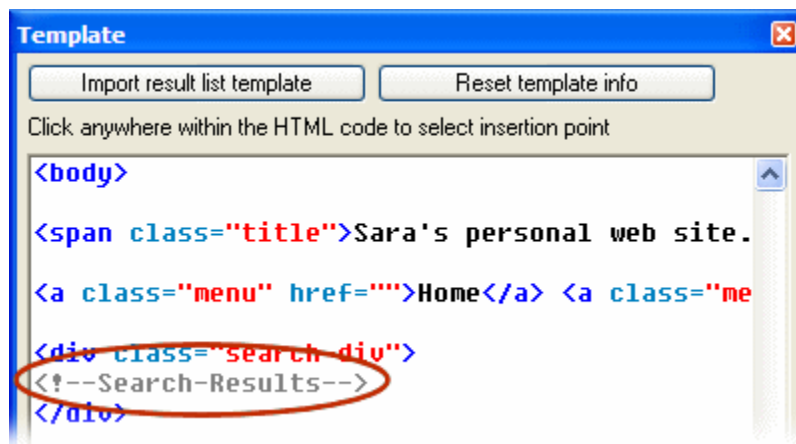
Click **OK** to close the dialog. The location entry appears in the **Locations to be indexed** panel of the main window (see the picture).




4. For the sake of the example, we will leave the default indexing options unchanged. This way the document sections like content, title, description, author etc. will be indexed.
5. Now you are ready to index the web site. Click the **Start Indexing** button you can see in the picture above. The **Log** dialog appears where you can follow the progress of indexing documents.
6. Once indexing is finished there should be some documents visible in the **Indexed documents** panel at the bottom of the main window (see below).




7. Now, it is time to change the appearance of the results page that displays matching documents when you run a query (we are not going to customize the search box layout in this tutorial). We will use a template file for the results page customization. Open the **Template** dialog by clicking the  icon in the [toolbar](#) or choosing **Template** from the **Window main menu**.
8. In the **Template** window click the **Import result list template** button and browse for the file with a layout that you want to use for the results page. The template is usually an HTML page that contains a graphical elements that you want to appear on the results page, like graphical header, footer, company logo, navigation menu etc. Once you choose the file its code appears in the main panel of the dialog. Now you just need to click inside the code in the place where you want to insert the results. The `<!--Search-Results-->` marker should appear there (see the picture below). Close the dialog.



9. Now you can try the search out. Open the Preview window by clicking the  icon in the toolbar and run a query to test it.



10. Once you are finished working on the appearance and testing your search engine it is time to generate it. For the sake of the example we will create a PHP based search engine and generate it on a local disk. If you want to set up automatic uploading to an FTP server read the [Uploading search engine to an FTP server](#) topic. Open the [Search Options](#) dialog and go to the **Output** panel.
11. Select the **Target Type** option and choose *PHP script*.
12. Expand the **Physical location of the generated (output) files** option and in the **Output directory for search box page files** field type a path where you want the page containing the search box code to be saved, for example `c:\mywebsitesearch`. We will leave the **Output directory for CGI/search scripts** field empty so the search script and database files are generated to the same location as the search box.

13. Expand the **HTTP paths used in output HTML code** option and in the **CGI/search scripts directory (HTTP path)** filed type an HTTP path that will point to the location on your server where you want to store the script and database files, for example `/search` or `http://www.mywebsite.com/search_`. This HTTP path is necessary because once you put the search box on your web site it needs to know how to access the search files.
14. Now, go to the main window and generate the search engine by clicking the  (generate to disk) icon in the [toolbar](#). When you browse to the `c:\mywebsite\search` folder you can see the generated files. Now, you can use your favorite FTP program to upload the files to the server hosting your personal web site or, if you are using e.g. FrontPage or Dreamweaver for site development, you can copy the files to the working directory of your web site project so the software automatically uploads the files for you when you publish your site for the next time.
15. If you want to embed the search box on multiple pages of your web site you can retrieve its HTML code from the generated search box file and paste it into the template file that you are using for your site development. As the result the search box will appear on your web site. Thanks to the HTTP path previously set in the **CGI/search scripts directory (HTTP path)** options wherever a page that contains the search box is located in your web site structure the search engine will work fine.

See also:

[Search engine types generated by the program](#)
[Customizing the results page with a template](#)
[Uploading search engine to and FTP server](#)
[Steps to create a search engine](#)

5.2 Indexing & searching in document sections

In this tutorial you will learn how to create a search engine that allows searching only in selected sections/parts of a document, like title, content, description or a custom meta tag.

Let's say that you want to index your web site on cars that contains HTML, MS Word and PDF documents. Each document, apart from the main content, has some additional data fields describing it. In case of HTML files this is title, author, keywords and, let's say, car-description, for example:

```
<html>
<head>
  <title>Audi R3 review.</title>
  <meta name="car-description" content="The A3 looks very stylish inside and out, and it's built to t
  <meta name="keywords" content="Audi, review, A3">
  <meta name="author" content="Sara Johns">
</head>
<body>
  .....
```

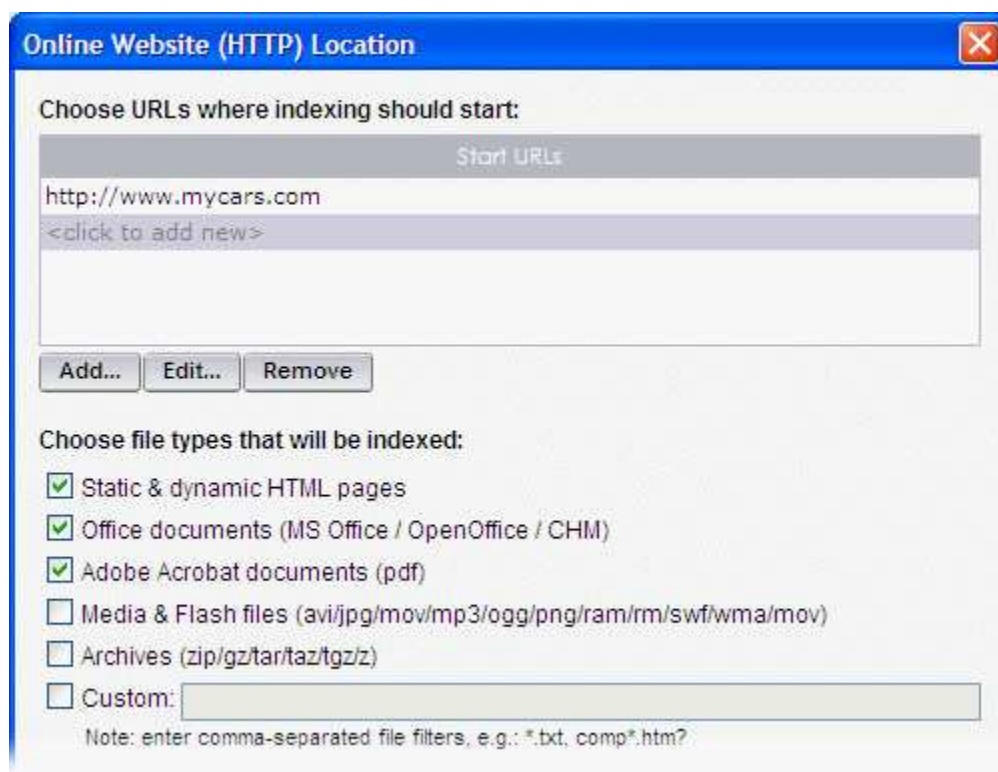
In case of MS Word and PDF documents these property fields include title, author and comments (you can add these fields when creating a Word or a PDF file).

Now, you want create a search engine that allows a visitor to your web site to choose in which document section they want to search, for example in a title. author or car-description field. Follow the steps below:

1. In the main program window click the **Add Location** button and choose **Online website via**

- HTTP.** The **Online Website (HTTP) Location** window opens.
- In the window click **Add...** below the **Start URLs** panel, type `http://www.mycars.com` and click **OK** to add this address as an entry URL where crawling your web site will start.

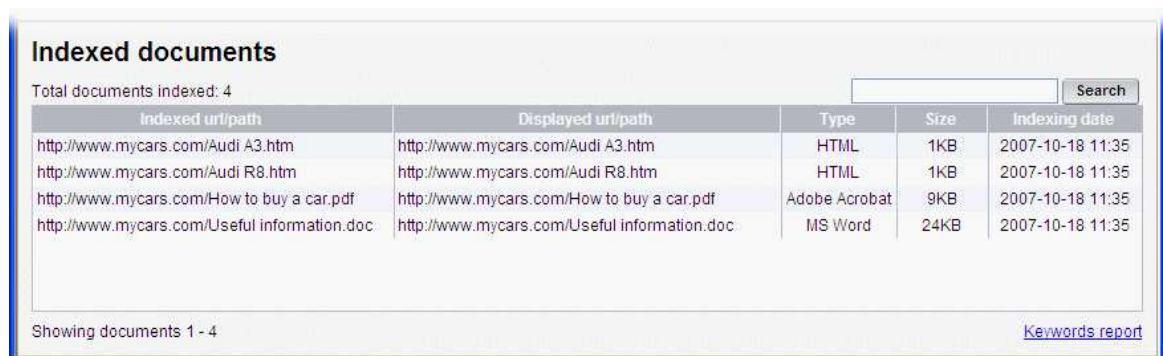
Note: For the sake of the example, an abstract URL `http://www.mycars.com` is used in this tutorial. This also can be another web site or a local folder with documents containing similar sections.
 - In the **Choose file types that will be indexed** panel select **Static & dynamic HTML pages**, **Office documents** and **Adobe Acrobat documents (pdf)** as you can see in the picture below. Click **OK** to close the dialog. The location entry appears in the **Locations to be indexed** panel of the main window.




- Now, it is time to choose which sections of a document will be indexed. As already mentioned we want to index the main content of a document, title, author, keywords, comments and car-description fields. Click the **Indexing Options...** button in the main window to open indexing options.
- In the dialog that opens you can choose which section of a document will be included into the search database. Assuming you have left the default settings *Document content*, *Document title*, *Meta author tag* and *Meta keywords tag* should already be added for indexing. To add the two remaining sections click **Custom meta tag** in the **Add more sections** panel on the right, type *comments* and click **OK**. This way you have chosen for indexing the comments meta field. Next, do the same for the *car-description* tag. Both options appear in the main panel as in the picture below. Click **OK** to close the dialog.

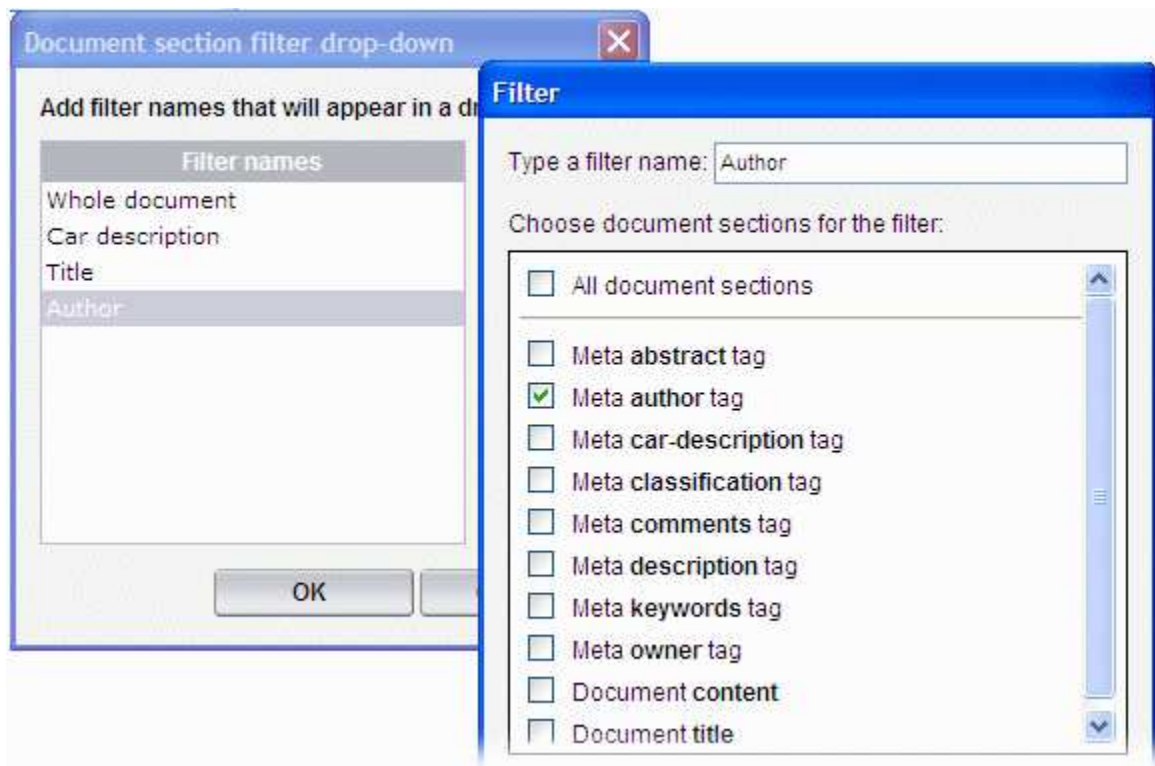


6. As you close the indexing options window an information dialog appears. It informs you that the changes you have made will not be reflected in the search database until you re-index your documents. This time click **No** to close it.
7. In the main window click the **Start Indexing** button. The indexing process starts and you can follow the progress in the **Log** dialog that automatically opens. Once indexing is finished there should be some documents indexed as you can see in the **Indexed documents** panel at the bottom of the main window (see the picture below).

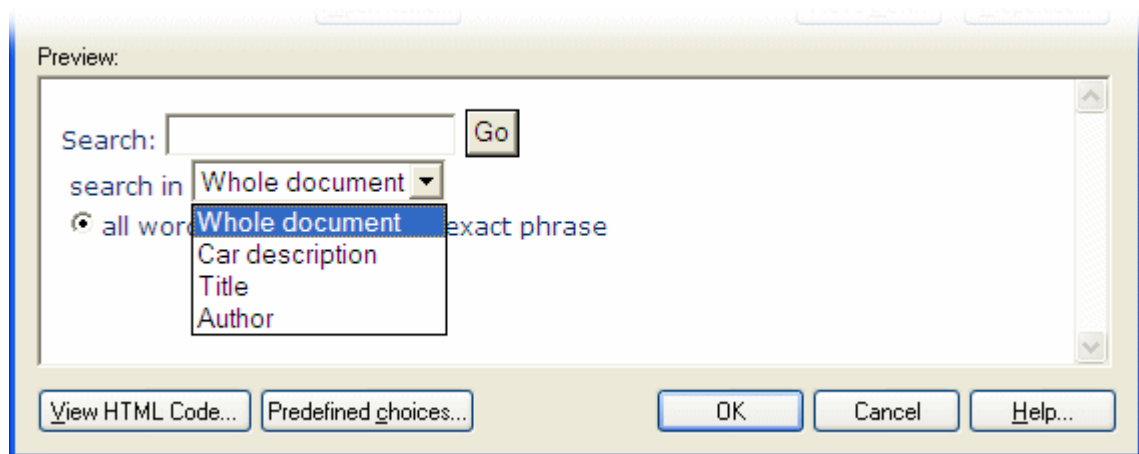


8. Open the **Search Box Builder** dialog by clicking the  icon in the [toolbar](#) or choosing **Edit Search Box...** in the **Search Engine** menu.

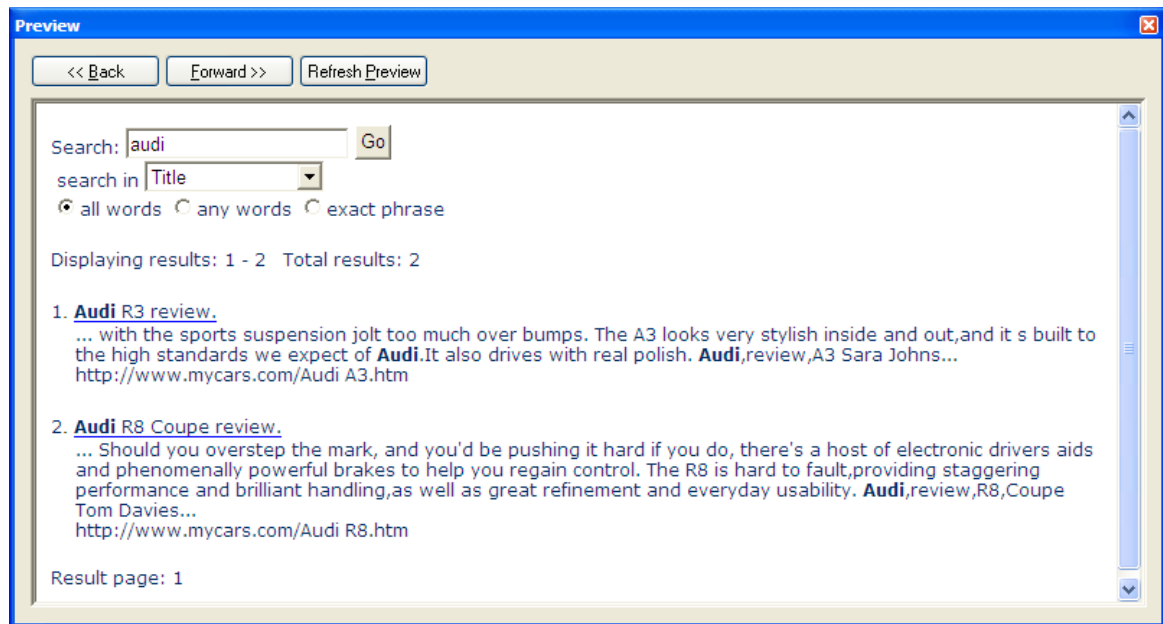
9. After the **Line break** element following the **Text-based search button** item in the **Search Box Items** panel on the right add a new **Custom HTML code** item by double-clicking it in the **Available items** panel on the left.
10. In the **HTML Source** dialog that opens type ` search in: `;
11. Next, add a **Document section filter drop-down** item. The properties dialog opens.
12. In the dialog click **Add** to insert a new drop-down entry option. In the **Filter** window that appears type *Whole document* as the filter name, select the *All document sections* option and click **OK**.
13. Repeat the operation to add a few more options to the drop-down list. For the first option type *Car description* as the name and select the *Meta car-description tag* option, next type *Title* and select *Document title* and finally for the last option type a name *Author* and select *Meta author tag* (see the picture below).



14. Click **OK** to close the **Document section filter drop-down** dialog and add one more **Line break** item after the drop-down element. As you can see in the preview panel at the bottom of the **Search Box Builder** dialog a new drop-down list is displayed with the options you have just inserted (see the picture below). Click **OK** to return to the main window.



15. Now, let's move to the **Preview** window and test the search engine (click the preview icon in the toolbar). As you can see in the picture below you can use the drop-down list to limit the search scope to one of the available areas of a document. If you want to search in the whole document you can always choose the *Whole document* option. On the other hand, if you want to search only in a document title simply select *Title* in the drop-down and start the search.



See also:

[Steps to create a search engine](#)
[Generating CGI based search engine](#)

5.3 How to create a more precise search

In this topic you will learn how to create a search engine that allows a visitor to your web site run a more precise query and get better matching results. You can do this by adding to your search box one or more available search filters and options so a user can limit the search scope to a smaller subset of documents.


In this sample you will extend the search functionality with:

Document section filter that allows to search only in e.g. a document title, description or an author, **Document type filter** where you can choose to search only in files of a specified type, for example HTML, Office or PDF documents, **URL filter** where you can select a department of your web site that you want to search, and **date filter** that limits the search scope to files matching selected date options,

Let's say that you have indexed an online web site that contains two departments:

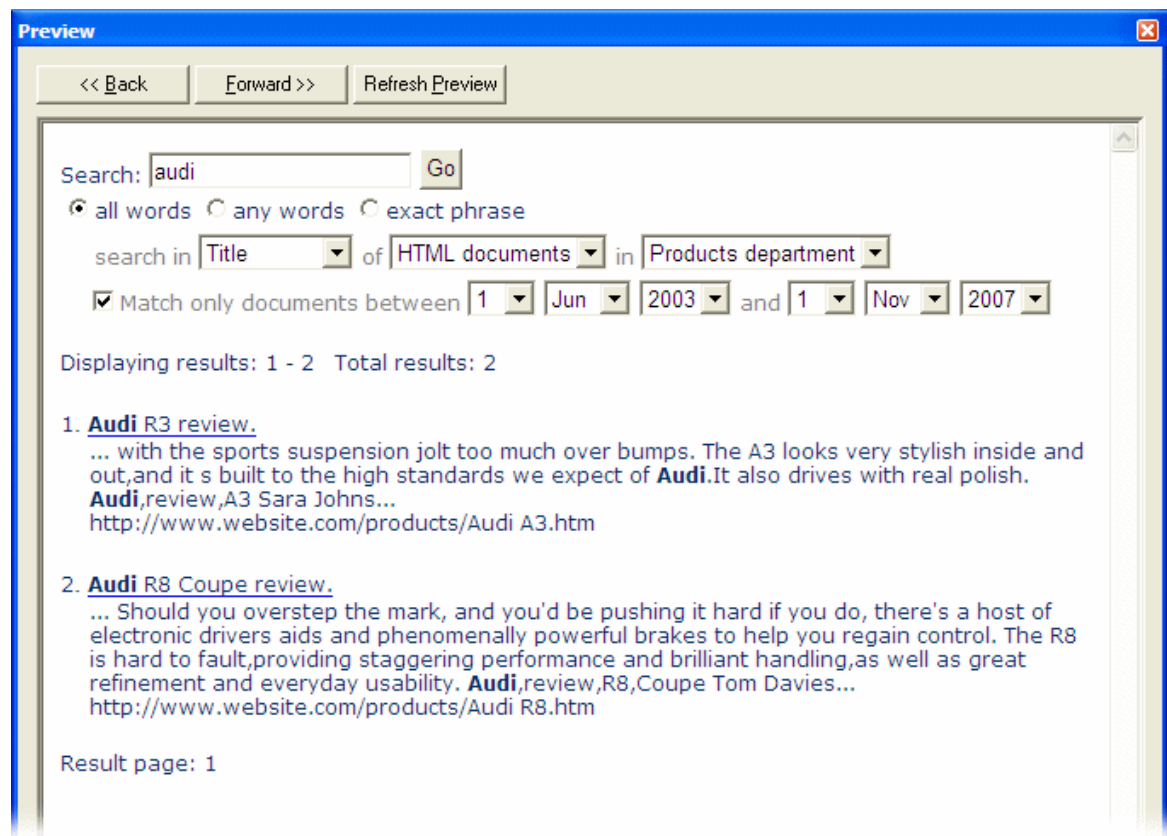
- <http://www.website.com/products> - that contains pages describing the products offered by your company
- <http://www.website.com/help> - where you can find a number of help documents

Follow the steps below:

1. Open the **Search Box Builder** dialog by clicking the  icon in the [toolbar](#) or choosing **Edit Search Box...** from the **Search Engine** menu.
2. After the **Search mode choice** element in the **Search Box Items** section on the right add a new **Line break** item by double-clicking it in the **Available items** panel on the left.
3. After the line break, add **Custom HTML code** and in the dialog that opens type `search in: ` and click **OK**.
4. Now, we will add a drop-down list that allows to choose a section of a document to search in. Add a **Document section filter drop-down** item from the **Available items** panel on the left. The properties dialog opens (see below).
5. In this dialog click **Add** to insert a new drop-down option. In the **Filter** window that appears type *All parts* as the filter name, select the *All document sections* option (see the picture below) and click **OK**.



6. In the same way add two more options. For the first one type a name *Title* and select *Document title* from the list and for the second one type *Description* and select the *Meta description tag* option. You can learn more about **Document section filter** in [Adding document section filter to the search box](#). Click **OK** to close the **Document section filter drop-down** window.
7. Then, add another **Custom HTML code** item and type `of `;
8. Next, we will add a filter that allows choosing a document type to limit a search scope. We will add options to search in all types of files, and only in HTML or PDF documents. Add a **Document type filter** element and in the dialog that opens type *ALL,HTM,PDF* in the first field and *All documents,HTML documents,PDF documents* in the second one. Click **OK** to close the dialog. You can see the newly added drop-down in the preview at the bottom of the **Search Box Builder** dialog.
9. Add one more **Custom HTML code** item and type `in `;
10. Now, we will add one more filter that allows to choose whether you want to search in the whole web site or only in the selected department. Double-click a **URL filter** item to add it and type *NONE,http://www.website.com/products,http://www.website.com/help* in the first edit field and *Whole website, Products department, Help department* in the second one. Close the dialog.
11. Add a **Line break** item so next search box elements are displayed in the next line.
12. Now, we will add an options that allows to specify a date of a document you want to find. Add an **Enable advanced data filter (checkbox)** item, type *Match only documents between* in the text field of the window that opens and click **OK** to close it.
13. Next, add **First allowed date filter** and type *1998* and *2007* as the first and the last available year and **Last allowed date filter** and enter the same values.
14. Between the two data filter items you can add a **Custom HTML code** element and type `and `;
15. Close the **Search Box Builder** dialog and go to the **Preview** window. As you can see in the picture below you have created a search box with a number of options that allow limiting a scope of the search to get more precise results and find documents that better match your criteria. In the sample below we are searching for HTML documents from the products department that contain a word *audi* in a title tag. Additionally, the matching documents have to be created or modified only between the chosen dates.



Note: If you want to learn how to generate a search engine to a local server or upload it to a remote server via FTP go to the [Generating CGI based search engine](#) or [Uploading search engine to an FTP server](#) section.

