

USERS GUIDE

Version 6.0

Zoom Search Engine
Custom website search engine



Copyright © 2000-2010 Wrensoft

Zoom Search Engine

Copyright © 2000-2010 Wrensoft

All rights reserved. No parts of this work may be reproduced in any form or by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of the publisher.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of information contained in this document or from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

Printed: March 2010

Table of Contents

Foreword	0
Part I Overview	8
1 Introduction.....	8
2 What's New.....	8
3 Features.....	9
4 System Requirements.....	12
5 How Zoom works.....	13
6 What should I do first?.....	14
7 Which search script platform should I use?.....	14
8 What will the end result look like?.....	15
9 Purchasing Zoom.....	16
10 Credits.....	17
Part II Indexing your website	19
1 Using the Zoom Search Engine Indexer.....	19
Spider mode	19
Offline mode	20
Start spider URL (spider mode only)	21
Start directory (offline mode only)	24
Base URL	24
Output directory	24
Platform	25
Start indexing	25
Stop indexing	25
Indexing status	25
Log	27
2 Configuring Zoom further.....	28
Scan options	29
Skip options	30
Spider options	32
Search page	35
Results layout	37
Indexing options	39
Limits	41
Authentication	42
FTP	45
Languages	45
Weightings	47
Content filtering	48
Categories	49
Sitemaps	49
Index log	51
Advanced	52
Synonyms	54

Recommended links	54
Custom Meta Search Fields	57
3 Step-by-step Wizard.....	57
Select file types step	58
Output folder and limit step	58
4 FTP	59
5 Categories.....	60
6 Custom Meta Search Fields.....	64
7 Search Statistics Report.....	67
8 Icons and thumbnails.....	69
Associating an image with a particular page	71
Associating icon images with a file type	71
Associating thumbnail images with files	73
Customizing the appearance of your icons or thumbnails	76
9 Image indexing	77
10 Incremental indexing	79
Updating an existing index	80
Add start points to existing index	80
Add list of new or updated pages	81
View or delete pages from existing index	81
Command-line parameters for incremental indexing	82
11 XML/RSS output (CGI only).....	82
12 Broken link detection.....	84
13 Configuration (ZCFG) files.....	84
14 Command line parameters.....	85
15 Scheduling and automatic indexing.....	86
16 Using the 64-bit Indexer.....	88
17 Plugins (Indexing PDF, DOC and other document formats).....	88
File formats supported	89
Unsupported