See also:

- [Steps to create a search engine](#)
- [Generating CGI based search engine](#)
- [Uploading search engine to an FTP server](#)
- [Indexing document sections](#)

Part



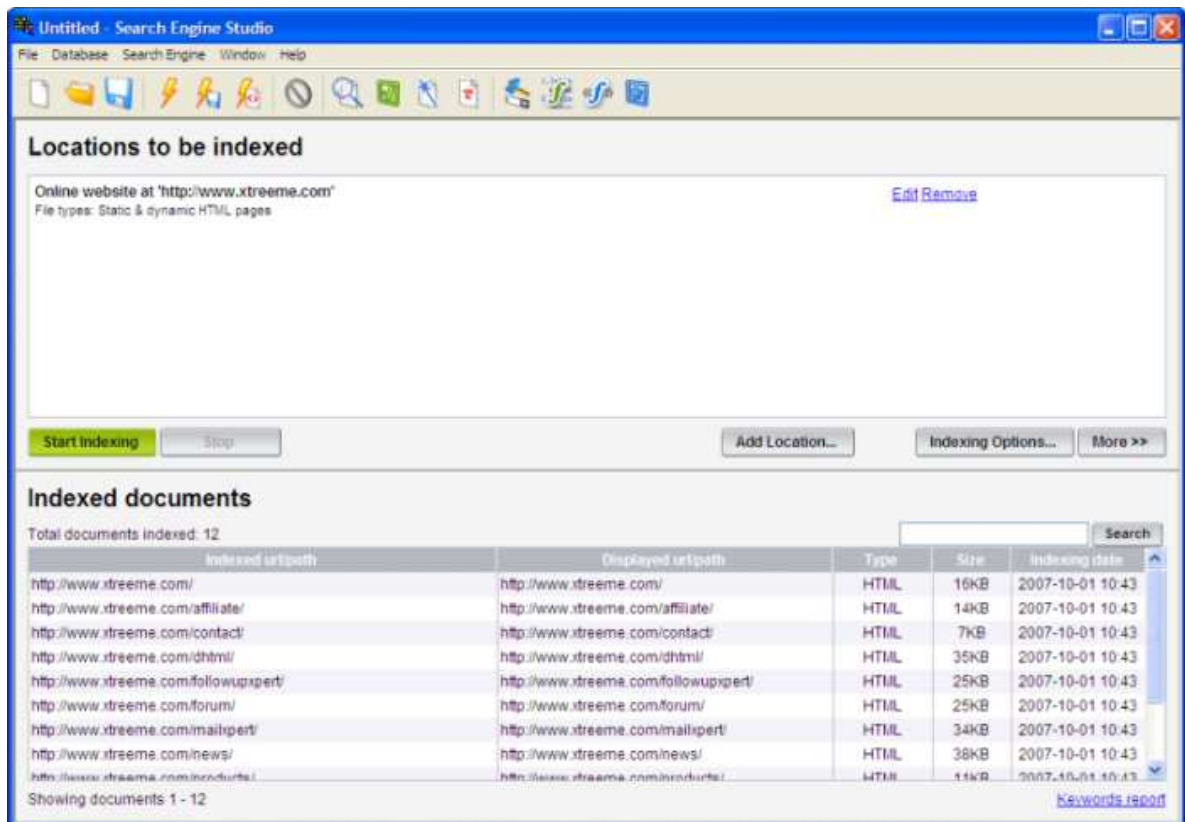
6 The User Interface

This section contains a reference of the program's user interface.

Choose a dialog's name in the *The User Interface* branch in the panel on the left to get a description.

6.1 Main window

This topic provides a brief guide to the main program window.



At the top of the window you can see the [main menu](#) and [toolbar menu](#). Click one of the links to read more about the commands and options available in these menus.

The Locations to be indexed panel

In this panel you can set up locations that you want to index. In the middle of this panel you can see currently added locations. You can clicking a link to the right of each location to edit or remove it.

Start Indexing

You can use this button to start the indexing process. Once you click it, the indexing of locations starts and you can follow the process in the [Log](#) window that opens.

Stop

When indexing is running you can click this button when you want to stop it.

Add Location...

By clicking this button you can open a wizard where you can choose one of four available indexing methods (HTTP, local folders, FTP and XML) and define its settings. Once you are finished the location's entry appears in the **Locations to be indexed** panel.

Indexing Options...

By clicking this button you opens the [Indexing Options](#) dialog where you can set up document sections (parts) that you want to index.

More >>

This button opens a menu where you can find advanced options for manipulating indexed documents, like re-indexing or removing single documents from pattern or XML file etc.

The Indexed documents panel

In this panel you can view a list of indexed documents. In the table containing information about indexed documents you can see their original path and the path that will be displayed in the search results (see [mapping](#) to learn more about his), document type (HTML, PDF, MS OFFICE etc.), size and the date when a document was indexed for the last time.

Keywords report

Clicking this link opens a dialog when you can view all indexed keywords.

Search

If you want to find a specific document you can use the search box in the top right corner of the panel. When you type a part of document's name or path and click **Search** only documents containing the specified text in their paths will be shown.

Show All

After you use the search option the table is showing only the matching documents. When you want to go back to all documents, click this button.

6.2 Main menu

At the top of the main window you can see the main menu bar. Here is the list of commands you will find in each drop-down.

File

This menu contains the basic file options for creating, opening and saving Search Engine Studio projects. Most of the options have corresponding icons in the toolbar, and several already have

keyboard shortcuts assigned to them.

New:	Creates a new project.
Open:	Opens an existing project file.
Save:	Saves the current project.
Save As:	Saves the current project under a different name.
Recent Projects:	Opens a project you have worked on recently.
Exit:	Exits the program.

Database

Here you can find commands and options related with indexing.

Stop indexing:	Stop currently running indexing. You can also use the Stop button in the main window.
Excluded Document Parts:	Opens Exclude Document Parts Rules dialog where you can define which parts of a document should be excluded from indexing.

Search Engine

You will find here options and dialogs where you can work on search pages appearance and generate a search engine.

Generate search engine:	Generates search engine files to the output location.
Stop generation:	Stops generation.
Edit search box:	Opens a dialog where you can customize layout and functionality of your search box.
Result Formatting:	Opens a window that allows you to define what should be shown in a resulting document's preview.
Result Page Overview:	Opens Result Page Overview where you can tailor for your needs the layout of the search results page.
Password Manager:	Opens a window where you can enter passwords that will be used by the indexer's HTTP scan.
Foreign Language Support\ Character Translation:	Opens Character Translation Table where you can add translation of special characters.

Window

Here you can open tool windows.

Assistant:	Starts assistant that guides you through the process of indexing and generating a search engine.
Options:	Opens Search Options where you set up properties of the search engine.
Preview / Search Test:	In this window you can see the preview of the search engine and run sample queries to test it.
Template:	Here you can import an existing template file to customize the search results page.
Log Window:	Opens a window where you can see error, warning and status messages which inform you about the indexing or generation process.
Log Analyzer:	Starts an Add-on program where you can analyze log files created on server that contain all search queries.

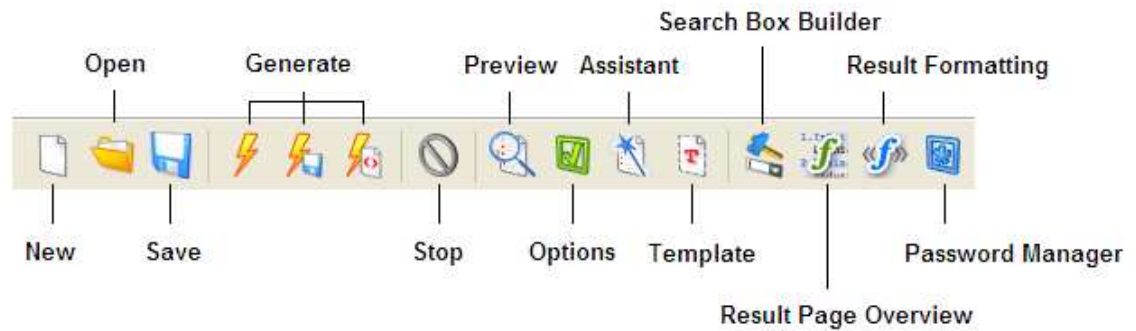
Help

In this menu you can find resources related with technical support and program usage.

Program Help:	Opens this help file.
Search Engine Studio Home Page:	Takes you to the program's home page at http://www.xtreeme.com/search-engine-studio .
Purchase Search Engine Studio:	Opens a page where you can purchase a full version of the program.
Technical Support Page:	Opens online technical support page.
Send an Email to Technical Support:	Use this option to send an email to technical support.
About Search Engine Studio:	Opens the About dialog where you can find the program version and enter document limit license key.

6.3 Toolbar

The toolbar menu gives your a quick access to the most frequently used dialogs and options.



New: Creates a new document.

Open: Opens an existing project file.

Save: Saves the current project.

Generate: Allows you to choose to generate search engine to disk or FTP or as copy&paste code.

Generate to disk: Generates search files and search box page to local disk or FTP.

Generate copy & paste code: Generates search files and presents the search box code which can be copied & pasted.

Stop: Stops currently running indexing.

Preview: Shows the preview window.

Options: Opens **Search Options** dialog.

Assistant: Starts assistant that guides you through the process of indexing and generating a search engine.

Template: Opens **Template** window where you can import an existing template file to customize the search results page.

Search Box Builder: Opens a dialog where you can customize layout and functionality of your search box.

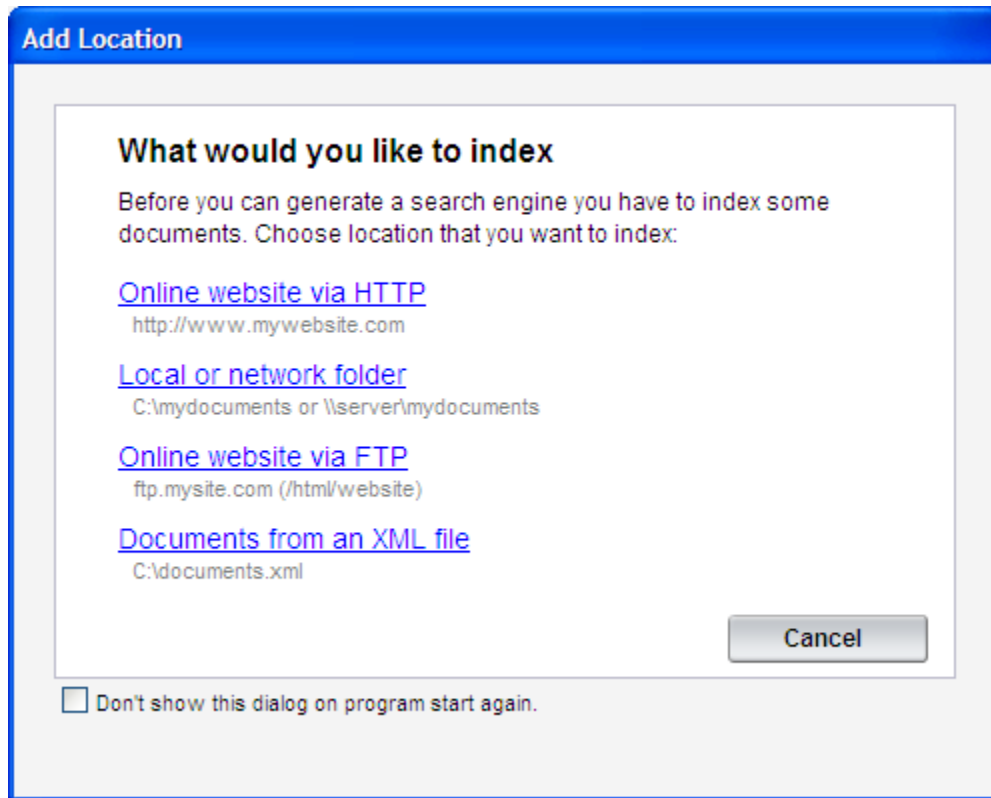
Result Page Overview: Opens a window where you can tailor for your needs the layout of the search results page.

Result Formatting: Opens a window that allows you to define what should be shown in a resulting document's preview.

Password Manager: Opens a window where you can enter passwords that will be used by the indexer's HTTP scan.

6.4 Add Location dialog

In this dialog you can choose a location of the documents that you want to index and include into your search engine.



There are four possible options you can choose:

- | | |
|------------------------------------|--|
| Online website via HTTP: | crawls through an Internet or Intranet website using the HTTP protocol |
| Local or network folder: | reads files located within a specified root directory on a local or network disk |
| Online website via FTP: | reads files using the FTP protocol |
| Documents from an XML file: | reads a list of documents to be indexed from an XML file |

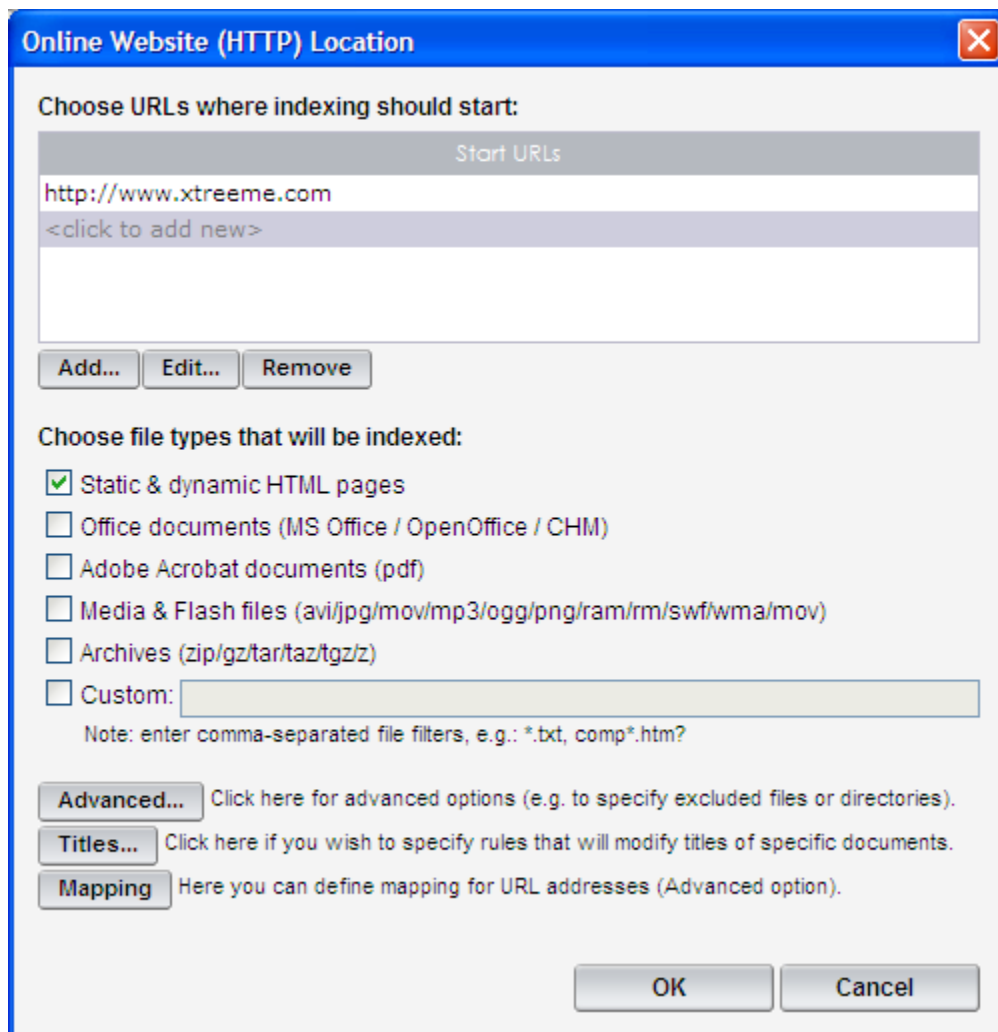
See also:

[Online Website \(HTTP\) Location dialog](#)

[Local/Network Folder Location dialog](#)
[FTP Location dialog](#)
[XML Location dialog](#)

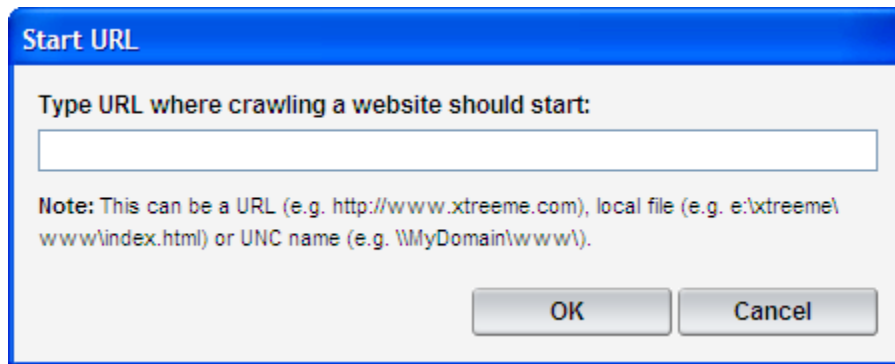
6.5 Online Website (HTTP) Location dialog

In this window you can add a URL of an online or intranet web site that you want to add to your search. This method uses HTTP protocol which allows to index dynamic web pages that first have to be processed by a server, like PHP, ASP, JSP or ColdFusion. If you are using this method Search Engine Studio, starting with a specified start URL(s), crawls through the whole web site and indexes each document matching the chosen indexing rules, like document type, URL or title filter etc.



Start URLs

In this panel you can add URLs where crawling a web site should start. Clicking the **Add...** button opens a **Start URL** dialog (see below) where you can type a URL you want to add. You can always edit or remove each URL by using the **Edit...** and **Remove** buttons.



Document types

This section allows you to choose which document types will be included in your search. You can select to index Static & dynamic HTML pages, MS Office, CHM and PDF documents and more. Additionally, you can define custom document filters. This allows you to index, for example, text documents (*.txt) or pages matching file name pattern (comp*.htm?).

Advanced

Clicking this button opens [Advanced Options dialog](#) where you can configure the indexer according to your specifications.

Titles

This button opens [Titles Transformation Rules dialog](#) where you can enter one or more rules transforming title names pulled from the indexed documents.

Mapping

This is an advanced option that usually is not needed when indexing a web site. It allows you to map one base URL into another one. You can read more about this in [Online Documents Mapping dialog](#) and [Mapping](#).

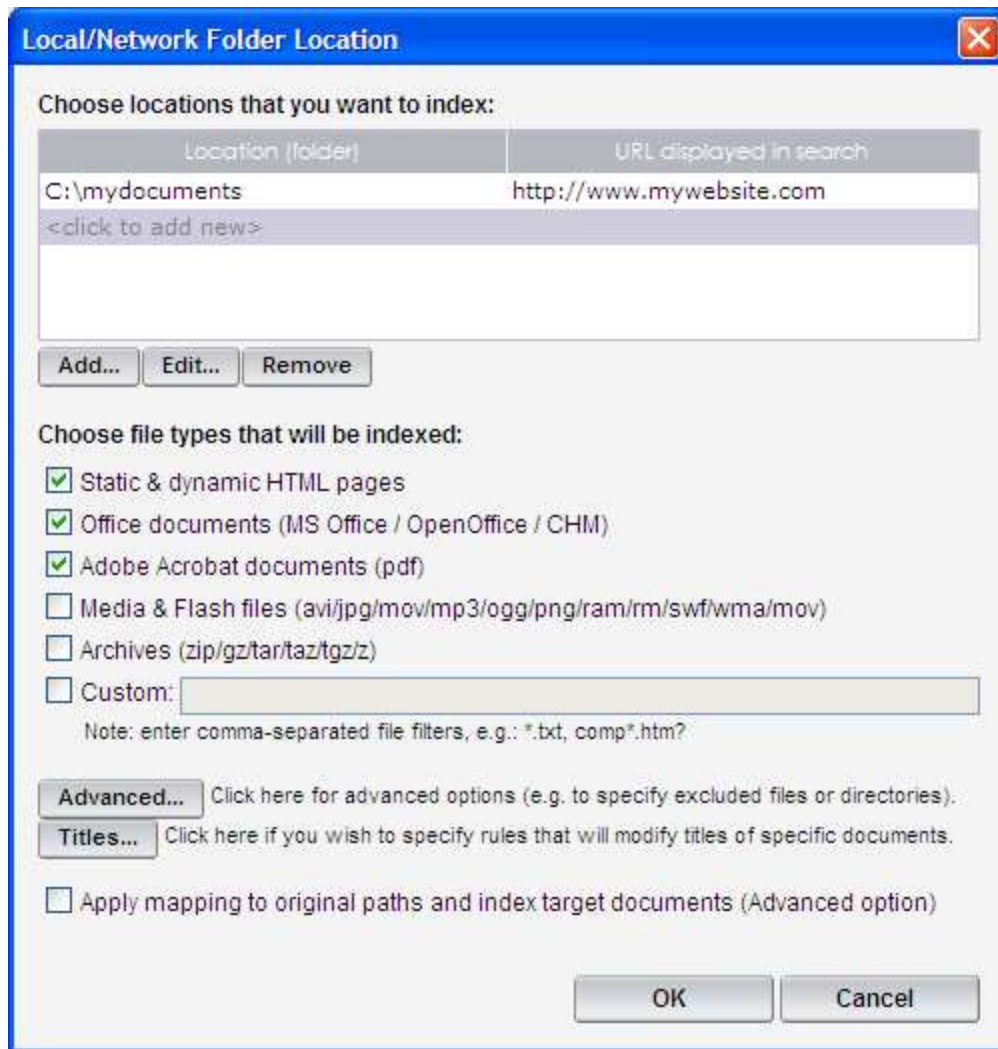
See also:

[Indexing online web site](#)

6.6 Local/Network Folder Location dialog

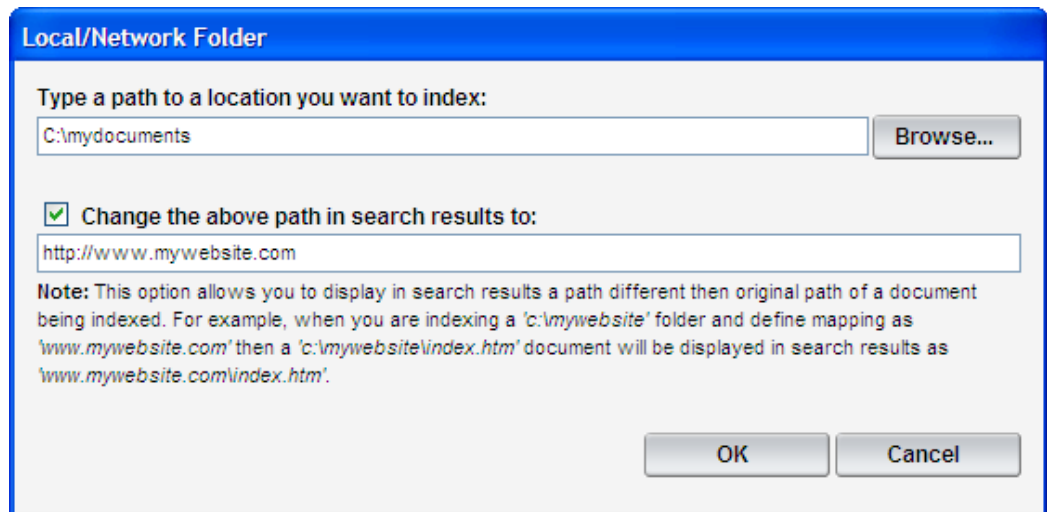
This is the place where you can set up indexing document residing on your local or network disk. This method is usually used when you have a copy or mirror of your online documents on your local or

network disk or when you are creating an offline search for a CD-ROM/DVD. When using this method Search Engine Studio reads one by one documents residing in a specified local/network folder and adds them to the search database.



Locations/folders

This is the place where you can add folders that you want to index. Additionally, for each folder you can type a URL that will be displayed instead of an original path in search results (you can read more about this in the [Mapping](#) topic). To add a folder click the **Add...** button that opens the **Local/Network Folder** dialog (see the picture below). You can also use the **Edit...** and **Remove** buttons to edit each folder or exclude it from indexing.



Document types

This section allows you to choose which document types will be included in your search. You can select to index Static & dynamic HTML pages, MS Office, CHM and PDF documents and more. Additionally, you can define custom document filters. This allows you to index, for example, text documents (*.txt), pages matching file name pattern (comp*.htm?) etc.

Advanced

Clicking this button opens [Advanced Options dialog](#) where you can configure the indexer according to your specifications.

Titles

This button opens [Titles Transformation Rules dialog](#) where you can enter one or more rules transforming title names pulled from the indexed documents.

Apply mapping to original paths and index target documents

This is an advanced option that allows you to index target/online documents instead of their local copies. For example, when you are indexing a 'c:\documents' folder and define '<http://www.website.com>' as the mapping then once a 'c:\documents\index.php' file is found in this folder mapping is immediately applied and the target '<http://website.com/index.php> file is indexed instead.'. This option is especially useful when your documents contain dynamic web pages that should rather be indexed through HTTP protocol.

See also:

[Indexing local or network documents](#)

6.7 FTP Location dialog

In this dialog you can set up indexing online documents using FTP protocol. Once you define your FTP server address, user name and password and enter remote folders that should be indexed, Search Engine Studio opens these folders and indexes all documents residing there.

FTP Location

Specify details of FTP server that will be used for indexing:

FTP server address:

User name:

Password: Remember password

Use proxy server (firewall)

Choose remote locations (folders) that you want to index:

Remote Location (folder)	URL displayed in search
html-pages/	http://www.website.com
<click to add new>	

Choose file types that will be indexed:

Static & dynamic HTML pages

Office documents (MS Office / OpenOffice / CHM)

Adobe Acrobat documents (pdf)

Media & Flash files (avi/jpg/mov/mp3/ogg/png/ram/rm/swf/wma/mov)

Archives (zip/gz/tar/taz/tgz/z)

Custom:

Note: enter comma-separated file filters, e.g.: *.txt, comp*.htm?

Click here for advanced options (e.g. to specify excluded files or directories).

Click here if you wish to specify rules that will modify titles of specific documents.

FTP server settings

FTP server In this field you should type the address of the FTP server that you want to use for indexing.

address:

User name: This is the FTP account user name that you want to use.

Password: Here you type the account password. You turn on the **Remember password** option if you don't want to be asked for the password each time you start indexing.

Use proxy server: You should select this option if you are behind a firewall and cannot connect directly to the FTP server (click the **Proxy settings** button to specify detailed options).

Remote Location (folder)

This is the place where you can add remote the folders on your server that you want to index. Click **Add...** to open a **Remote Folder** dialog (see below) where you can type a path of a remote folder and a URL that will be show instead in search results (you can read more about this in the [Mapping](#) topic). You can also use the **Edit...** and **Remove** buttons to modify folders or remove them.

Remote Folder

Type a remote location you want to index:

html-pages/

Change the above path in search results to:

http://www.website.com

Note: This option allows you to display in search results a path different then original path of a document being indexed. For example, when you are indexing a 'mywebsite' folder and define mapping as 'www.mywebsite.com' then a 'mywebsite\index.htm' document will be displayed in search results as 'www.mywebsite.com\index.htm'.

OK Cancel

Document types

This section allows you to choose which document types will be included in your search. You can select to index Static & dynamic HTML pages, MS Office, CHM and PDF documents and more. Additionally, you can define custom document filters. This allows you to index, for example, text documents (*.txt), pages matching file name pattern (comp*.htm?) etc.

Advanced

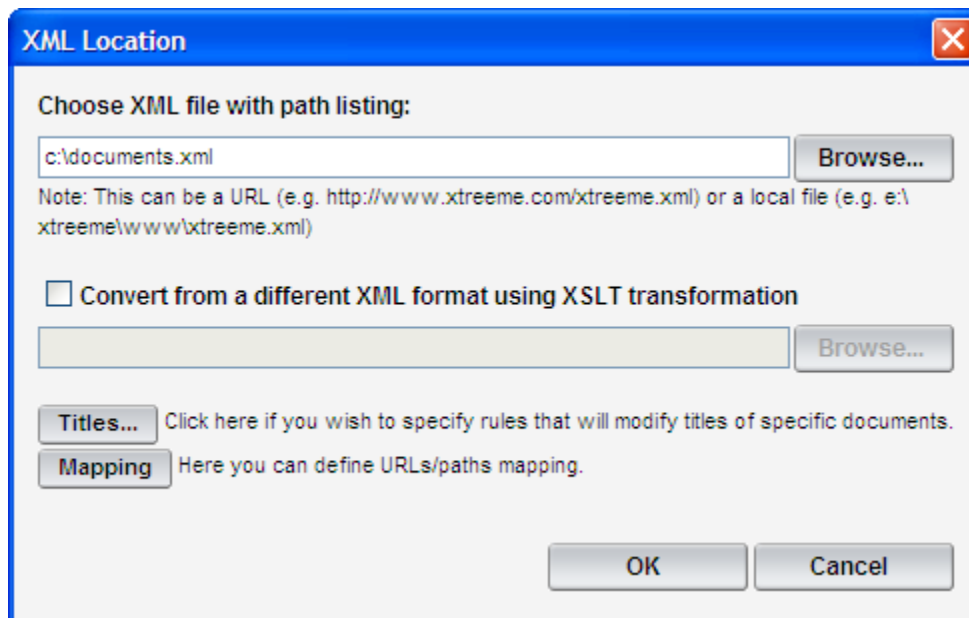
Clicking this button opens [Advanced Options dialog](#) where you can configure the indexer according to your specifications.

Titles

This button opens [Titles Transformation Rules dialog](#) where you can enter one or more rules transforming title names pulled from the indexed documents.

6.8 XML Location dialog

In this window you can choose an XML file containing paths of the documents that you want to index. Additionally, you can use an XSLT transformation that converts the XML from a different format.



XML file path

In this field you should type a path to an XML file containing paths of documents that you want to index. To learn more about the format of this XML see the [Indexing documents from XML file](#) topic.

Convert from a different XML format using XSLT transformation

If your XML file is using a different format than the one accepted by Search Engine Studio you can turn this option on and type a path to a file with XSLT transformation that will carry out the conversion.

Titles

This button opens [Titles Transformation Rules dialog](#) where you can enter one or more rules transforming title names pulled from the indexed documents.

Mapping

This is option allows you to map one base URL into another one. You can read more about this in [Documents Mapping dialog](#) and [Mapping](#).

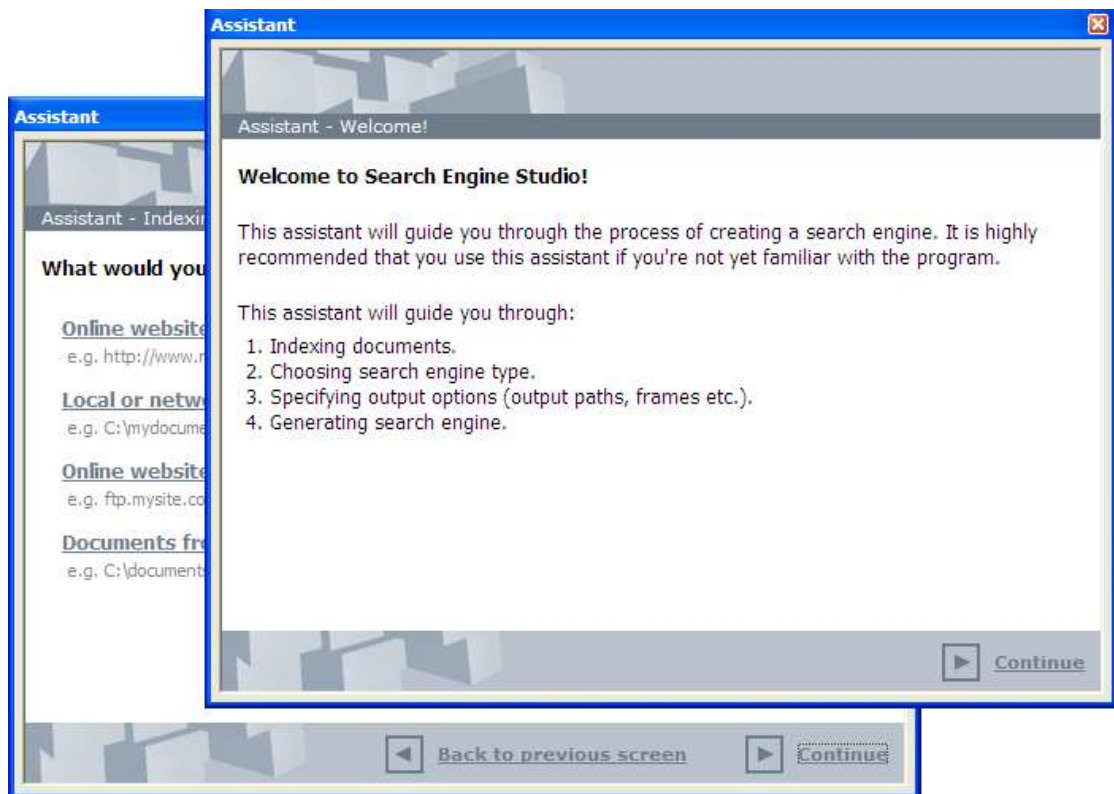
See also:

[Indexing documents from XML file](#)

6.9 Assistant dialog

The Assistant window opens automatically when you start Search Engine Studio for the first time. You can also start the assistant at any time by clicking the icon in the program [toolbar](#) or choosing **Assistant** from the **Window** menu.

The assistant is the right place to get started with the program. Although it doesn't contain all functionality which is accessible from the rest of the program's user interface, following the assistant's commands will let you build a new search engine from scratch and insert it into an existing page, template or frameset configuration.

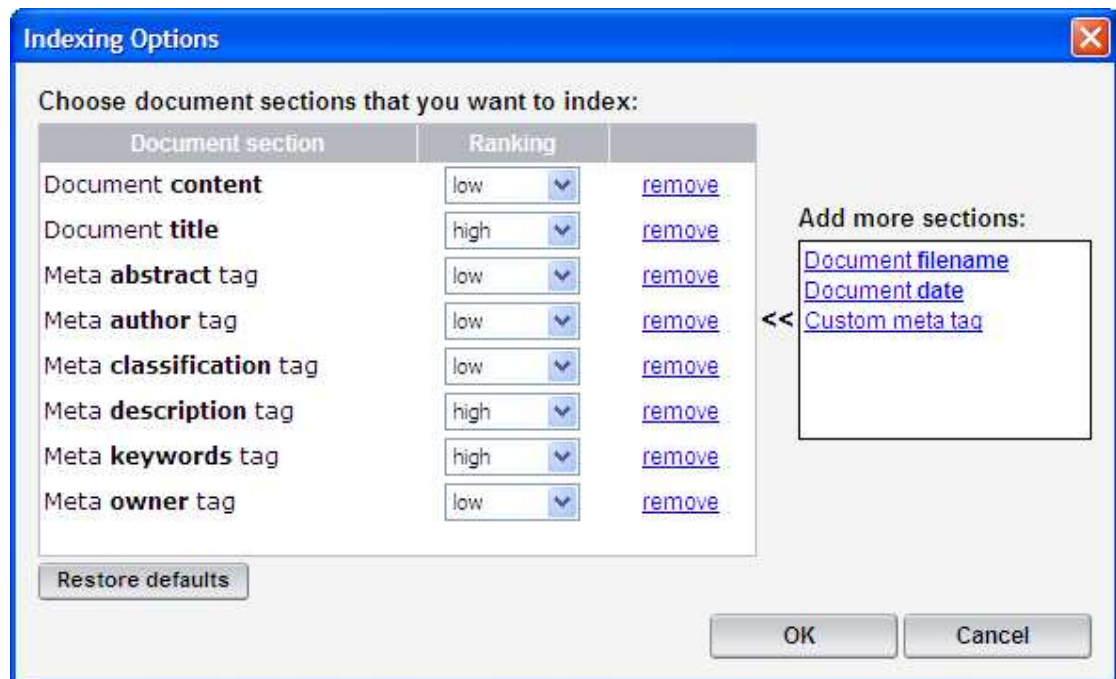


6.10 Indexing Options

In this dialog you can set up the parts of a document that you want to index and include into your search engine. Usually, you want to index document content, title and some meta tags like author, keywords or description. You can also choose to index document date or custom meta-tags. Additionally, you can assign each section a ranking factor (values from *low* to *very high*) which

determines the influence of this section on total score for given document. A document's score determines its position on the list that is displayed on the search results page.

To read more about indexing document sections and creating a search box that allows to search by them see the [Indexing & searching by custom document sections](#) topic.



Document sections list

In the main table of this dialog you can view the document sections that are currently selected for indexing. The **Document section** column lists a name of a sections. In the **Ranking** column you can set a ranking that determines the influence of keywords indexed in this section on final score of the document. By clicking the **remove** link in the third column you can delete a section so it will not be indexed.

Add more sections

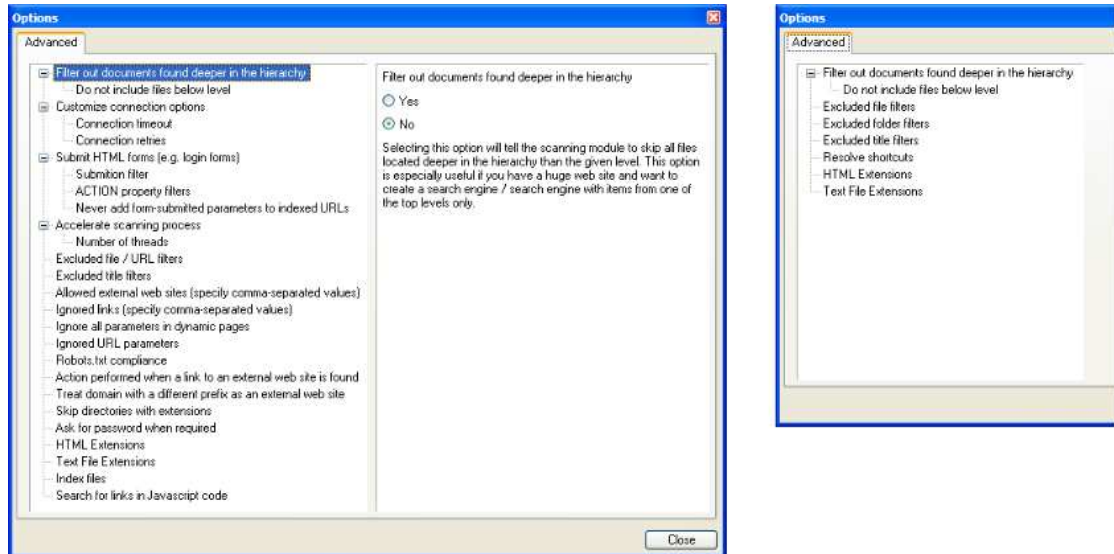
In this box you can add a new sections that you want to index. When you want to index a custom meta tag you should choose the **Custom meta tag** option and type a name of this tag in the window that opens. One you click **OK** an appropriate entry appears in the main table.

The **Restore defaults** button

Click this button when you want to return to the default settings.

6.11 Advanced Options dialog

In this dialog you can configure the indexer according to your needs. This window opens when you click the **Advanced...** button in the [Online Website \(HTTP\) Location dialog](#), [Local/Network Folder Location dialog](#) or [FTP Location dialog](#) dialog. Still, for online web site location this window contains different options then for a local/network or FTP location.



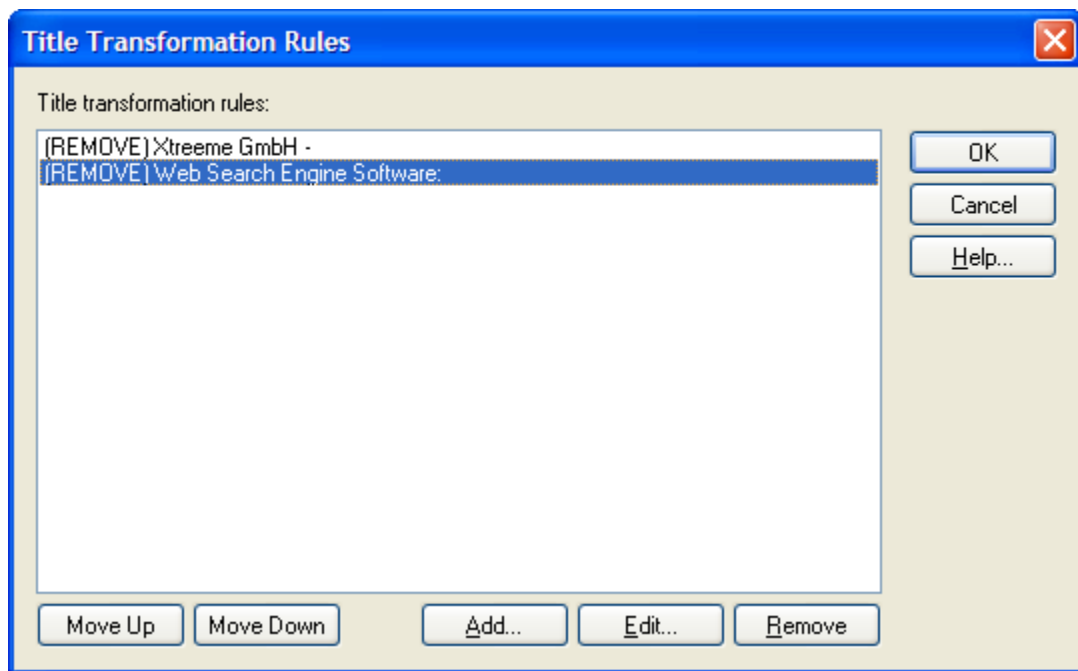
The picture on the left presents options for online HTTP scanning and the picture on the right options for local/network folder indexing. As you can see additionally to some common options like filtering out documents found deeper in the hierarchy or excluding documents by title, the online web site dialog on the left contains also options related with HTTP protocol, like ignoring URL parameters, action performed when linking to an external web site etc.

When you select an option a short description appears in the panel on the right so you can learn more about its functionality.

6.12 Titles Transformation Rules dialog

This window is accessible from all four location dialogs.

Here you can enter one or more rules which transform title names pulled from the indexed documents. This is especially useful when your document titles always contain a common part, like a company name which you don't want to be displayed in the search results.



You can add one of the following four rule types:

- Add Text: Use this rule to add specified text to the beginning or end of titles. If you specify a text in the option **Only if text contains**, the rule will be applied only to documents whose titles contain given text.
- Remove Text: Use this rule to remove specified text from the beginning or end of titles.
- Replace Text: This can be used to replace parts of titles with another text. You can also use this rule to remove parts of texts (in such case 'Insert following text' should be left empty).
- Advanced Transformation: This rule type can be used to perform advanced modifications of the titles using regular expressions.

Regular Expressions

Regular expressions are patterns that describe character combinations in text. They can be used to describe concepts such as 'words that start with letter T' or '3-digit numbers'. The format expression can then be used to reformat a title text matching given regular expression.

The following list specifies special characters you can use in the 'regular expression' field, their meanings (what types of character combinations they match), and usage examples. Note: to treat a special character as the actual character (e.g. a * character in the title), you'll need to precede it with backslash, e.g. '*note'.

- .: Any single character, e.g.: .ox will match box in 'blue box'
- ^: Beginning of text, e.g.: ^A matches A in "A Cool Title" but not in "This Is A Cool Title"
- \$: End of text, e.g.: e\$ matches e in 'A cool title' but not in 'Title ABC'
- \b: Word boundary, e.g.: \ba will match a in 'active' but not in 'rabbit'
- \B: Anything other than a word boundary, e.g.: \ba will match a in 'active' but not in 'rabbit'
- +: The preceding character 1 or more times, e.g.: ex+ will match ex in 'extra' and in

- 'exxtra' but will NOT match e in 'error'
- *: The preceding character 0 or more times, e.g.: ex* will match ex in 'extra' and in 'exxtra' but will also match e in 'error'
- ?: The preceding character 0 or 1 times, e.g.: it?a will match ita in italy and ia in 'ian'
- [xyz]: Any of the characters enclosed in the bracket. Ranges can be specified with a hyphen, e.g. [a-c] is the same as [abc], e.g.: [a-c] matches a in 'rage' and c in 'cosmos'
- [^xyz]: Oposite of [xyz] -- this will match any character NOT enclosed in the bracket. You can also use the hyphen to specify character ranges
- \w: Any alphanumeric character or underscore. Same as [A-Za-z0-9_]
- \d: A digit; same as [0-9]
- \D: A non-digit character; same as [^0-9]
- a|b: Either a or b, e.g.: color|colour will match both "The color is blue" and "The colour is blue"
- {x}: Precisely x occurrences of the preceding character, e.g.: e{2} matches ee in 'Xtreeme'
- {x,y}: Between x and y occurrences of the preceding character, e.g.: e{1,2} will match both 'ee' and 'e' in 'Xtreeme'

In order to be able to use the **Format expression** field, you need to use parentheses to determine groupings within the regular expression (to be referred to in the format expression). Using \$1, \$2, \$3, ... in the **Format expression** field refers to the first, second, third, and later parenthetical groupings.

For example, if you have one or more documents with the title starting with 'Home page of...' and you want to change such titles to 'This is the home of...' you'd need to add two Regular Expression rules:

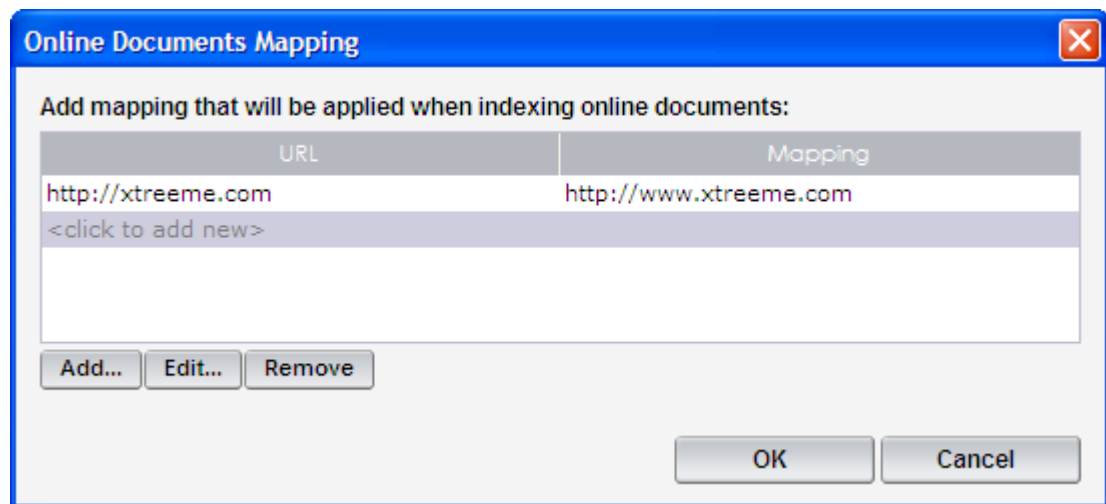
- 1. Regular expression: "^Home". Format expression: "home" -- this rule will replace all titles having 'Home' as the first word with 'home'.
- 2. Regular expression: "(home) page". Format expression: "This is the \$1". This rule will replace the expression "home page" (only if present at the beginning) with "This is the home". Note that \$1 replaces 'home' in the regular expression.

Note that the order in which the rules are executed is important. Reversing the order of the two rules above would not work correctly. Any subsequent rule always takes the result of the previous rule as the input text.

6.13 Online Documents Mapping dialog

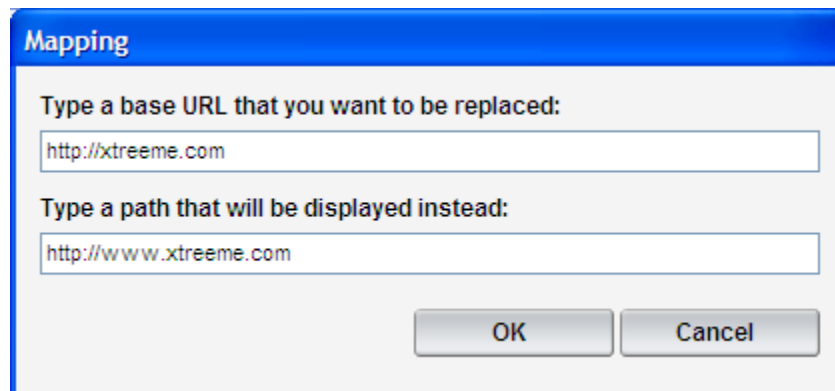
This dialog is opened from [Online Website \(HTTP\) Location dialog](#). You can read more about mapping in the [Mapping](#) topic.

The main table of the window shows added mapping rules. As you can see in the sample below <http://xtreeme.com> will be mapped into <http://www.xtreeme.com>.



Add...

You can add a new mapping entry by clicking the **Add...** button. This opens a dialog shown below, where you can type a base URL that you want to replace and the one that should be displayed instead.



Edit...

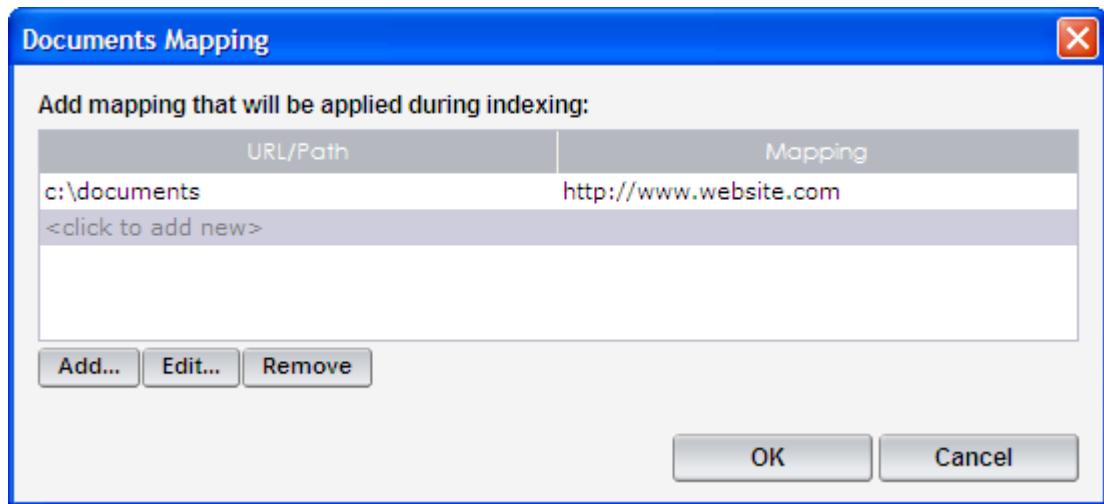
You can use this button to modify a mapping entry. Select a mapping that you want to edit and click this button to open the properties window. You can also double-click on an entry to edit it.

Remove

You can use this button to remove a selected mapping entry.

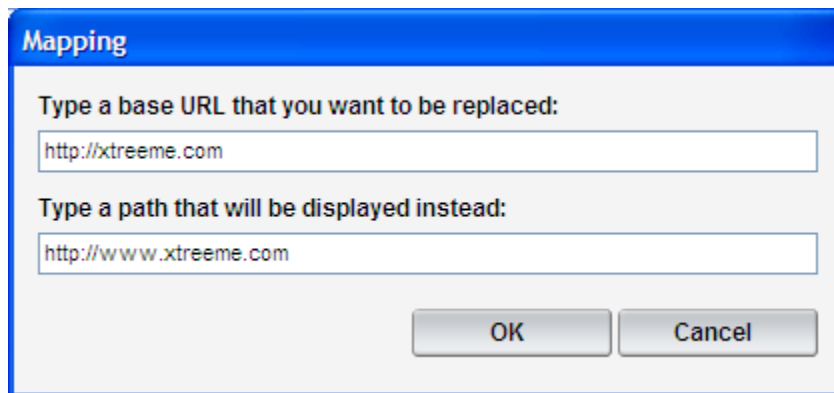
6.14 Documents Mapping dialog

This dialog is opened from [XML Location dialog](#). When documents that you are indexing through an XML file are located on a local or network disk you may need to set up mapping that will display a different path in search results. You can read more about this in [Mapping](#).



Add...

You can add a new mapping entry by clicking the **Add...** button. This opens a dialog shown below, where you can type a base URL that you want to replace and the one that should be displayed instead.



Edit...

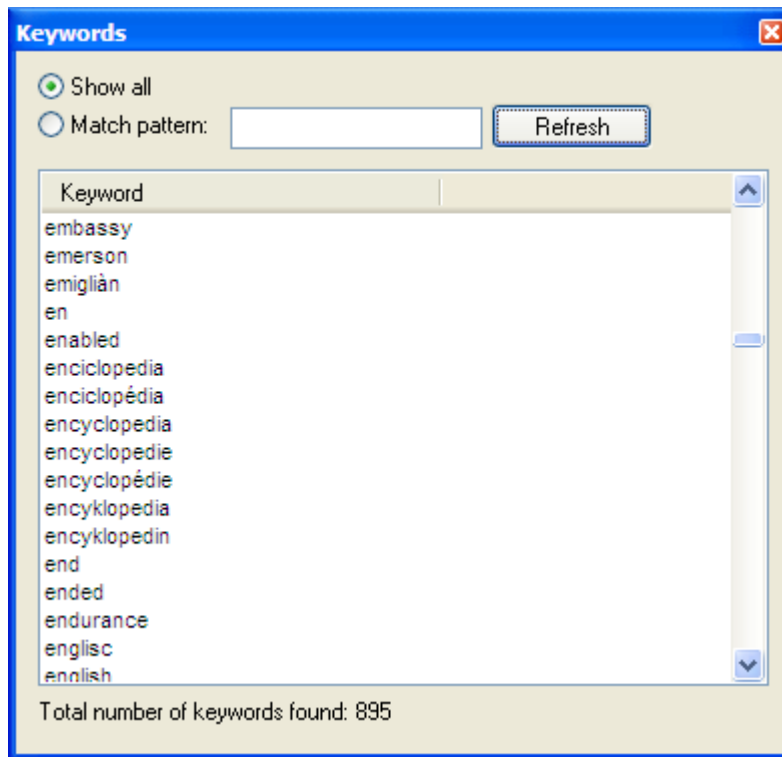
You can use this button to modify a mapping entry. Select a mapping that you want to edit and click this button to open the properties window. You can also double-click on an entry to edit it.

Remove

You can use this button to remove a selected mapping entry.

6.15 Keywords dialog

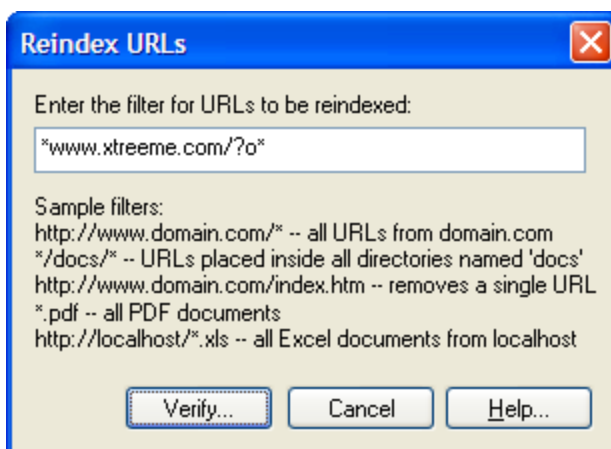
In this window you can view all indexed keywords.



You can choose to show all keywords by selecting **Show all** or display only those that match a given pattern. Note that a pattern can contain * and ? characters. You can show, for example, only keywords matching 'ency*' that in the sample presented on the picture above would display enciclopedia, *encyclopedia*, *encyklopedia* etc.

6.16 Re-index URLs (from pattern) dialog

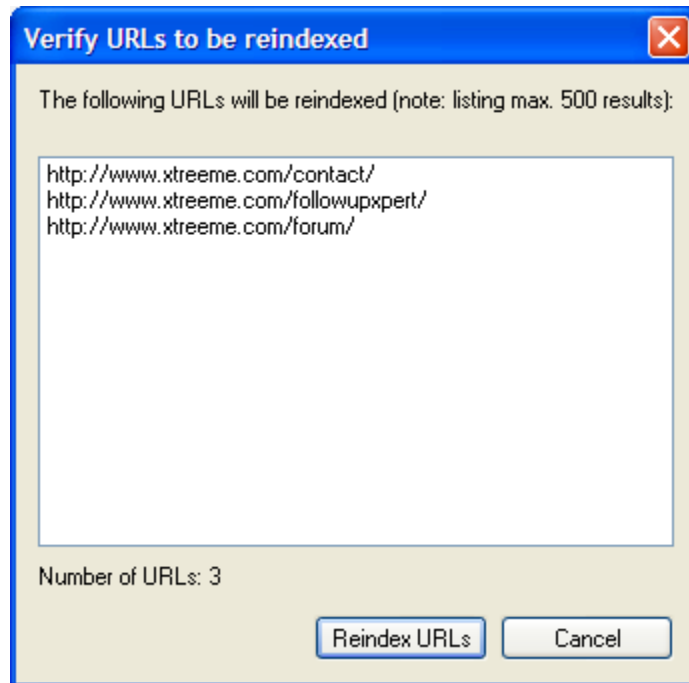
This dialog opens when you choose the **Re-index documents from pattern** option from the **More >>** drop-down in the main window.



In this dialog you can specify which documents should be re-indexed. When you specify a pattern, all documents that match it will be re-indexed. For example, when you type **domain*.

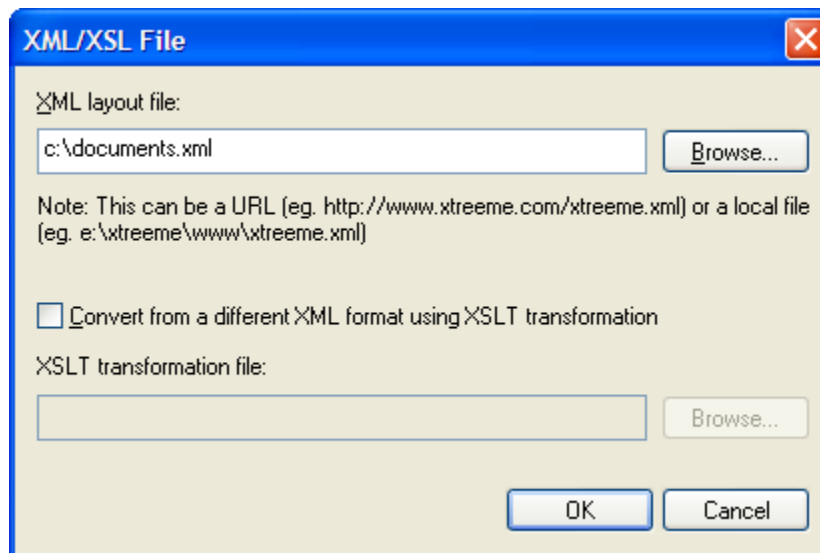
`com/help/*.pdf` all PDF files in the *help* section will be re-indexed.

Once you click the **Verify...** button a dialog with a list of documents matching the pattern is shown (see the picture below). When you are sure that you want to re-index the listed documents you can start by clicking the **Reindex URLs** button.



6.17 Re-index URLs (XML file) dialog

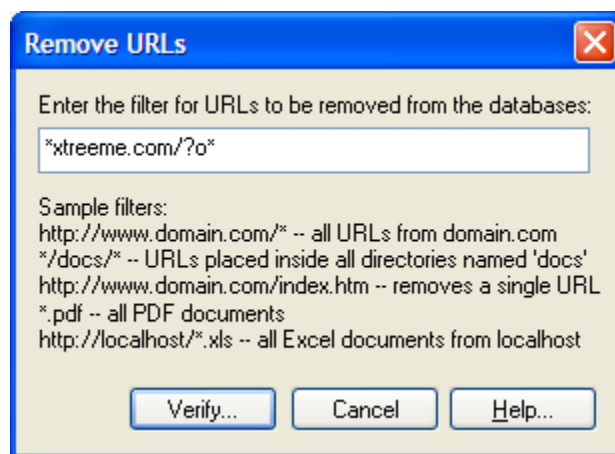
This window opens when you choose the **Re-index documents from XML file** option from the **More >>** drop-down in the main window.



Here you can choose an XML file containing paths of documents that you want to re-index. You can learn more about the format of this XML in the [Indexing documents from XML file](#) topic. If you have a different format of an XML you can use an XSLT transformation by turning on the **Convert from a different XML format using XSLT transformation** option and typing a path to the XSLT file.

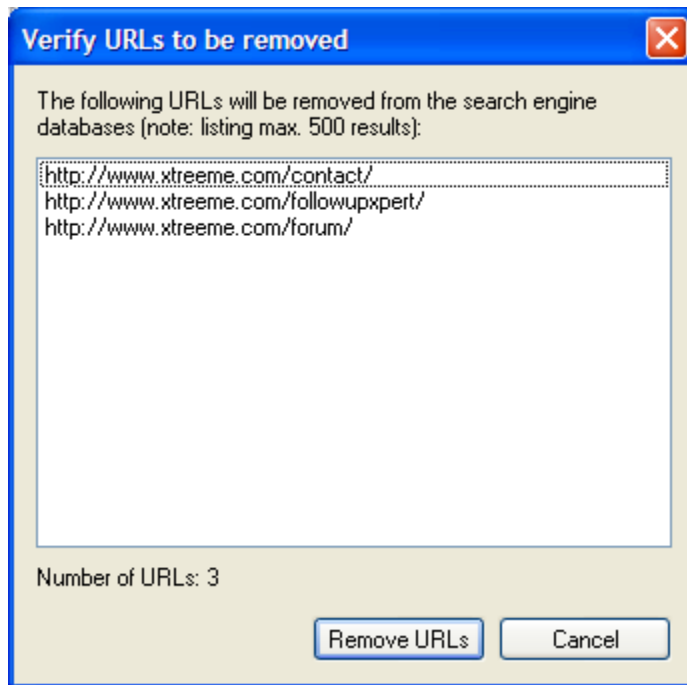
6.18 Remove URLs (from pattern) dialog

This window opens when you choose the **Remove documents from pattern** option from the **More >>** drop-down in the main window.



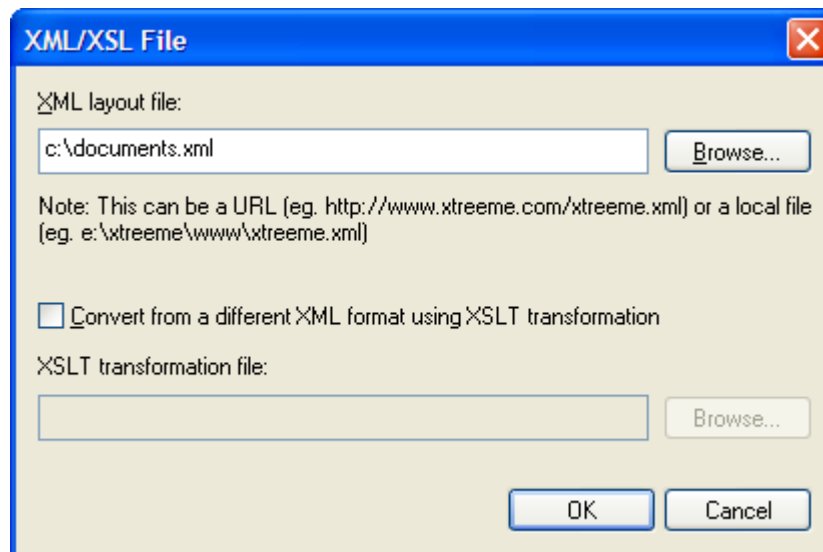
Here you can type a pattern that will match documents that you want to remove. For example, when you type `*domain.com/help/*.pdf` all PDF files in the `help` section will be removed.

Once you click the **Verify...** button a dialog with a list of documents matching the pattern is shown (see the picture below). When you are sure that you want to remove the listed documents you can do this by clicking the **Remove URLs** button.



6.19 Remove URLs (XML file) dialog

This dialog opens when you choose the **Remove documents from XML file** option from the **More >>** drop-down in the main window.

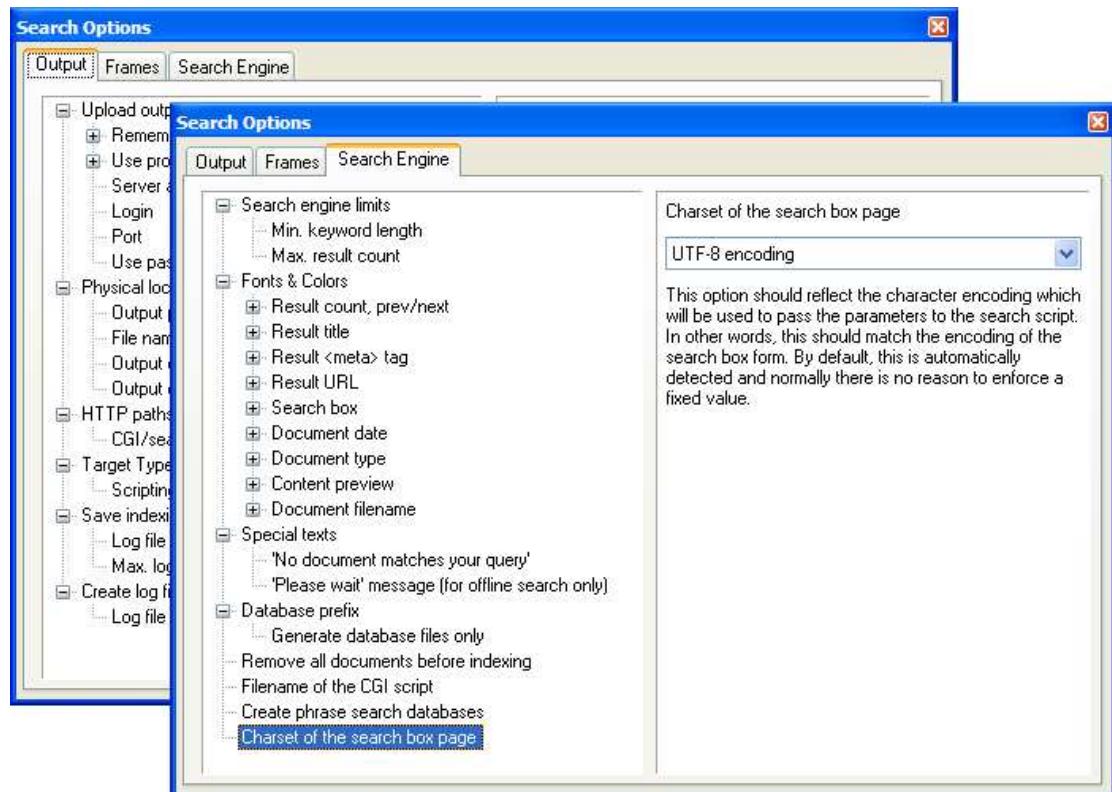


Here you can choose an XML file containing paths of documents that you want to remove. You can learn more about the format of this XML in the [Indexing documents from XML file](#) topic. If you have a different format of an XML you can use an XSLT transformation by turning on the **Convert from a different XML format using XSLT transformation** option and typing a path to the XSLT file.

Note that the removal operation cannot be undone.

6.20 Search Options dialog

This dialog is the place where you set miscellaneous options of the search engine. You can open this dialog by clicking the icon in the program [toolbar](#) or choosing **Options** from the **Window** menu.



This window is divided into three tabs: **Output**, **Frames** and **Search Engine**. Below you will find a detailed options list.

Output

Output is where you specify where the output files should be saved and which search engine type is being generated.

Output / Upload output to an FTP server / Remember password

Select this option if you want Search Engine Studio to store your in and password. Please note that storing your FTP password is insecure!

Output / Upload output to an FTP server / Use proxy server (firewall)

Select this option if you are behind a firewall and cannot connect directly to the FTP server (click the 'Proxy settings' button to specify detailed options)

Output / Upload output to an FTP server / Server address

Specify the address of an FTP server, e.g: ftp.xtreeme.com. Do not include the ftp:// protocol prefix

Output / Upload output to an FTP server / Login

Your login name for the FTP connection.

Output / Upload output to an FTP server / Port

Port for the FTP connection. Usually the default value of 21 will be correct.

Output / Upload output to an FTP server / Use passive mode

This option allows to turn on a mode where client establishes connection. This can be necessary with some firewalls running on your machine which don't allow connections from the outside.

Output / Physical location of generated (output) files

Options in this category are used to set the physical location of files generated by the program. This is either a local/network location (e.g. c:\dir\... or \\server\dir\...) or an FTP location (if using the FTP upload option -- e.g. /home/abc/www/...)

Output / Physical location of generated (output) files / Output page location same as template file used

This option will automatically set the location of the output page according to the imported template file. Using this option will result in the original template file being replaced by a similar page but with output code inserted by the program. You cannot use this option if not using a template file.

Output / Physical location of generated (output) files / File name of the search box page

Specify the name of the output HTML file (default: search engine.html). Please specify file name only without a path.

Output / Physical location of generated (output) files / Output directory for search box page files

Directory where the search box page will be saved, either on local/network disk or (if you selected the option 'Upload output to an FTP server') on an FTP server. If you leave the next option (output dir. for CGI/scripts) empty, all output files including the search scripts and databases will be also saved in this directory. If specifying an FTP directory, you can specify an absolute path (e.g. '/usr/www/users/userName/search') or a path relative to the default folder for the FTP server (usually your home directory; e.g. 'public_html/search'). If saving the files on local/network disk, you can use the 'Browse' button to specify the target directory or enter a directory name such as c:\inetpub\wwwroot or \\MyDomain\www.

Output / Physical location of generated (output) files / Output directory for CGI/ search scripts

Directory where the script and database files will be saved, either on local/network disk or (if you selected the option 'Upload output to an FTP server') on an FTP server. For PHP and offline searches, you should normally leave this option empty or set it to the same directory as in 'Output directory for search box page files'. For Java servlet output, this should be the 'lib' subdirectory of 'WEB-INF'. If specifying an FTP directory, you can specify an absolute path (e.g. '/usr/www/users/userName/cgi-bin') or a path relative to the default folder for the FTP server (usually your home directory; e.g. 'public_html/cgi-bin'). If saving the files on local/network disk, you can use the 'Browse' button to specify the target directory or enter a directory name such as c:\inetpub\scripts or \\MyDomain\www\scripts.

Output / HTTP paths used in output HTML code

Options in this category are used to set paths used on HTML pages to access files (images, scripts etc.) generated by the program. These must be relative or absolute paths, just like URLs specified in the tag, e.g. http://server/dir/ or /dir/ or ../dir

Output / HTTP paths used in output HTML code / CGI/search scripts directory (HTTP path)

Specify the HTTP path pointing to the directory containing the search engine script (the program will save this file in the directory specified in the option 'Output directory for CGI/search scripts'). For PHP and offline outputs you should normally leave this option empty. For CGI outputs, this should normally be set to '/cgi-bin'. For Java servlet outputs, this should normally be set to '/servlet'.

Output / Target Type

Choose the type of the target for the search engine to be generated. When creating a search engine for a website, choose either the PHP target (compatible with 99.9% of webservers) or one of the highly optimized Windows / UNIX (e.g. Linux, FreeBSD) binary CGI scripts -- please consult your ISP if you're not sure whis OS is used by your ISP. You may also create an off-line search engine for your local searches or CD-ROM / DVD distributions.

Output / Target Type / Scripting language wrapper

This option is useful if you're using a CGI script or JavaServlet output types and you want to embed dynamic content on the search results pages using a server-side scripting language such as ASP or PHP. This allows you to e.g. embed ads on the results page based on the keywords being searched. If this option is used the search will be referenced as a script (e.g. ASP, PHP), and then this wrapper script will internally run the actual search script (CGI).

Output / Save indexing/generation log to file

Use this option to save all messages from the log window to a file on disk.

Output / Save indexing/generation log to file / Log file path

Specify location of the log file

Output / Save indexing/generation log to file / Max. log file size (in KB)

Maximum size of the log size in kilobytes

Output / Create log file on server

Use this option to create a log file on server. All search queries performed by visitors will be logged allowing you to analyze the most frequent queries or queries without results. You can use the Log Analyzer application (it comes with Search Engine Studio) to analyze the log file.

Output / Create log file on server / Log file path

Specify location of the log file. You can specify file name only (e.g. log.slg) to create the log file in the default directory for the search script or you can specify a relative or absolute path (e.g. /usr/abc/priv/logs/search.slg). Note that when using an absolute path this is a real path on server and not FTP or HTTP path. Please also note that there is no size limit for the log file. You should download and remove the log file from the server periodically. It is very important to set appropriate read/write permissions for the parent directory of the log file. By default the search script might not be able to create/write to files on the server.

Frames

Frames is where you specify frame-related options.

Frames / Search box frame

This is the name of the frame into which the search box will be inserted

Frames / Result list frame

This is the name of the frame where the search results will appear

Frames / Result target frame

This is the name of the frame where the URLs from the search result list will be opened. You can also enter one of the special window names ("_blank", "_top" or "_parent"), e.g. "_blank" to have the URLs open in a new browser window.

Search Engine

Search Engine is where you specify the looks of the search engine (colors, fonts, etc.) as well as its functionality.

Search Engine / Search engine limits / Min. keyword length

Determines min. keyword length required to add keyword to search database.

Search Engine / Search engine limits / Max. result count

Maximum number of results in query. Increasing this value may result in longer execution time of the CGI script.

Search Engine / Fonts & Colors / Result count, prev/next / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Result count, prev/next / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Result count, prev/next / Font style

Choose font style.

Search Engine / Fonts & Colors / Result count, prev/next / Text color

Choose font color.

Search Engine / Fonts & Colors / Result count, prev/next / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Result title / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Result title / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Result title / Font style

Choose font style.

Search Engine / Fonts & Colors / Result title / Text color

Choose font color.

Search Engine / Fonts & Colors / Result title / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Result <meta> tag / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Result <meta> tag / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Result <meta> tag / Font style

Choose font style.

Search Engine / Fonts & Colors / Result <meta> tag / Text color

Choose font color.

Search Engine / Fonts & Colors / Result <meta> tag / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Result URL / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Result URL / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Result URL / Font style

Choose font style.

Search Engine / Fonts & Colors / Result URL / Text color

Choose font color.

Search Engine / Fonts & Colors / Result URL / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Search box / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Search box / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Search box / Font style

Choose font style.

Search Engine / Fonts & Colors / Search box / Text color

Choose font color.

Search Engine / Fonts & Colors / Search box / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Document date / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Document date / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Document date / Font style

Choose font style.

Search Engine / Fonts & Colors / Document date / Text color

Choose font color.

Search Engine / Fonts & Colors / Document date / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Document type / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Document type / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Document type / Font style

Choose font style.

Search Engine / Fonts & Colors / Document type / Text color

Choose font color.

Search Engine / Fonts & Colors / Document type / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Content preview / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Content preview / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Content preview / Font style

Choose font style.

Search Engine / Fonts & Colors / Content preview / Text color

Choose font color.

Search Engine / Fonts & Colors / Content preview / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Fonts & Colors / Document filename / Font face

You can enter one or more fonts (separated by colons)

Search Engine / Fonts & Colors / Document filename / Font size

Enter font size (e.g. x-small, 10px or 12px)

Search Engine / Fonts & Colors / Document filename / Font style

Choose font style.

Search Engine / Fonts & Colors / Document filename / Text color

Choose font color.

Search Engine / Fonts & Colors / Document filename / Class

You can enter a class instead of defining font properties. The class must be defined on the search box page (in Search Box Builder) or the results page (in Results Page Overview) using the <style> tag.

Search Engine / Database prefix

This option should be set only when creating different sets of databases (so the user can choose between documents from different areas of a web site). When choosing this option you must make sure that the search box contains a database choice option containing the same prefix as entered in this option. You should enter a name containing letters and numbers only (e.g. entering 'products' will result in creation of database files named productsdbxx.sdb).

Search Engine / Database prefix / Generate database files only

If this project is not your main search project but only serves as a way to generate a separate set of database files (with a prefix), you should select this option so that the program doesn't unnecessarily generate duplicate output files (search script, HTMLs etc.). The only output files generated will be .sdb files which should be placed in the same directory as the other .sdb files from the main project.

Search Engine / Remove all documents before indexing

Here you can decide whether the search databases will be cleared from currently indexed content before a new indexing starts. In most cases clearing the database is the most frequently used mode.

On the other hand, when you turn this option off no data will be removed before indexing. In this case, when a file is being indexed, Search Engine Studio checks if this document is already present in the databases and if its content has changed since last indexing to determine, whether it should be re-indexed or not. If so, then the old content related with this document is removed and the document is re-indexed.

In other words, in this mode only those documents that have been modified are re-indexed. This can speed up indexing process significantly but only in the case when a small part of documents changes. If most of your documents are modified, clearing the databases at the beginning will be much faster then updating modified documents one by one during indexing.

Search Engine / Filename of the CGI script

Enter the name of the CGI script file. Leaving this option empty will result in the default filename being generated (e.g. sessearch.php for the PHP version or sessearch.cgi for the CGI version). Please note: you should only change the default value if your web server doesn't properly interpret the default extension suggested by the program (e.g. when the .cgi extension is reserved for Perl scripts).

Search Engine / Create phrase search databases

Select this option if you want the program to create databases which are needed in order for phrase search to work correctly (e.g. the search query "search engine" would return only the documents which contain the word 'engine' following 'search'). If this option is not selected, such queries are treated as if the quotation characters weren't used.

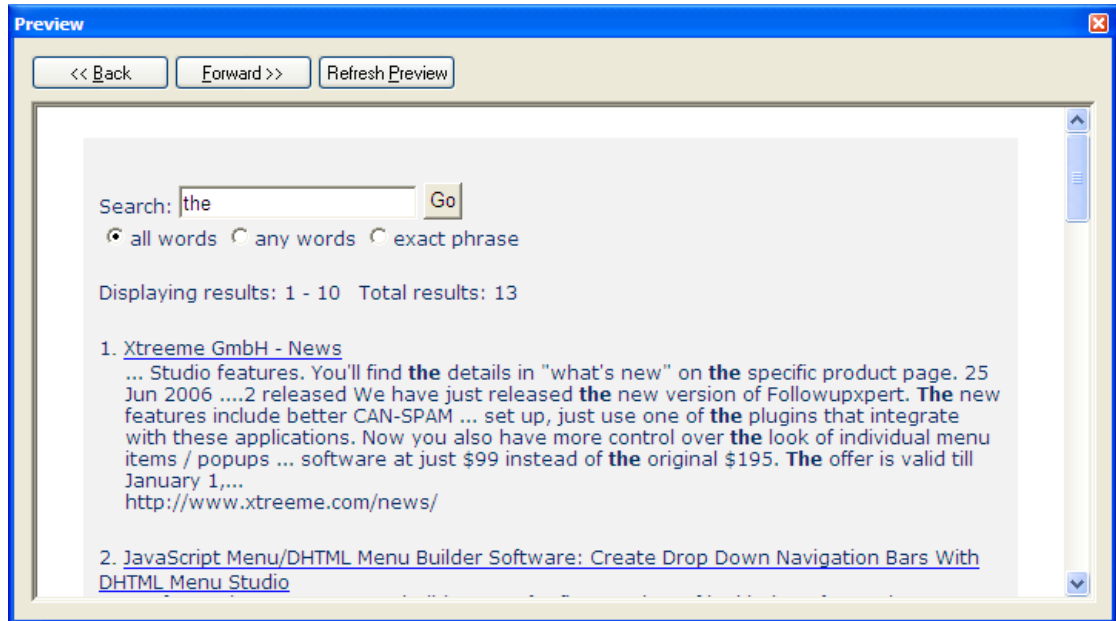
Search Engine / Charset of the search box page

This option should reflect the character encoding which will be used to pass the parameters to the search script. In other words, this should match the encoding of the search box form. By default, this is automatically detected and normally there is no reason to enforce a fixed value.

6.21 Preview dialog


You can open this dialog by clicking the icon in the program [toolbar](#) or choosing **Preview / Search Test** from the **Window** menu.

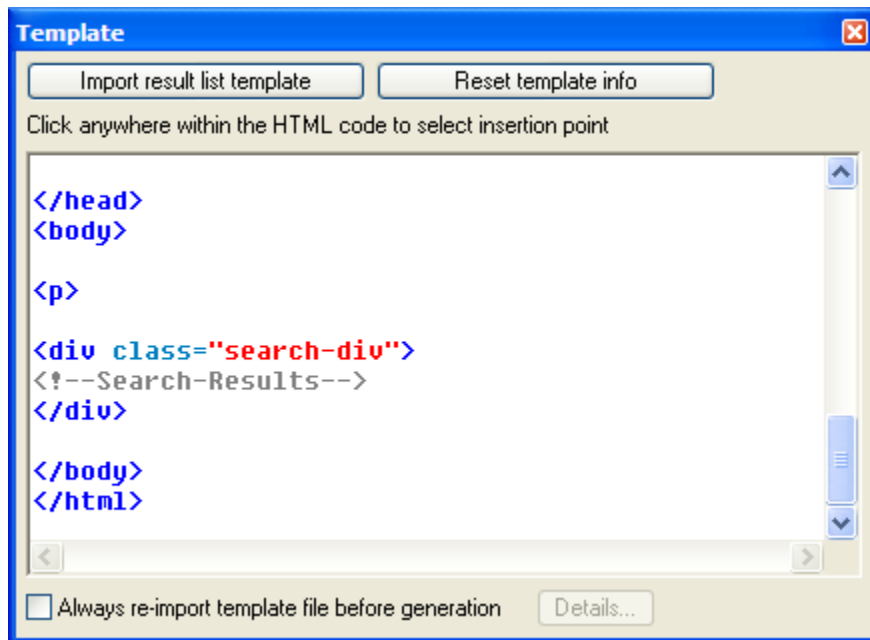
This window shows a preview of the search engine. At the same time, it's also a place where you can test your search engine by entering a search query. The displayed results will be identical to the results of the final search engine, whether it's a CGI script, a PHP script or an offline ActiveX/Java-based version.



6.22 Template dialog

In this dialog you can customize appearance of your search results page by choosing a template file

and specifying the insertion point for the results. You can open this dialog by clicking the  icon in the program [toolbar](#) or choosing **Template** from the **Window** menu.



Import result list template

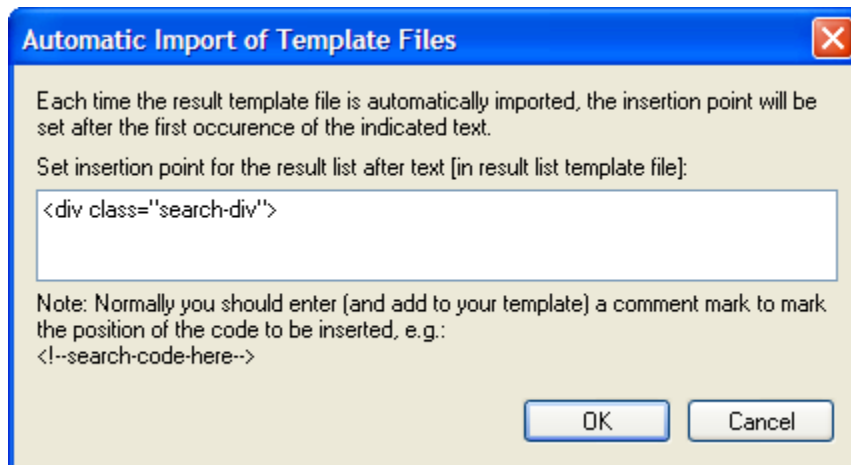
Clicking this button to browse for a file that will be used as a template of your search results list. Once you choose a file its code appears in the main panel of the dialog. Then you should click in the HTML code in the place where you want to insert the results. The `<!--Search-Results-->` tag appears there to mark the spot (see the picture above).

Reset template info

Click this button to reset information about the template file.

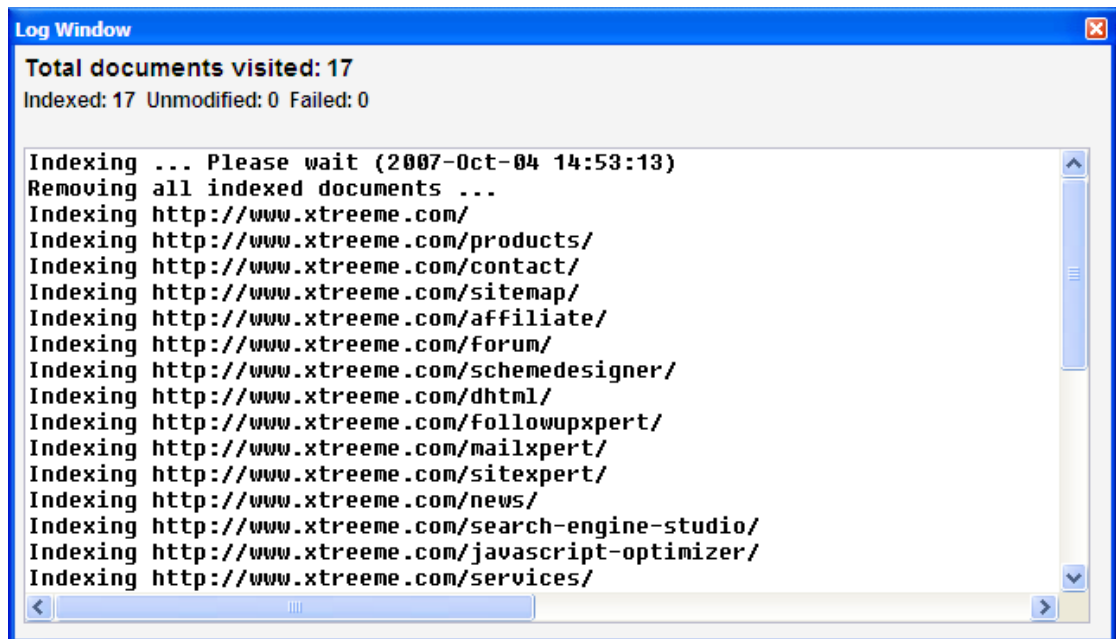
Always re-import template file before generation

You can use this option to automatically re-import a template file. Once you choose a template file, instead of clicking within the code to select a spot for the results you can turn this option on. In this case the list of results will be inserted after the text you can type in the dialog that opens when you click the **Details...** button (see the picture below). From now on, each time you generate your search engine the template file will be re-imported and the insertion point for the results list will be set in the appropriate place.



6.23 Log window

In this window you can follow the progress of indexing documents or generating a search engine. This dialog opens automatically when you start indexing, generation or run another command, like re-indexing or removing documents. You can also open it by choosing **Log Window** from the **Window** menu.



This window displays error, warning and status messages which inform about the indexing or generation process. Please note that in **Options / Output / Save indexing-generation log to file** you can specify to save the contents of the log window to a file. This is especially useful when running Search Engine Studio in the background.


Additionally, in the log you can see the time when indexing was started, when it was finished and the total time of indexing.

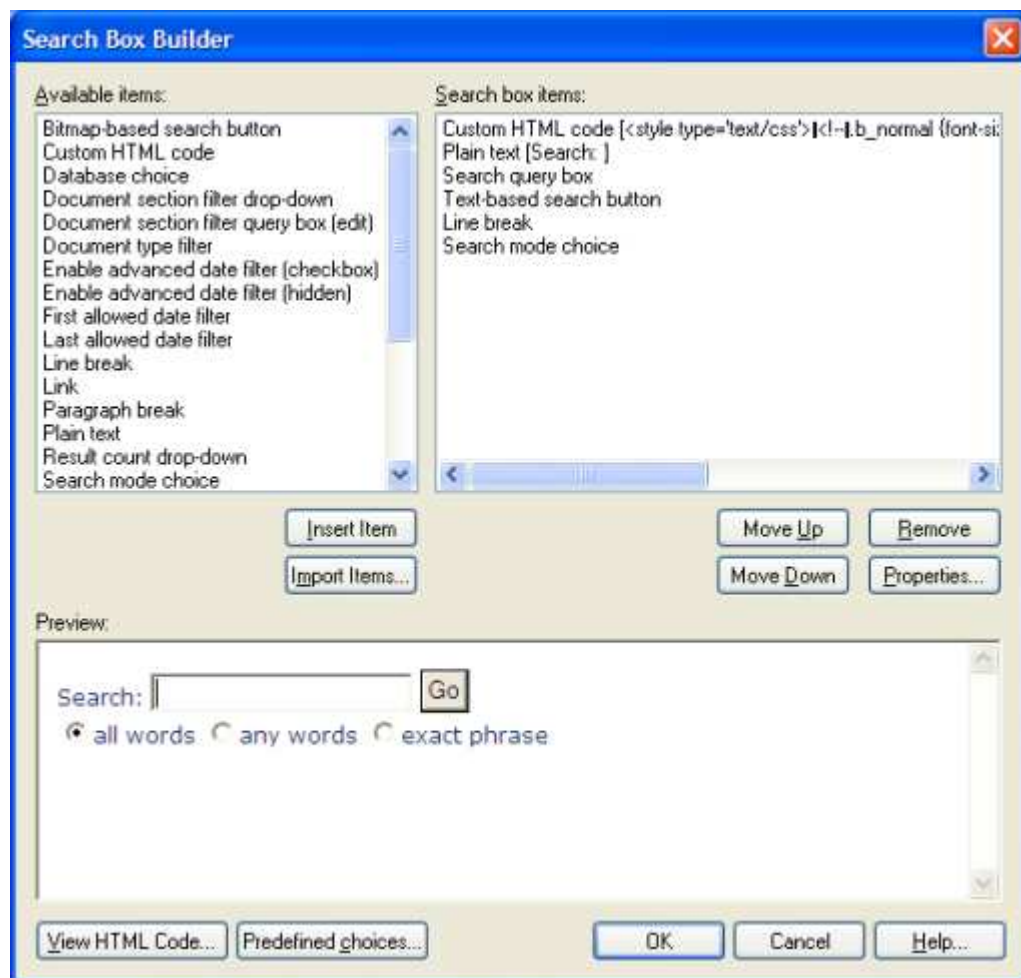
Total documents visited

In this section you can see how many documents were visited and how many of them were indexed or recognized as unmodified. Then unmodified documents is related with the "changed-only" indexing mode. You can read more about this in the description of the **Search Engine / Remove all documents before indexing** option in the [Search Options dialog](#) topic.

6.24 Search Box Builder dialog



You can open this dialog by clicking the  icon in the program [toolbar](#) or choosing **Edit Search Box** from the **Search Engine** menu.



Available items

In this panel you can pick one of the available search box elements. When you want to add an item you can double-click it or select it and click the **Insert Item** button. Once added, the item appears in the **Search box items** list on the right. Please note that one of the available items is a **Custom html code** element which allows insertion of any HTML code in case you want to customize the search box even further, for instance with JavaScript, Flash animations etc.

Search box items

This window lists currently active search box items that create an appearance of your search box. You can edit an item by double-clicking it or by selecting it and clicking the **Properties...** button. Once you open the properties dialog you can change its option and learn more about its functionality. You can also use **Move Up** and **Move Down** to change a position of a selected item. Your changes are automatically applied and visible in the **Preview**.

Preview

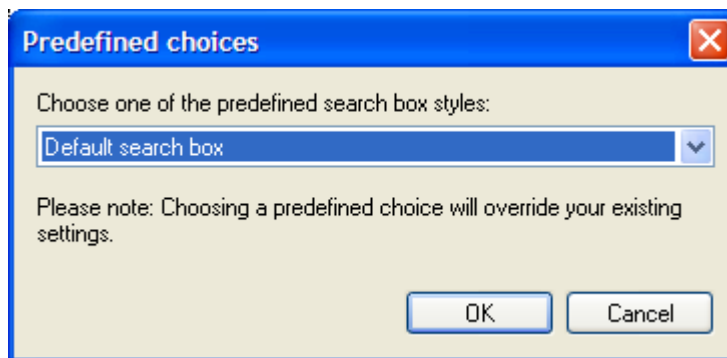
This panel is displayed at all times at the bottom of the window and shows the current look of your search box. When you modify your search box this is reflected in this preview in real-time.

View HTML Code...

At any time you can use this button to see HTML code of the search box. The code reflects your current items and properties configuration.

Predefined choices...

You can use this button to bring up one of the predefined search box styles. The dialog you can see below opens and you can choose one of the styles from the drop-down list.



See also:

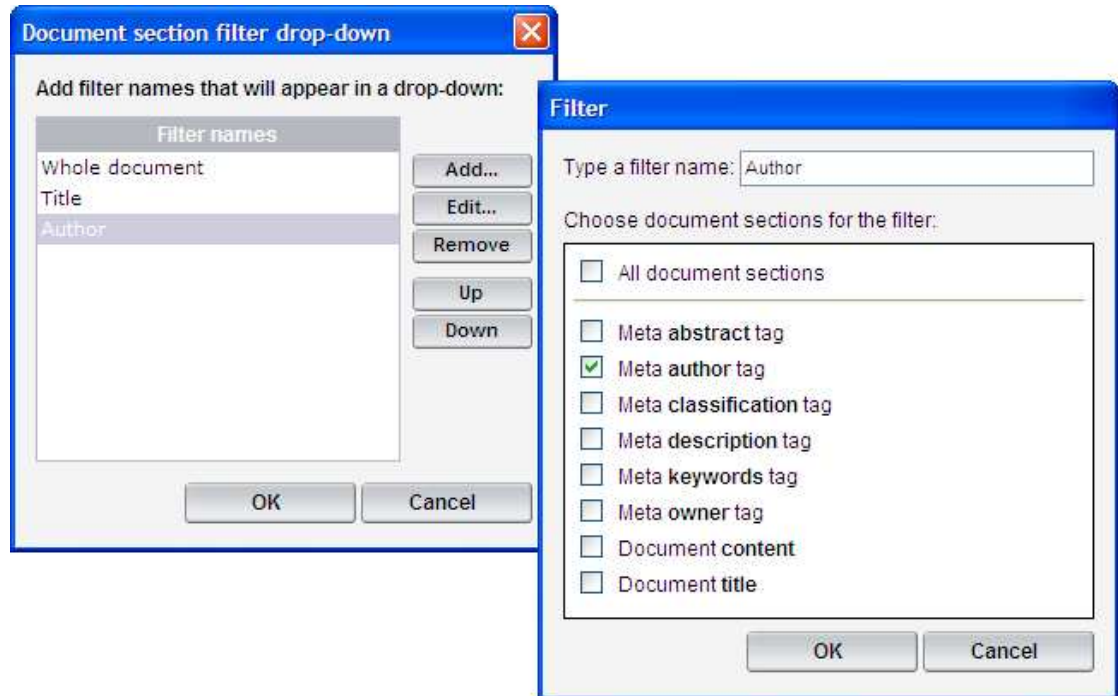
[Quick start tutorial: Adding document section filter to the search box](#)
[Document section filter drop-down dialog](#)
[Document section filter query box \(edit\) dialog](#)

6.25 Document section filter drop-down dialog

This dialog opens when you add a **Document section filter drop-down** item in [Search Box Builder](#).

This is the place where you can add to your search box a filter in a form of a drop-down list that allows the user of your search engine to limit the search scope to one or more document section, for example a document title or the *keywords* meta tag. In this case, in the results list there will be displayed only

those documents that have the query word you are searching for in the chosen section, e.g. title. Of course, you can also choose to search in the whole document which is usually the default behaviour. You can read more about this in the [Indexing document sections](#) topic.

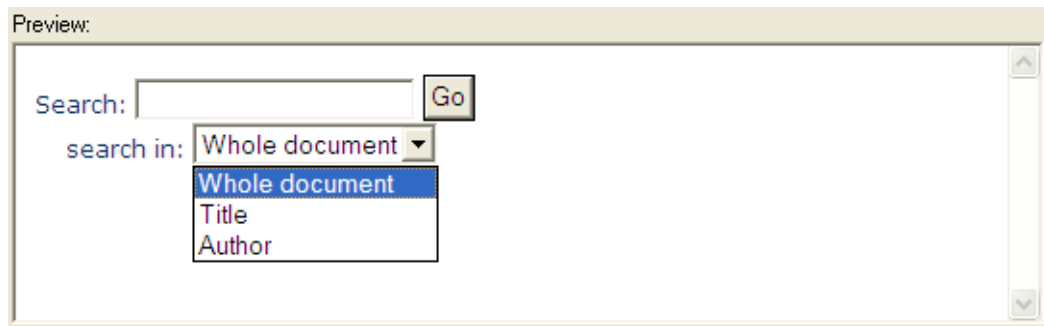


Filter names

In the main window of the dialog (the left picture above) you can see currently added options of the drop-down list. They appear in the final list in the same order as you can see them here. To change the order of an item you should select it and use the **Up** and **Down** buttons.

Add, Edit

You can use these two buttons to add a new filter item or edit an existing one. When you click one of these buttons the filter dialog appears (the right dialog above). Here you can type a name of an option that will appear in the drop-down list and select which document sections will be 'attached' to it. For example, by typing 'Author' and selecting the **Meta author tag** section, this item will appear in the drop-down filter and once chosen the search will be limited to an author meta tag of indexed documents. You can see a sample drop-down containing three options in the picture below.



Remove

When you want to remove an option you should select it and click the **Remove** button.

6.26 Document section filter query box (edit) dialog

This dialog opens when you add a **Document section filter query box (edit)** item in [Search Box Builder](#).

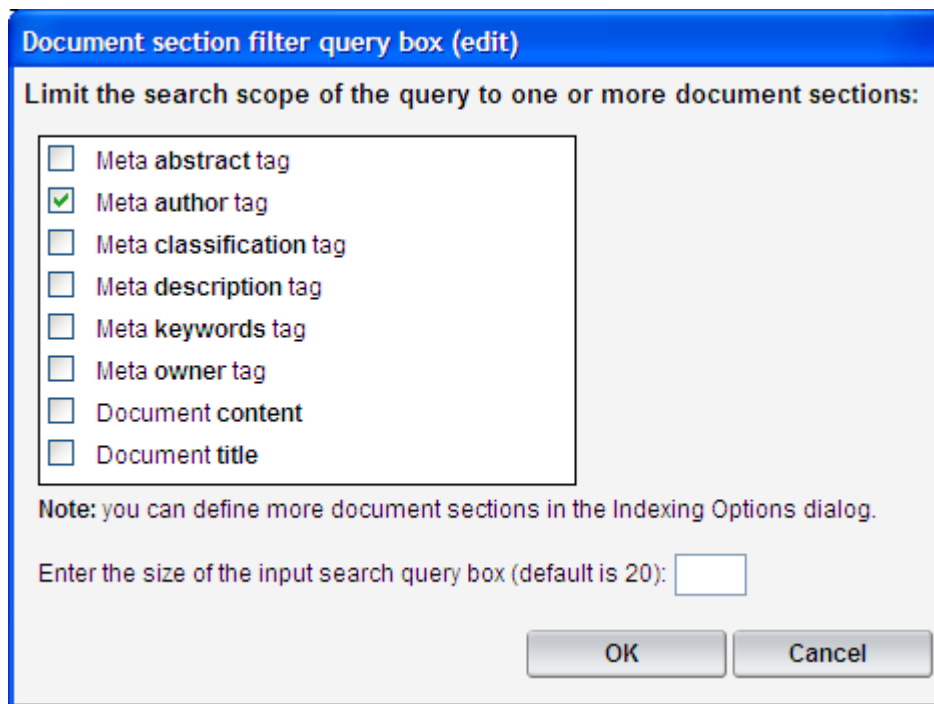
Here you can add to the search box an edit box that will limit a search scope of the query to one or more document sections. For example, when you add this item and choose *Meta author tag section* then once a user types a word and runs a query the search will be limited to an author meta tag of indexed documents. Note that you can also choose more than one section for an edit box.

Note that you can add multiple edits in the **Search Box Builder** window to create a search form (see Picture 1. below). You can read more about document sections in the [Indexing document sections](#) topic.



Picture 1. Preview of a search box containing two *Document section filter query box (edit)* items.

When you add a *Document section filter query box (edit)* element the dialog presented on the Picture 2. appears.



Picture 2.

Limit the search scope...

In the main part of the dialog you can choose which document sections will be assigned to the edit box. You can select one or more options.

Enter the size of the input search query box

Here you can choose a size of the edit box by typing a desired number of sings that will be shown. The default value is 20.

6.27 Result Page Overview dialog

You can open this dialog by clicking the icon in the program [toolbar](#) or choosing **Result Page Overview** from the **Search Engine** menu.

This window allows you to customize the overview of the results page, that is, how the information about results should be presented (total number of results, currently displayed results), how the result navigation should be organized (previous, next and specific result page links) as well as where the actual results will be shown.

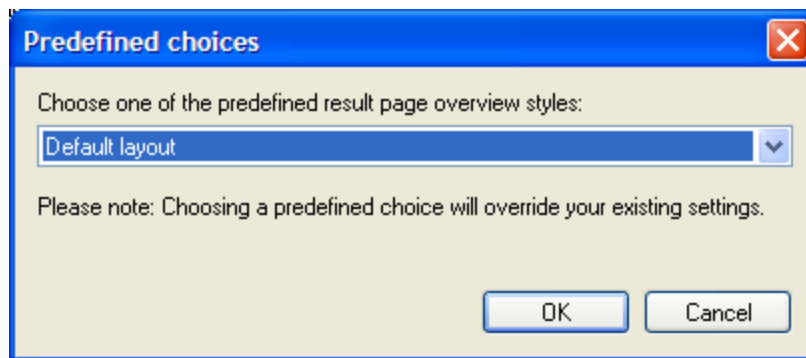
Note that you also can customize appearance of a [single result](#) as well as use a [template file](#) that will be used by the results page, which allows you to fully control the presentation of the results.

Sample preview

This panel is displayed at all times at the bottom of the window and shows the current look of your results area. When you modify your items or add a new one this will be reflected in this preview in real-time.

Predefined choices

You can use this button to bring up one of the predefined layout styles. The dialog you can see below opens and you can choose one of the options from the drop-down list.



Special constructs and conditional statements

In order to include a pre-defined element in the result page overview template, such as: first/last displayed result, total result count, link to previous/next result page etc. you can add it from the **Available items** list. On the other hand, sometimes (e.g. when importing the template from a text file) you might want to manually specify these elements, e.g. inside a **Custom html code** element. In such a case, you can use the following special constructs:


- (resultcode) - inserts code for the search results for current page (please note that the formatting of results can be modified in the 'result formatting' window)
- (resultcount) - total number of results found
- (resultsperpage) - number of results displayed per page
- (firstresult) - index of the first displayed result on current page
- (lastresult) - index of the last displayed result on current page
- (currentpage) - index of the currently displayed result page
- (lastpage) - index of the last result page
- (linkprev) - URL for the previous result page
- (linknext) - URL for the next result page
- (linkpageN) - URL for the Nth result page, please note that you'll need to replace N with the actual number, e.g. (linkpage1), (linkpage2) etc...
- (searchquery) - the query string entered by the user
- (nbpagel[A][B]) - link to a neighboring page. If [A] is set to P, it will be a previous page ([B] determines the offset from the current page, e.g. 1 for previous page, 2 for yet one more page earlier etc). If [A] is set to N ([B] is also the offset), then this will be a link to one of the next pages following currently visited page. Example: if current page is 10, then: nbpagelP1 - link to previous page (9), nbpagelN3 - link to page 13
- (nbpagen[A][B]) - number of a neighboring page. If [A] is set to P, it will be a previous page ([B] determines the offset from the current page, e.g. 1 for previous page, 2 for yet one more page earlier etc). If [A] is set to N ([B] is also the offset), then this will be

the number of one of the next pages following currently visited page. Example: if current page is 10, then: nbpageNP1 - number of the previous page (9), nbpageNN3 - will be represented by 13

Other than the constructs listed above, you can further customize the result overview by adding conditions. You can do this by embedding any HTML code that you want inserted conditionally into a `{{[cond]][code]}` structure. In such a case `[code]` will be outputted only if condition `[cond]` is satisfied. Here are all possible conditions to be used in place of `[cond]`:

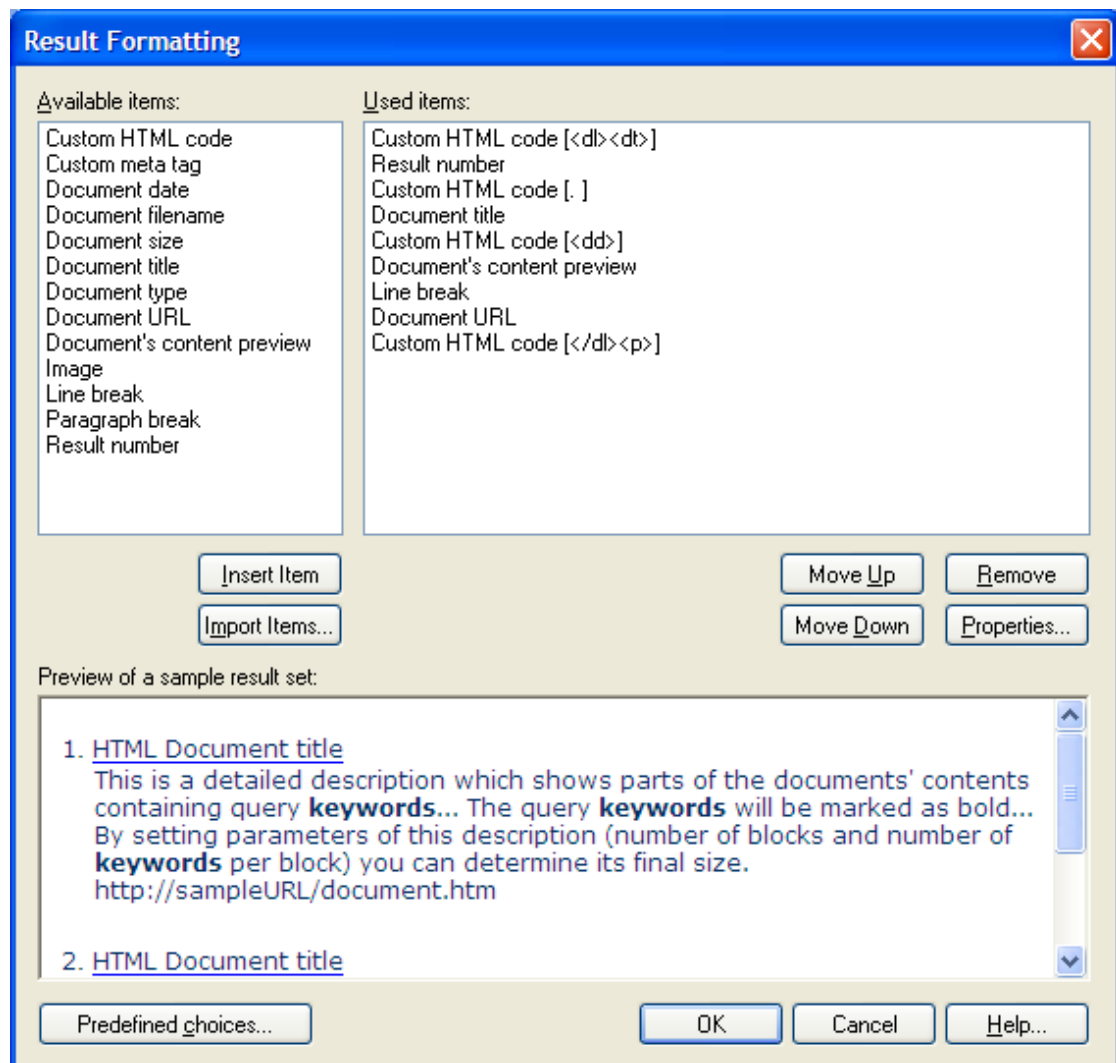
- IFNL[X] - true if current page plus X is not the last result page. In particular, `{{IFNL0|this is not the last page}}` will output the text 'this is not the last page' if the current result page is not the last one. This condition is useful if you want to display links to next results
- IFGT[X] - true if current page number is greater than [X]. For example `{{IFGT2|page 3 or higher}}` will display the text for pages 3,4,5 etc.
- IFEX[X] - true if page number [X] of search results exists
- IFNE[X] - true if page number [X] of search results doesn't exist
- IFPG[X] - true if current page number equals [X]
- IFNP[X] - true if current page number does not equal [X]

6.28 Result Formatting dialog

You can open this dialog by clicking the  icon in the program [toolbar](#) or choosing **Result Formatting** from the **Search Engine** menu.

This window allows you to fully customize the way the results will be displayed. You can display regular HTML items such as images, tables etc. or search-specific elements: document's date, filename and title, document type and size, URL, content preview, result number. You can also display custom meta tag items.

Note that you also can customize appearance of the [result page](#), also by using a [template file](#), which allows you to fully control the presentation of the results.



Available items

In this panel you can pick one of the available elements. When you want to add an item you can double-click it or select it and click the **Insert Item** button, Once added the item appears in the **Used items** list on the right. Please note that one of the available items is a **Custom html code** element which allows insertion of any HTML code in case you want to customize the results even further, for instance with tables, images, JavaScript, Flash animations etc.

Please note that the predefined items, such as document title or URL, will be formatted according to the settings defined in **Options / Search Engine / Fonts & Colors** or a chosen class. If you wish to add these items without any additional formatting, for example in a **Custom HTML code** item, you can use the following special constructs:

- (URL) - document URL
- (URLEx) - extended version of document URL -- can contain additional info (at this time only in case of PDFs; the query string is added to the URL so that the keywords can be easily found inside the PDF document)
- (doccontents) - document's content
- (doctitle) - document's title

- (doctype) - document type, e.g.: PDF / Word / Excel / HTML
- (docdateXXX) - document date where XXX is any text containing date formatting characters: DD = day, MM = month / number, MMM = month / text, YYYY = year
- (docname) - filename
- (resnumber) - result number
- (docsize) - document size

For example, adding a custom HTML code element containing: "(resnumber). (docdateMMM DD, YYYY) (doctitle)" will generate results looking like this:

1. (Jan 01, 2000) First document
2. (Feb 02, 2004) Second document

Used items

This window lists currently active items that create an appearance of the results. Each item has a set of properties which include a conditions section and custom properties of a given item. You can edit properties of an item by double-clicking it or by selecting it and clicking the **Properties...** button. To learn more about the properties of each see the **Properties** section below.

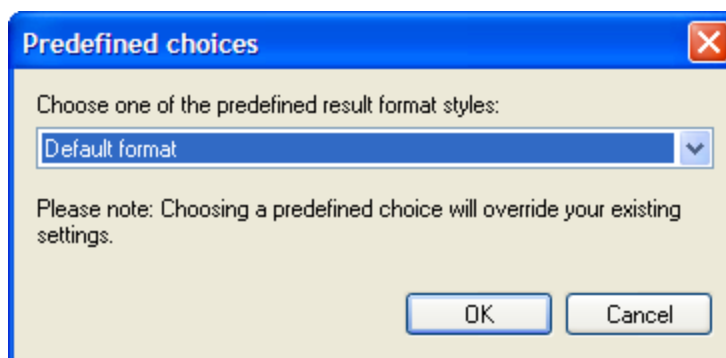
You can also use the **Move Up** and **Move Down** buttons to change a position of the selected item. Your changes will be automatically applied and visible in the preview.

Preview of a sample results set

This panel is displayed at all times at the bottom of the window and shows the current look of your results. When you modify your items or add a new one this will be reflected in this preview in real-time.

Predefined choices

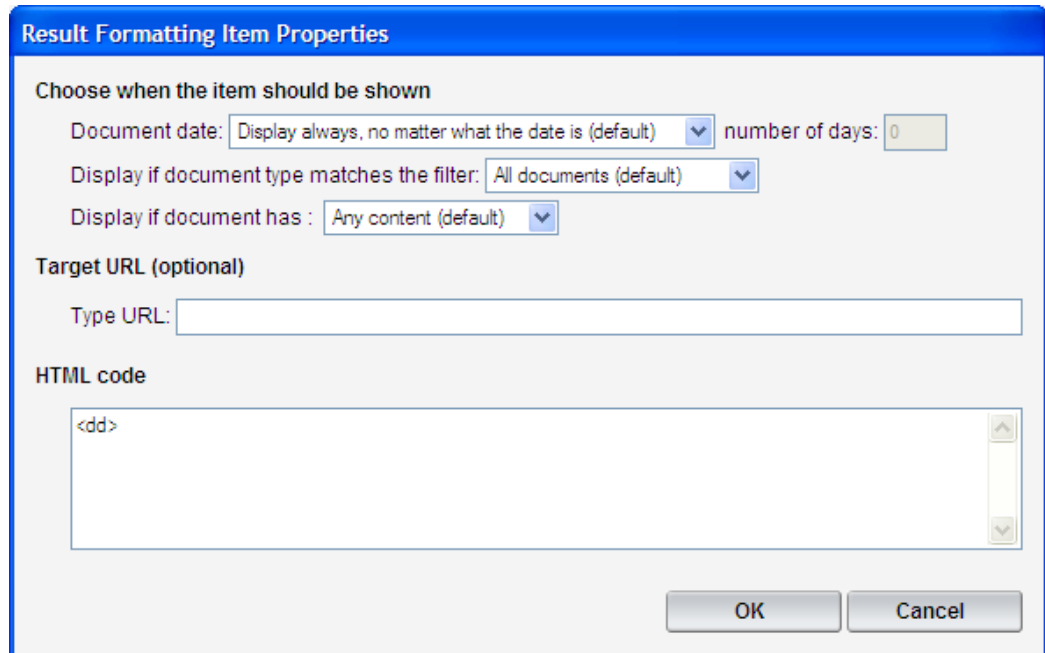
You can use this button to bring up one of the predefined layout styles. The dialog you can see below opens and you can choose one of the available options from the drop-down list.



Properties

When you click this button the **Result Formatting Item Properties** dialog for the currently selected item opens (see below). Properties of each element contain a **Choose when the item should be shown** section where you can decide when a given item is displayed or ignored based on document's date and type (this is especially useful to e.g. display a NEW

icon next to new documents). You can also display an item under a condition that the current result contains a chosen section, for example a title or a selected meta tag. You can see a sample presenting conditions usage in [Quick Tutorial: Conditional displaying items in a result](#).



Picture. Properties window of the *Custom HTML code* item.

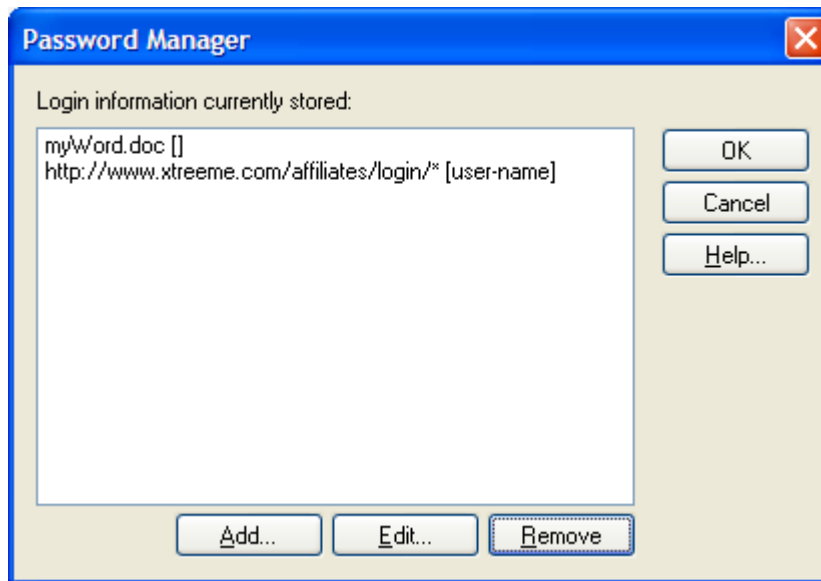
Another common property is **Target URL**. When specified, the given item will be embedded into a link (<a> element) pointing to the specified URL. Note that if you enter (*URL*) or (*URLEx*) the link will point to the URL of the resulting document.

Additionally, some items contain custom options. For example, the *Custom HTML code* item (see the picture above) contains an *HTML code* field where you can type a code that will be displayed.

6.29 Password Manager dialog

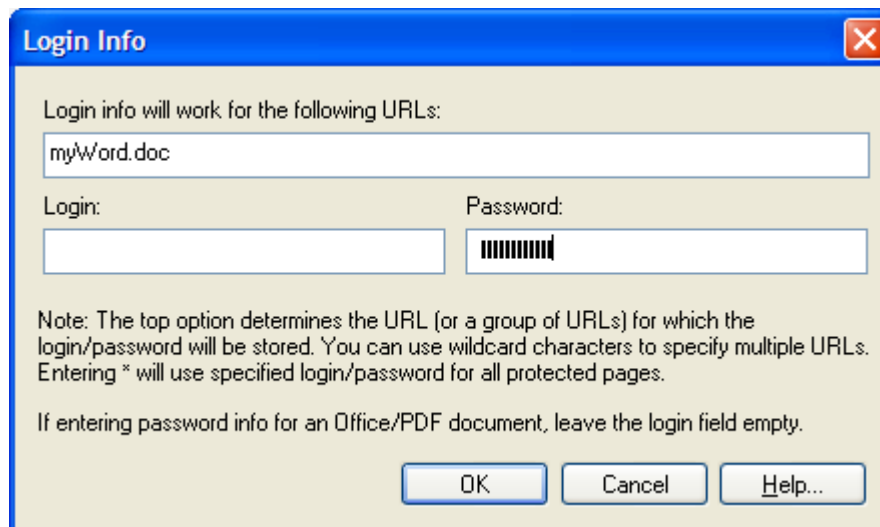
You can open this dialog by clicking the icon in the program [toolbar](#) or choosing **Password Manager** from the **Search Engine** menu.

Password manager is where you can enter passwords used by the indexer's HTTP scan. These passwords will be used to access protected areas of a web site. Please note that you can disable entering password-protected areas in advanced options of **Online Website (HTTP) location** (the **Ask for password when required** option). By default, the program tries all matching login information entered in the password manager and if this fails, a window will pop-up asking for password each time it's required. Note that you can also use *password manager* to allow indexing password protected documents like PDF or MS Office files.



Adding login information

You can use the **Add...** and **Edit...** buttons to add a new item or edit an existing one. Once you click one of these buttons the **Login Info** dialog opens. In the dialog you can specify login information that will be applied during indexing.



To enter login information, you need to enter:

- URL specifier -- this will determine for which URLs the login info will be used, e.g. if you enter *, the login info will be used for all password protected pages (if it fails, the next matching entry in the list will be used); if you enter: `http://www.website.com/private/`, the login info will be used for all web pages located within the private directory on website.com domain. You can also enter a specific location, e.g. `http://www.website.com/dir/`
- Login -- this is the login (user name)

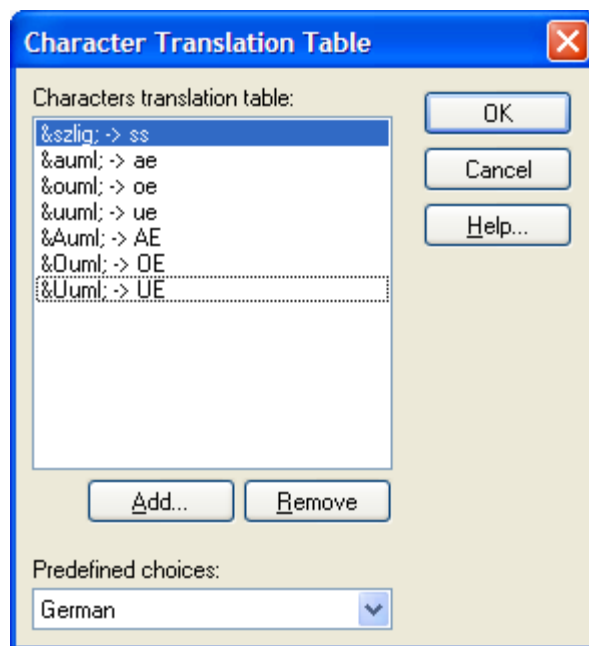
- Password -- password to be used

If entering password information for PDFs or MS Office documents, you can only specify the password and leave the *login* empty. This should be the user password (password needed to open the document).

6.30 Character Translation Table dialog

You can open this dialog by choosing **Foreign Language Support > Character Translation** from the **Search Engine** menu.

This window is useful when creating a search engine either for documents written in a language containing special characters or when creating a search engine interface with modified texts.



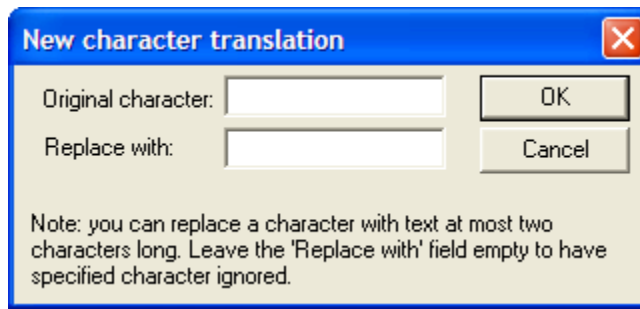
This dialog is the place where you can add translations of special characters, e.g. entering a pair: 'ü' -> 'ue' will ensure that when the user enters 'über' and 'ueber' into the search box, both search queries will be treated in the same way. You can also use this window to tell the program to ignore certain characters (when the target character is unspecified).

Characters translation table

In this panel you can see currently defined character translations that will be applied when running search queries. At any time you can edit an item by double-clicking it.

Add...

Use this button to add a new character translation. Once you click it the window below opens.



In the **Original character** field you can type a character that should be replaced and in the **Replace with** field a text (no longer than two characters) that should be used instead.

Remove

Click this button to remove a selected item.

Predefined choices

This drop-down list contains predefined sets of character translations for a few languages. Once you choose one of the available options these translations appear in the **Characters translation table** overriding previous options. For example, when you choose *German* a set of character translations like 'ü' -> 'ue' or 'ä' -> 'ae' is added.

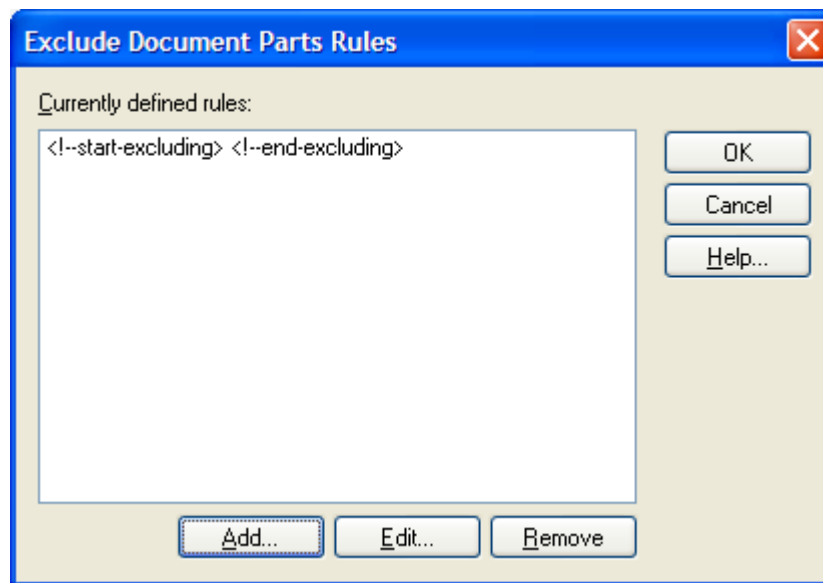
6.31 Exclude Document Parts Rules dialog

You can open this dialog by choosing **Exclude Documents Parts** from the **Database** menu.

There are situations when you need to prevent some document fragments from being indexed. For example, when each page of your web site uses the same footer or a navigation menu it is a good idea to exclude it from indexing. Otherwise, when searching for a word that appears e.g. in a menu that resides on all pages, then all documents will be listed in the results list which is not helpful.

This dialog is the place where you can specify which document fragments you don't want to index. You can add a rule containing a start marker (the text that defines the beginning of a fragment that will be excluded) and an end marker. As the result, the matching fragment will not be indexed or added to the search databases (see the sample below).

Note: exclusion rules should be used only for the main content of a document, that is the <body> tag. If you want to choose which elements of a document's header (title or custom meta tags) are indexed, you can do this in the [Indexing Options](#) dialog.

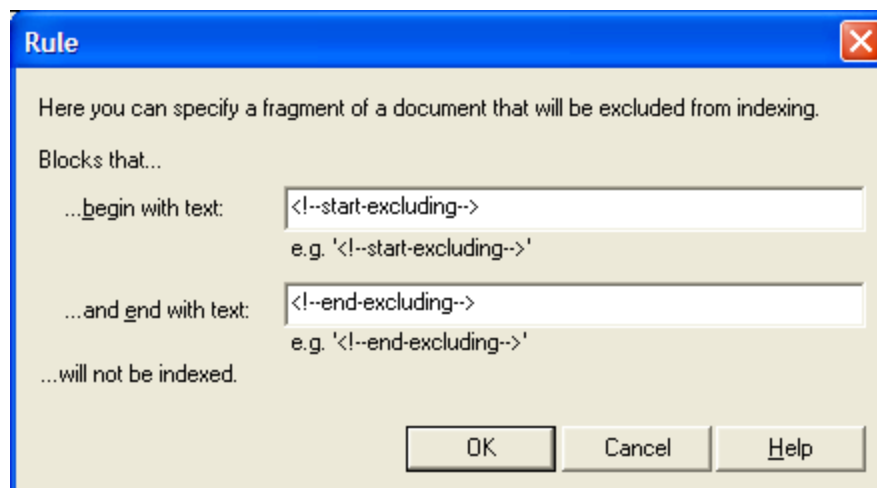


Currently defined rules

This panel lists all rules that will be applied to a document to exclude a matching fragment. The rules will be applied in the same order as they appear in this window.

Add...

Click this button to add a new rule. The **Rule** dialog opens (see the picture below). Here you can type the text that the fragment that you want to remove begins with and the one it ends with. The whole content between those two markers will not be indexed (see the sample below).



Edit...

Click this button to modify a selected rule.

Remove

Click this button to remove a selected rule.

Sample

Let's say that each page of your web site has a company name at the top of a page (see below) but you don't want this fragment to be included into your search engine.

```
<html>
.....
<body>
  Xtreme GmbH
.....
</body>
</html>
```

Add two comment tags around the area that you want to exclude to simplify creating the exclusion rule. It is usually a good idea to use unique comment tags to avoid unintentional conflicts with other parts of a code. In this case we add two comment tags around the company name.

```
<!--title start-->
  Xtreme GmbH
<!--title end-->
```

Now you just need to add a new rule (click the **Add** button) and set the beginning text to `<!--title start-->` and the ending text to `<!--title end-->`. As the result, the company name will not be indexed.

6.32 Proxy Settings dialog

In this dialog you can specify settings of a proxy server that you can use when indexing documents via FTP. Using proxy may be necessary when you are behind a firewall and cannot connect directly to an FTP server. This dialog opens when you click the **Proxy Settings** button in [FTP Location dialog](#).

A screenshot of a Windows-style dialog box titled "Proxy Settings". The dialog has a blue title bar with a close button (X) in the top right corner. The main area is light gray and contains the text "Specify proxy server (firewall) settings:". Below this text are five input fields: "Proxy host name:" (a text box), "Proxy user name:" (a text box), "Proxy password:" (a text box), "Proxy port:" (a text box containing the number "0"), and "Proxy type:" (a dropdown menu with "LOGIN" selected). At the bottom of the dialog are two buttons: "OK" and "Cancel".

Proxy host name

In this field you should type the name of the proxy server to connect through. For example: ftp.xtreeme.com. Do not include the ftp:// protocol prefix.

Proxy user name

User name used to authenticate the client on the proxy server. Note that not all proxy servers require this information.

Proxy password

Password used to authenticate the client on the proxy server. Note that not all proxy servers require this information.

Proxy port

The port number the proxy server is listening on; a value of zero specifies that the default port number should be used.

Proxy type

Here you can choose the authentication type of the connection. The available options are:

- **USER:** This value specifies that the client is not logged into the proxy server. The USER command is sent in the format `username@ftpsite` followed by the password. This is the format used with the Gauntlet proxy server.
- **LOGIN:** This value specifies that the client is logged into the proxy server. The USER command is then sent in the format `username@ftpsite` followed by the password. This is the format used by the InterLock proxy server.
- **OPEN:** This value specifies that the client is not logged into the proxy server. The OPEN command is sent specifying the host name, followed by the username and

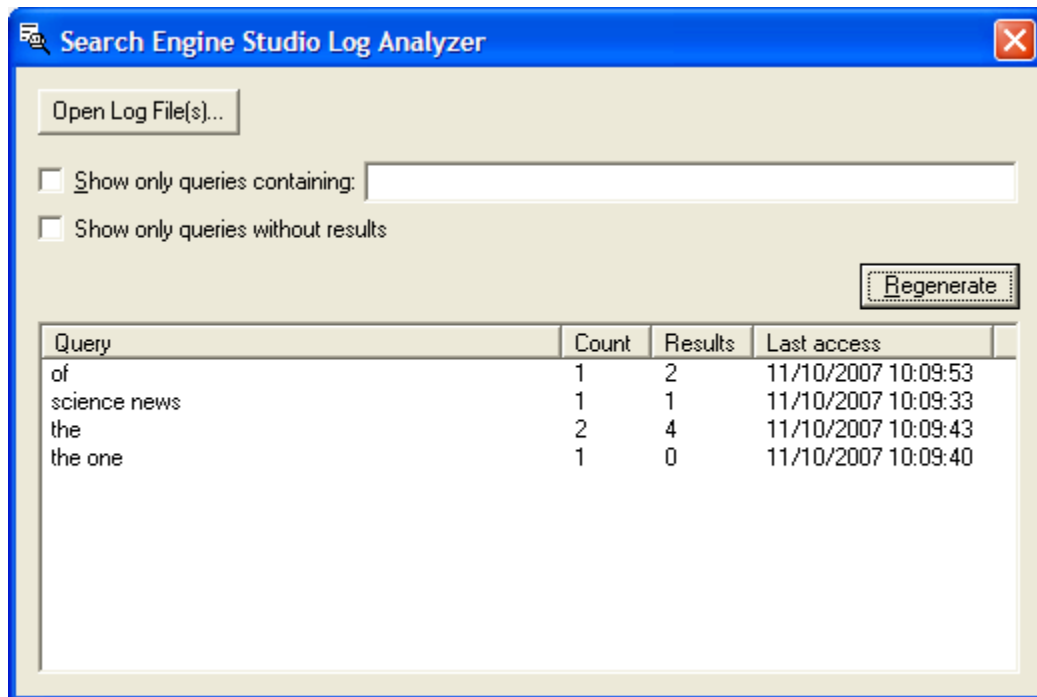
password.

- **SITE:** This value specifies that the client is logged into the server. The SITE command is sent, specifying the host name, followed by the username and the password.

6.33 Log Analyzer utility

You can open this utility by choosing **Log Analyzer** from the **Window** menu (see the [Main menu](#) topic) or from the *Windows Start Menu* by choosing **SES Log Analyzer** from the Search Engine Studio's group.

Before you can analyze a log you need to download the log file from the server to your local or network disk.



Open Log File(s)

Click this button to open one or more log files that you want to analyze.

Regenerate

This button can be used to refresh the report whenever you change any option. Once you click it, the report list at the bottom of the dialog shows log info according to the new settings.

Show only queries containing

This field can be used to limit the report only to search queries containing specific text. For

example, turning this option on and typing *science* will display only those queries that contain the word *science*.

Show only queries without results

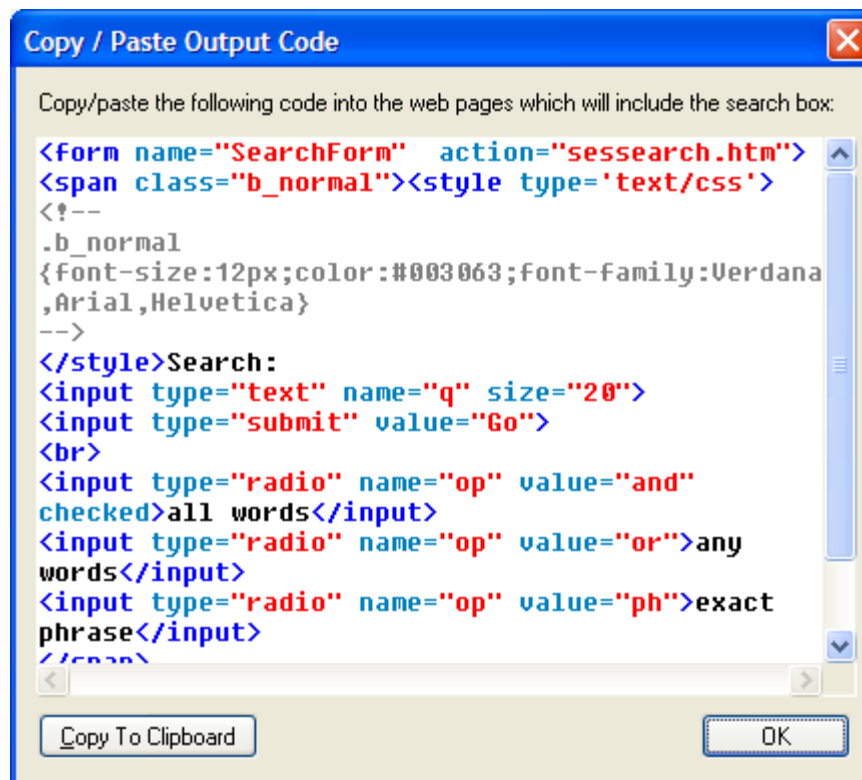
You can also select **Show only queries without results** if you're interested in search queries which didn't return any results.

Report list

The report list shows the search query (the **Query** row), the number of times the search query was entered (**Count**), the number of results returned (**Results**) and the time of the most recent query (**Last access**). You can click on a column's header to accordingly sort the results.

6.34 Copy / Paste Output Code dialog

This dialog is shown when you choose a generation in a copy&paste mode. This window presents an HTML code of the search box. You can copy this code and embed it on your web site to present the search box to your visitors.



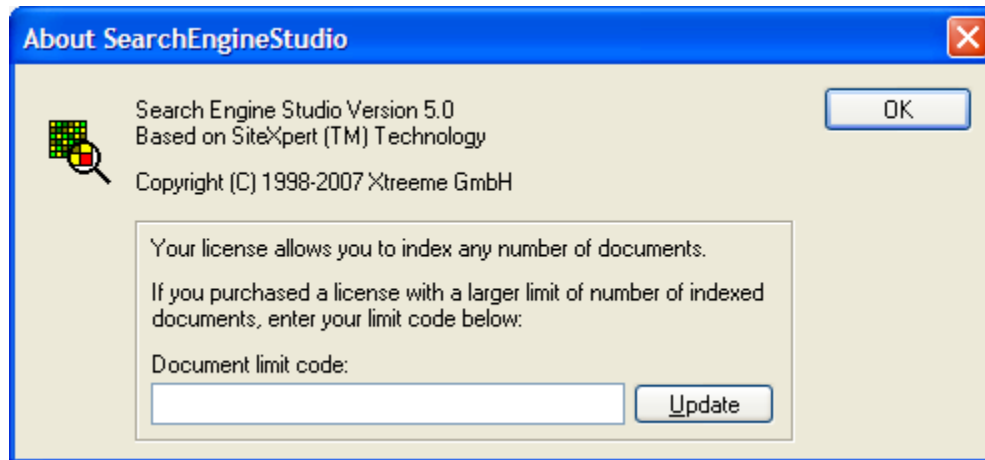
Copy To Clipboard

Click this button to put the search box code to clipboard so you can paste it in another program that you are using to develop your web site.

6.35 About dialog

You can open this dialog by choosing **About Search Engine Studio** from the **Help** menu.

This dialog displays information about Search Engine Studio, including the version of your copy.



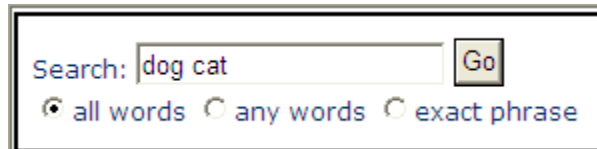
In the panel at the bottom of the window you can see what kind of limit for number of indexed documents you currently have. If you have purchased a license with a larger limit you can enter the registration code in the **Document limit code** edit box and click **Update**. An information about the new limit shows at the top of the panel.

Part



7 Search syntax

When you want to perform a search and find documents you are looking for, you basically type one or more keywords in the search box and click the 'go' button.



Picture. Simple search box.

Additionally, you can also use a number of available options and operators, like wildcards, Boolean operators or a phrase search, to make your search even more precise and get better matching results. This way you can search for documents that...

- contain ALL of the keywords you type in (the AND operator)
- contain at least one of the keywords you type in (the OR operator)
- contain the exact phrase you enter (phrase search)
- do NOT contain a specified keyword (the NOT operator)
- contain a word matching a specified pattern (wildcards)

Below you can find a more detailed description of each operator:

AND

This Boolean operator allows searching for documents that contains all of the specified keywords. For example, a search on `dog AND cat` will return results that contain the words `dog` and `cat` in the contents. Note that when the *all words* options that is available below the search box is selected (see the picture above), the use of the *AND* operator in a query is not necessary.

OR

This operator allows you to search for documents that contain a least one of the specified keywords. For example, a search on `dog OR cat` will yield results that contain the word `dog` or the word `cat` in the contents. Note that when the *any words* option that is available below the search box is selected (see the picture above) you can omit the *OR* operator.

NOT

Using this operator allows to search for documents that don't contain the specified word. For example, searching for `dog NOT cat` will return results that contain the word `dog` but don't contain the word `cat`.

Phrase search

You can create a phrase search by surrounding the keywords that build the phrase with quotation marks ("...") or by selecting the *exact phrase* option. Phrase search allows to find documents that contain the words in exactly the same order as typed in the query. For example, searching for "big

`car big` will return results that contain the word `big` just before the word `car`.

Wildcards (*, ?)

Wildcard characters are a special type of operators that you can use in your query to stand in for one or multiple characters. The `?` wildcard matches a single character and the `*` wildcard matches multiple characters. By using these operators you can search for words starting with, ending with, or containing a given letter sequence. For example, searching for `comput*` will yield results that contain words like `computer`, `computers`, `computing` and searching for `*tribut?` will match words like `contribute`, `attribute` or `distribute`.

You can of course use a combination of the operator presented above to create even more advanced search queries. Below you will find a few samples:

- `(dog and not cat) or (best and pet)` - this matches the documents that contain `dog` but don't contain `cat` or contain both `best` and `pet`.
- `car (door or window)` - this query will match documents that contain the word `car` and at least one of the words `door` and `window`.

See also:

[How to create a more precise search](#)

Part



8 Logging search queries on server

It is sometimes good to know what kind of information people visiting your web site search for and if they can find what they expect. Search Engine Studio gives you a possibility to log all search queries entered by users, together with the number of results that was returned and the date when the query was run.

You can turn on logging search queries in [Search Options dialog](#) in **Output / Create log file on server** and specify a path of the file that will store the information. If you specify a file name only (for example log.slg), the log file will be created in the same directory as the search script. Otherwise, you should specify an absolute path, such as `/usr/abc/private/log/log.slg`, where the log file can be created. Please note that this is a real path on your server, not an FTP or HTTP path.

It is also very important to set appropriate read/write permissions for the parent directory of the log file. Otherwise, the search script might have insufficient permissions to create/write to files on the server. Note that the log has no size limit so you'll be responsible for periodic removal of the log file.

Analyzing log files

Log files can be analyzed using the [Log Analyzer](#) utility that comes with Search Engine Studio. You can open this program by choosing **Log Analyzer** from the Window menu (see the [Main menu](#) topic) or from the Windows Start Menu by choosing **SES Log Analyzer** from the Search Engine Studio's group.

Before you can analyze log information you need to download the log file from the server to your local or network disk. To learn more about viewing the log data go to the [Log Analyzer utility](#) topic.

Part



9 Search engine types generated by the program

Search Engine Studio allows you to create the following types of search engines:

- **Binary CGI script:** This is the recommended target type if you want to create a search engine for a large number of documents. Custom, binary CGI scripts can be used on most web servers (ask your ISP to find out if they're supported on your server), whether Windows (IIS Server), Unix (Apache Server on Sun SPARC, Linux, FreeBSD) or MacOS (WebSTAR or Apache) based. The CGI version of the script is highly optimized, therefore it works fast even when hundreds of thousands of documents are indexed. Please note that some ISPs allow running CGI scripts only in a designated area of a web site (usually inside the 'cgi-bin' directory).
- **PHP script:** This is the recommended target type if you want to create a search engine for small and medium-size web sites. In order for this search engine type to work, you will need to save the generated files on a web server supporting PHP scripts. PHP scripts can be used on virtually any web server. The disadvantage of the PHP script is that it is much slower than the CGI version when a very large number of documents is indexed. Another disadvantage is that not all of features (e.g. wildcard, advanced expression search) are implemented in this version.
- **Offline search:** This search target is useful for generating search engines for CD-ROM / DVD distributions. The search engine is implemented as both ActiveX control and a Java applet -- the search page detects whether the ActiveX version can be used and in case it can't (for non-IE or non-Windows based browsers), it switches automatically to the Java version. This ensures that the offline search will work for practically all up-to-date browsers including IE, Netscape, Mozilla, Opera, Mac Safari.
- **Java servlet:** This server-side search type is a good alternative if your ISP / web server doesn't allow you to run binary CGI scripts but supports Java servlets (e.g. Apache Tomcat or Sun's J2EE).

Part



10 Search Engine Studio Service

Search Engine Studio Service is an NT/XP Service that allows you to schedule automatic updates of a search engine on your machine, even when you are not logged on your system account. This way, once you set the service up, you can have your search engine updated periodically at the specified time interval, on selected days, or whenever the contents of your source documents are modified.

To get more information about the service go to the [Introduction & requirements](#) section.

10.1 Introduction & requirements

Search Engine Studio Service is an NT/XP service that allows you to schedule automatic updates of a search engine on your machine, even when you are not logged on your system account. This way, once you set the service up, you can have your search engine updated, depending on your settings, periodically, on selected days, or whenever contents of selected folders are modified.

You can easily manage installation and running of the service or adding and modifying tasks through the **Search Engine Studio Service Manager** utility. You can start the manager from the **Windows Start** menu by choosing the **SES Service Manager** option from the program's group.

Requirements:

- Search Engine Studio version 5.0 or higher installed on your system.
- Windows NT, 2000, XP, 2003 or Vista for installation and running the service.
- Administrator account for installation of the service.

How to get started

- Go through the [Installation of the service](#) and [Adding & editing tasks](#) sections to get familiar with the service installation and management.
- See the [tutorial](#) to learn about the process of setting up and running the service.

Note: Search Engine Studio Service is sold separately and is not a part of the Search Engine Studio package. For purchase related info go to <http://www.xtreeme.com/search-engine-studio/order/>

If you don't wish to use the service, you can still use the standard Windows scheduler for simple tasks (this will not allow you to e.g. update the search automatically whenever a new document is added, but will allow periodical updates). Please see this [FAQ question](#) for more info. You can also read more about automatic updates in the [Automatic updates of search engine databases](#) topic.

10.2 Why should I use the service?

Search Engine Studio is a program that allows you to create a complete search engine: index a set of documents, define appearance of the results page and generate it to your internet or intranet site. Once you have all necessary indexing, appearance and output options defined you can run the generation manually in the Search Engine Studio main window. You usually repeat the generation every time you want to refresh your search engine, for example when the contents of your documents changed.

Search Engine Studio Service is a utility that allows you to set up automatic updates of your search

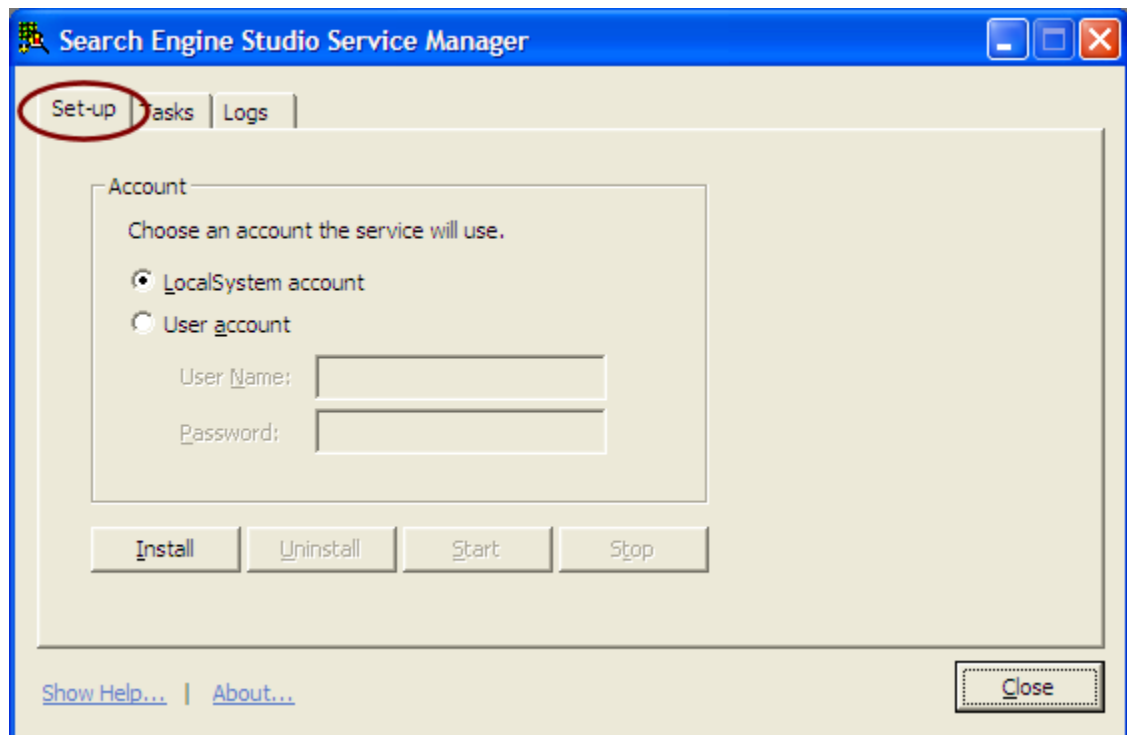
engine that will run even when you are not logged on your system account. By choosing Periodical, Day Scheduled or Event task you can set up an automatic update that will run at the specified time interval, on selected day or whenever the content of your documents changes.

You can learn more about this in the [Automating search engine updates with Search Engine Studio Service](#) tutorial.

You can learn more about the process of creating a search engine in the [Steps to create a search engine](#) topic.

10.3 Installation of the service

Before the service can be started you have to install it on your system. You can do this in the **Set-up** section of **Search Engine Studio Service Manager** (see the picture below). Note, that first you need to run the set-up file to copy the service files onto your disk.



Follow the steps below to install the service:

1. Start the **Search Engine Studio Service Manager** (see the note for Vista user below).
2. Select the **Set-up** tab.
3. In the **Account** section choose the appropriate option that you want the service to use. In most cases, if you want to generate and store your search engine on a local computer, you can select the **LocalSystem** option. On the other hand, if you want to automatically upload your search engine to an FTP server or perform other operations that require network permissions you should use the **User account** option, since the LocalSystem account doesn't have a network

access privileges. In this case you have to type a user name of the account that will be used and the password.

4. Click the **Install** button.

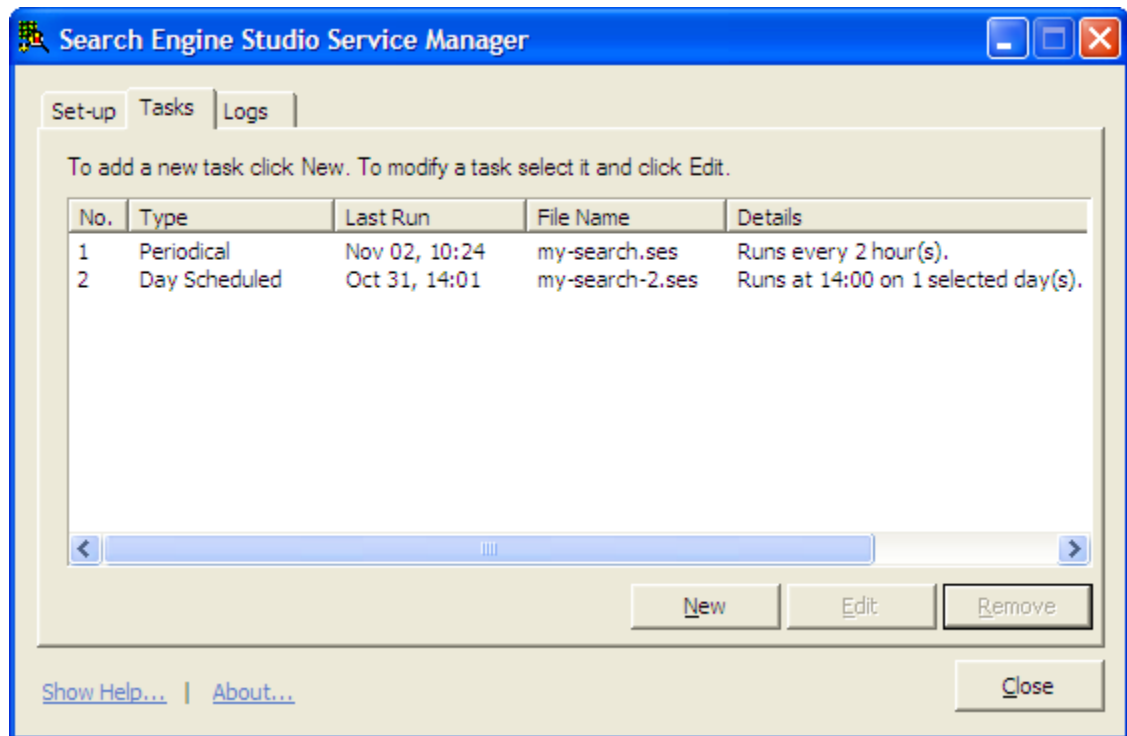
Note for Vista users: Because of new security options introduced in Windows Vista you should start the manager with administrator privileges. You can do this by right-clicking the **SesServiceManager.exe** file (or the shortcut in the program's group) and choosing the **Run as Administrator** option. Otherwise, the application will not have the permissions necessary for installation and running the service.

10.4 Adding & editing tasks

In this section you will learn how to add tasks to perform an automatic updates of your search engine. You can add, modify and remove tasks in the **Tasks** section of **Search Engine Studio Manager** (see the picture below).

There are three types of tasks available in Search Engine Studio Service:

- Periodical - allows you to run your search engine generator at the specified time intervals, for example every 12 hours.
- Day Scheduled - in this task you can select days and time when you want to refresh your search engine. For example, you can specify to update your search engine every Tuesday and Friday at 22:00.
- Watch - this task allows you to watch specified folders and start the search engine generator when their contents changes. When a file residing in one of these folders is modified, added or removed, the update will start. Note that this task works only with local and network folders.



Follow the steps below to add a task:

1. Go to the **Tasks** tab and click the **New** button.
2. In the **New Task** window that opens choose a task that you want to add and click **OK**.
3. In the **Task Properties** dialog you should select a Search Engine Studio document (.ses file) that you want to automate and select the execution priority level. In the properties dialog for the **Periodical** task you can also set up a time interval at which the update will run and for the **Day Scheduled** task you can choose days and time of an update.
4. When you are finish click the **OK** button. The new task entry appears in the main panel of the **Tasks** tab.
5. Once you are finished adding tasks you can start the service by clicking the **Start** button in the **Set-up** tab.

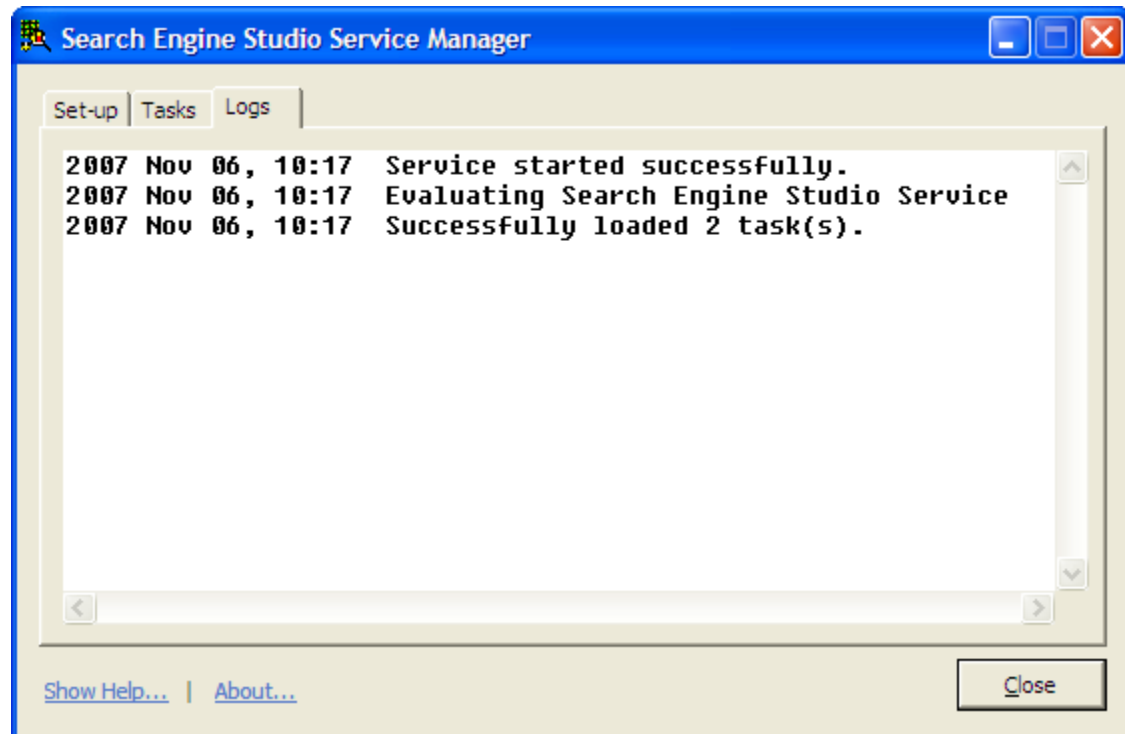
If you want to modify a task settings, follow the steps below:

1. Double-click a task that you want to modify or select it and click the **Edit** button.
2. In the **Task Properties** window that opens you can make your modifications.
3. Once you are finished and click **OK** the changes will be applied.

Note: You cannot add, edit or remove tasks while the service is running. If you want to modify your tasks you have to stop the service in the **Set-up** tab and after you finish doing modification start it again.

10.5 Viewing the log

In the **Logs** tab of Search Engine Studio Service Manager you can view the service activity. This includes messages about running the service, starting and finishing tasks or information about any problems that occur when they are running.



10.6 Tutorial: Automating search engine updates with Search Engine Studio Service

In this tutorial you will set up an automatic updates of a search engine by using Search Engine Studio Service.

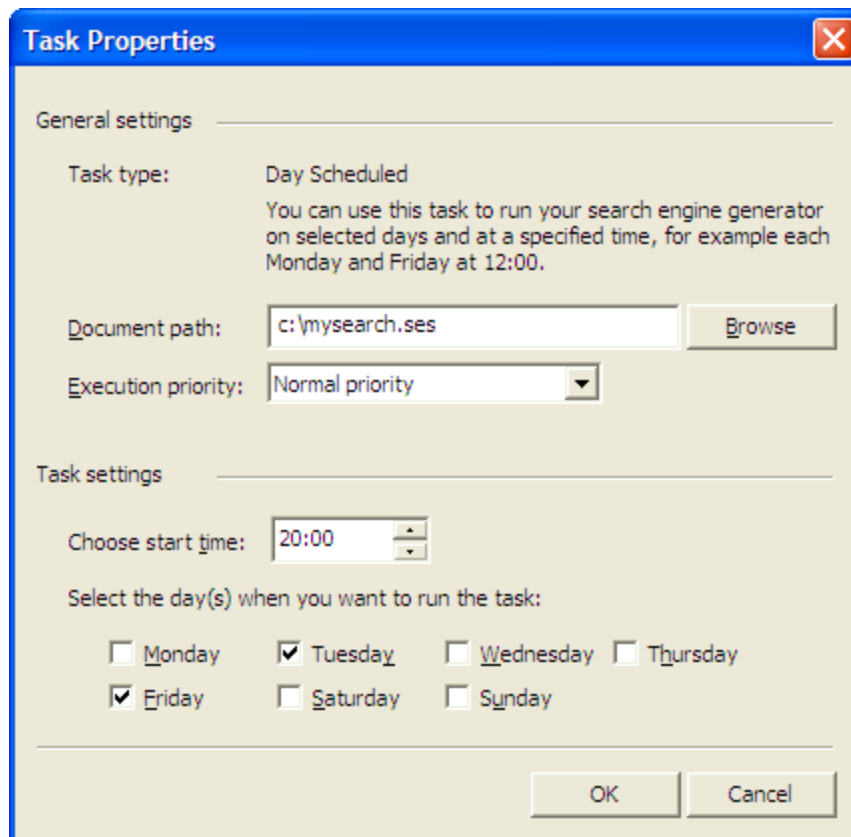
Note: before you can add and run tasks the service has to be installed on your system. You can learn more about this in the [Installation of the service](#) topic.

Let's say that you have already prepared a Search Engine Studio document where you have defined a location of the documents for indexing, you have chosen appearance of your search results page and set up all generation options, like search engine type, output folders, uploading etc. Let's say that the document is saved as *c:\mysearch.ses*.

Now, you will set up an automatic updates for this search engine. Let's say that you want to update it twice a week, on Tuesday and Friday at 20:00.

Follow the steps below:

1. Start **Search Engine Studio Service Manager** and choose the **Tasks** tab.
2. Click the **New** button to add a task. In the **New Task** window that opens choose the **Day Scheduled** option and click **OK**.
3. The **Task Properties** dialog opens (see the picture below). In the **Document path** field, type a path to the .ses document that you want to automate, in this case `c:\mysearch.ses`.
4. For the sake of the example, leave the default value in the **Execution priority** drop-down.
5. In the **Task settings** section at the bottom of the window select **Tuesday** and **Friday** and set the **Choose start time** option to 20:00.



6. Once you click **OK** the new task entry appears in the tasks list.
7. Next, go to the **Set-up** tab and click the **Start** button to run the service.
8. As you can see in the **Logs** tab the service is started and the tasks initialized.
9. Now, every Tuesday and Friday at 20:00 your documents will be automatically re-indexed and the search engine will be regenerated according to the setting specified in the `c:\mysearch.ses` document. Later on, you can view the log to learn about running the task and if the generation was successful.

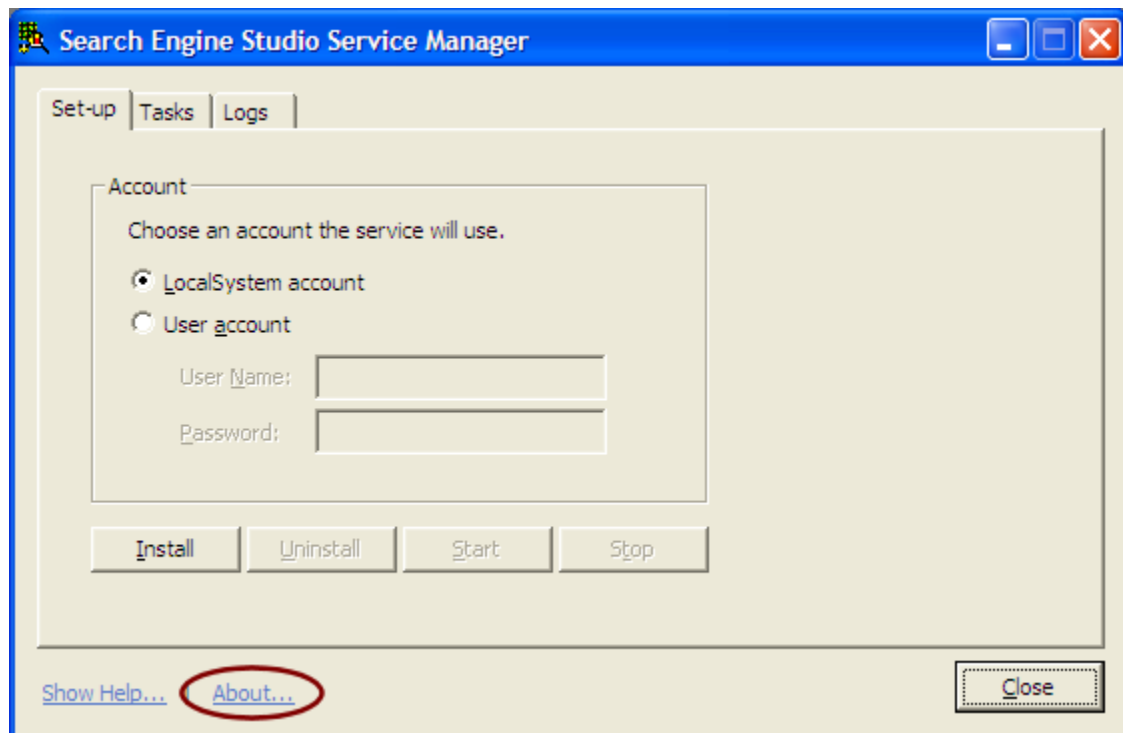
You can learn more about creating a search engine in Search Engine Studio in the [Steps to create a search engine](#) topic.

10.7 How to enter the registration key

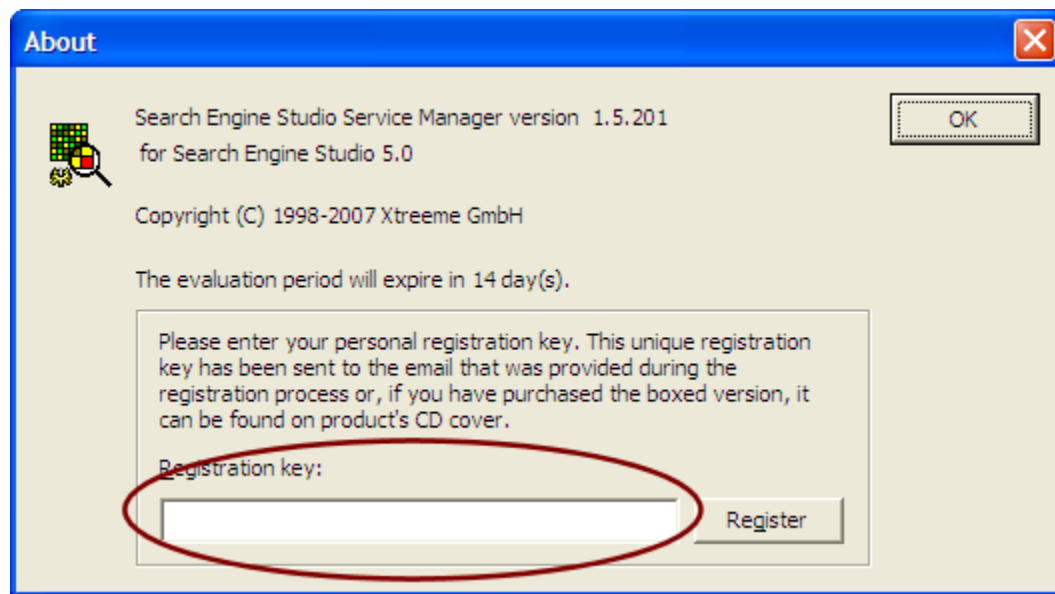
Once you purchase Search Engine Studio Service you receive an email with a unique registration key to the email address specified during the buying process.

Follow the steps below to enter the registration key:

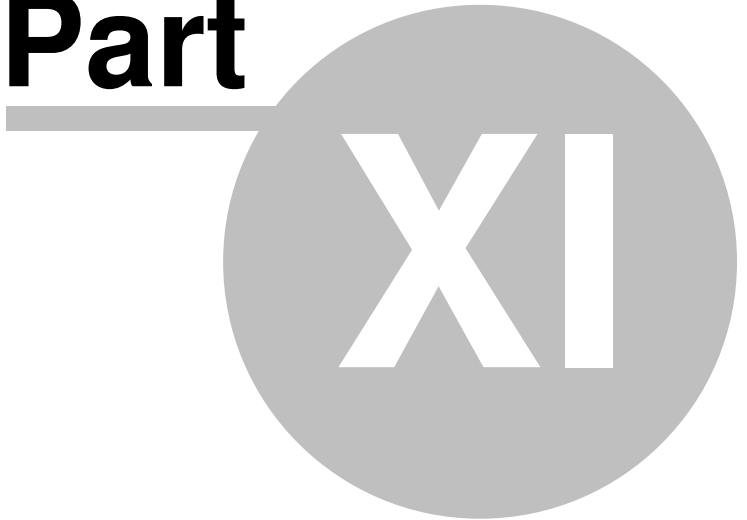
1. Open **Search Engine Studio Service Manager** by choosing the appropriate shortcut from the program's group in the Windows Start menu.
2. Open the program's **About** dialog by clicking the **About...** link on the main window (see the picture below).



3. In the dialog that opens enter your registration key in the field at the bottom of the window and click the **Registration** button. Now your copy of the service is registered.



Part



11 Advanced topics

This section contains topics related with advanced issues of Search Engine Studio.

Click on of the topic in the panel on the left.

11.1 Search script parameters

The search script accepts the following parameters passed using GET or POST methods. Note that only the 'q' parameter is required:

- q: this is the search query
- sec: this parameter will filter out (not list in the results page) all documents whose URL does not start with the value of this parameter, e.g. specifying: 'sec=http://www.xtreeme.com/sitexpert/' will only search for documents within the 'sitexpert' directory of xtreeme.com.
- strt: specifies the first result to be listed (this value is zero-based, meaning that 0 stands for the 1st result, 1 for the 2nd etc.)
- cnt: specifies the max. number of results displayed on a single page
- sdy, sdm, sdd: specifies the year (sdy), month (sdm) and day (sdd) which will filter out all documents created (or modified) earlier than the given date, e.g. 'sdy=2003&sdm=3&sdd=1' will only list documents created on or after 1st March 2003.
- edy, edm, edd: specifies the year (edy), month (edm) and day (edd) which will filter out all documents created (or modified) later than the given date, e.g. 'edy=2004&edm=2&edd=1' will only list documents created on or before 1st February 2004.
- df: when set to '1' or 'on', the start and end date filters (sdy,sdm,sdd,edy,edm,edd) will be active; otherwise they will be ignored
- pvdc: specifies the max. number of days since today's date -- all documents created before that date (i.e. more than XX days ago) will be filtered out. For example specifying 'pvdc=14' will only list documents created (or modified) within the previous 2 weeks
- vdt: used to filter out documents based on the document type. This value is 65535 (or 0) to include all document types in the search results or it can be a sum of the following values: 1 for PDF documents, 2 for Word documents, 4 for Excel documents, 8 for PowerPoint documents, 32 for media files, 64 for archives, and 16 for HTML (and text) documents. For example: 'vdt=14' will only list MS Office documents (14=8+4+2); 'vdt=17' will list HTML and PDF documents (17=16+1)
- pref: this parameter can be specified in order to change the default database set (the dbXX.sdb files) to another set, that is the .sdb files having this prefix (i.e. instead of dbXX.sdb, the pref_dbXX.sdb files will be used where 'pref_' stands for the prefix). This is very useful when creating custom database sets (e.g. to allow the user to specify the search scope).
- dir: you can use this parameter to set the directory where the database files (.sdb) as well as settings files (header.htm, footer.htm and sessearch.t) are located. This parameter is similar to the 'pref' parameter but sometimes allows more flexibility, e.g. placing search databases in different directories or on different servers can be necessary to provide security for Intranet searches. Please note that the value of this parameter must end with a slash or backslash character; e.g. on a Unix server, this could be "/usr/www/~username/cgi-bin/dir3/", for IIS this could be "c:\inetpub\wwwroot\dir3\" and for an Intranet search this could be "\\server\search\dir3\"
- sort: this parameter can be used to change the order in which the results are listed. By default the results are sorted according to their ranking (value of 0) but you can change it to: 1 -- sort by URL, 2 -- sort by file type, 3 -- sort by file size, 4 -- sort by document date. The second sorting criterium is always document ranking.
- wrap: (CGI and Java servlet versions only) if this parameter is used, all internal links on the results page to other result pages will not point to the CGI script itself but instead to the URL specified by

this parameter (all other search parameters will be passed to this URL as well). This is useful if you want to add server-side code to the results pages -- please see options/output/target type/script language wrapper for more info.

- op: determines if the 'q' parameter specifies 'any words' search (when set to 'or'), 'exact phrase' search (when set to 'ph'), or 'all words' search (any other value)
- qor: search query for 'any words' search
- qph: search query for 'exact phrase' search
- enc: encoding used for the search query (this is the encoding of the page with the search box); the possible values are 'utf-8' (for the UTF-8 encoding) or 'iso-8859-1' (for 8-bit ANSI encodings). If not specified, the encoding is detected automatically (see options/search engine/charset of the search box page for more info).
- fqN: document area-filtered search query, where N is one of the following values or a sum of them (to specify multiple filters): 1 - document contents, 2 - description, 4 - title, 8 - meta keywords, 16 - meta abstract, 32 - author, 64 - filename, 128 - other; e.g. fq3 would limit the search scope to the contents and descriptions of documents

11.2 Extensions - indexing non-HTML documents

Search Engine Studio extensions are used by the program during the scanning process for indexing non-HTML documents. During indexing, the scanning module collects content of each document and puts it into the search database. When the program attempts to scan a document that is not HTML, for example a Microsoft Office or a PDF file, then some kind of conversion is necessary to before the file can be indexed. This conversion is carried out by Search Engine Studio extensions.

In other words, the extension is a utility that can convert a non-HTML document, like PDF, MS Word or a JPEG into an HTML file that next can be indexed. Information and data that will be extracted from a given document depends on its type. For example, in case of a PDF document this will be its main content, meta information like author, title or keywords. On the other hand, in case of an MP3 file this will be a title, author or lyrics if available in the file. In any case, Search Engine Studio tries to extract all relevant data and creates a temporary HTML document that next can be indexed.

Extensions included in the program

Search Engine Studio comes with a number of extensions supporting many popular document formats. This includes:

- Office documents (ppt/xls/doc/pps/chm/odp/ods/odt/p/wpd)
- Adobe Acrobat documents (pdf)
- Media & flash files (avi/jpg/mov/mp3/ogg/png/ram/rm/swf/wma/wmv)
- Archives (zip/gz/tar/taz/tgz/z)

Adding a custom extension

Note: this is an advanced option and in most cases it requires creating a conversion program (exe) in one of programming languages, like C or C++.

When you have a document that you would like to include in your search engine but this format is not supported by the extensions that come with Search Engine Studio, you can create your own extension that will convert the file into an "indexable" HTML format. Please note that this does not

have to be a well-formatted HTML document. It only should contain the <head>, <title>, <body> tags and unformatted content in the body section. It is also good to add a <meta name="content-type" ...> tag with information about the character set that is used for the document.

To add a custom extension you need an executable that converts the source format of a document into an HTML document. This extension utility should follow the rules below:

- It should accept as the parameters: a source document path (pointing to the file that is supposed to be converted) and an output HTML document path (where the extracted content will be saved as an HTML file).
- The program should return code 0 when the conversion was successful and any other value when it failed.

As the result of running the program, the source document's content should be retrieved and saved as an HTML file under the specified output path parameter.

Once you have the executable you should copy it to the *Extensions* folder inside the installation folder, e.g. *C:\Program Files\SearchEngineStudio\Extensions*. In the same folder you should create a text file with a name *xxx.cnv*, where *xxx* is the file format extension that you want to be converted (e. g. for the MP3 extension this would be *mp3.cnv* and for MS Word document *doc.cnv*). The file should contain a command line format that runs the executable. You should use the tags below as markers of the parameters in the command line format:

- (doc-path) - put this marker at the position where the input document path parameter should be passed to the conversion executable
- (output-file-path) - this marker should go where the output HTML file path is expected
- (temp-folder) - this optional marker can be placed at the position where a path to a temporary folder is expected. You can use this marker when your program needs a temporary location e.g. for extracting files etc. to carry out conversion

Sample

Here you can find two samples of a *.cnv* file that match extensions of MP3 documents and ZIP archives (both the extensions come with Search Engine Studio):

mp3.cnv
Mp3ToHtml.exe (doc-path) (output-file-path)

zip.cnv
ZipToHtml.exe (doc-path) (output-file-path) (temp-folder)

Note that the second command line format contains the (*temp-folder*) marker, since the *ZipToHtml.exe* program expects a temporary folder path as the third parameter.

When the scanning module attempts to index e.g. a *c:\mysite\mymusic.mp3* file, then *Mp3ToHtml.exe* is executed with the MP3 file path as the first parameter. Next the extension extracts meta data, like author, title, lyrics etc. from the file, saves all of them in an HTML document at the path specified by the second parameter and returns code 0 (success). Once the conversion is finished the indexing module of Search Engine Studio can index the temporary HTML file and add its content to the search engine database. Of course, all the indexed data will refer to the original MP3 file and not to the temporary HTML file used.

Using multiple conversion executables for one document format

If you want to define more than one executable for the same file format you just need to specify multiple command line formats in the `.cnv` file that defines the extension.

Sample:

```
mp3.cnv
Mp3ToHtml.exe (doc-path) (output-file-path)
Mp3ToHtml2.exe (doc-path) (output-file-path)
```

In the sample above there are two utilities that can be used to convert an MP3 document, `Mp3ToHtml.exe` and `Mp3ToHtml2.exe`. When an MP3 document is supposed to be converted, the first program (`Mp3ToHtml.exe`) is executed. If the conversion is successful the operation is finished and the second utility (`Mp3ToHtml2.exe`) is not used. On the other hand, if the first executable fails then the second one gets its chance.

11.3 Automatic updates of search engine databases

Search Engine Studio can automatically update the search engine databases (by re-indexing documents and uploading the updated database files). In order to take advantage of this option, you will need to use the command-line parameters of the program.

Here's how the program should be executed in the automatic mode:

```
SearchEngineStudio.exe <.ses document path> [option1] [option2] ...
```

The first parameter should be the path of the `.ses` (Search Engine Studio) document for which re-indexing and update should be performed. A number of other parameters (`option1`, `option2` ...) can be specified:

- `/d` - removes all previously indexed documents.
- `/r` - re-index all locations (according to settings specified in the `.ses` file).
- `/g` - generate output files (to local disk or by uploading to an FTP server, according to settings specified in the `.ses` file)
- `/q` - quiet mode; this makes the program run in the background (the UI will be invisible) and exit when all processing is finished

- `/s1`, `/s2`, `/s3` - changes program's execution priority; `/s1` will run Search Engine Studio with idle priority (so it doesn't slow down other programs) and `/s3` will run it with highest priority; by default a normal priority is used (`/s2`)
- `/rxml"<path>"` - takes a list of document paths from an XML file pointed by `path` (e.g. `/rxml"c:\document-paths.xml"`) and removes them. To learn more about the format of this XML see the [Indexing documents from XML file](#) topic.
- `/rxsl"<path>"` - you can use this option to convert using XSLT transformation the XML specified in the `rxml` parameter from a different format.
- `/ixml"<path>"` - re-indexes documents from an XML file pointed by `path` (e.g. `/ixml"c:\document-paths.xml"`). To learn more about the format of this XML see the [Indexing documents from XML file](#) topic.
- `/ixsl"<path>"` - you can use this option to convert using XSLT transformation the XML specified in the `ixml` parameter from a different format.

For example, typing:

```
SearchEngineStudio.exe c:\website.ses /r /q /g <enter>
```

will run the program, open the *website.ses* file, re-index all documents in the background, upload the output files to FTP server (or local / network disk) and exit.

Note: Starting with version 4 of Search Engine Studio you can use the [Search Engine Studio Service](#) to accomplish this task without the need to set up scheduling in external applications. Other advantages of using the SES Service are that the service can run without any user being logged on (important if e.g. the computer reboots) and that you can set it up to update the search engine not only periodically but also automatically, whenever documents get updated, deleted, or added).

11.4 Indexing acceleration by updating only changed documents

Usually, when you start indexing Search Engine Studio goes through all documents of the locations that you have chosen and indexes them. This way each document is indexed and added to the search database. Search Engine Studio also enables you to turn on a mode that allows re-indexing only those documents that have been modified since last indexing process.

You can turn this mode on by setting the ***Remove all documents before indexing*** option in the ***Search Engine*** tab of the ***Search Options*** dialog to *No*. This way the next time you start indexing Search Engine Studio goes again through all documents but time it uses a CRC sum to check if a document's content has changed. As the result, the document's data will be updated only when it was modified.

Generally, it is good to use this option to speed up the indexing process when only a small part of indexed documents changes. In this case updating a few documents will take much shorter then indexing all documents again. On the other hand, if a bigger amount of documents was modified you shouldn't use this indexing mode since clearing the whole database at the beginning will be much faster then updating them one by one during indexing.

11.5 Indexing document date

In the [Indexing Options](#) dialog you can choose to index document date. Usually, this is a date of the document file. On the other hand, in some cases it may appear useful to control the date manually. You can do this by adding a date meta-tag to your HTML document's header. As the result, this date will be used in the search database.

Search Engine Studio accepts date in the following format:

```
date: <meta name="date" content="2005-08-01">, Or <meta name="date" content="2005-08-01T08:00:00+00:00">
```

```
dc.date: <meta name="dc.date" content="2005-08-01">
```

Part



12 Customer and Technical Support

Here you can learn how to get more information about Search Engine Studio and how to find solutions for problems that you may have when using the program.

Frequently Asked Questions

On our web site you can find a [Frequently Asked Question](#) section that contains solutions for most frequent problems that may occur when using Search Engine Studio.

Discussion Forum

You can also take a look at our discussion forum at <http://www.xtreeme.com/forum/> where you can find many help topics about indexing, working on search engine appearance and generating. You can also add your own questions at the forum and get opinions and answers from other users of Search Engine Studio or members of our customer support.

Technical Support Email

If you can't find solution for your problem on our [FAQ](#) page or on the [discussion forum](#) you can always send us an email. You can do this at http://helpdesk.xtreeme.com/helpdesk/new_ticket by filling out the Request Assistance form.

Don't forget to choose the *Search Engine Studio* option in the **Product** drop-down on the form so your question directly goes to the Search Engine Studio team. It is also a good idea to mention in the message a version of your Search Engine Studio and the system you are running.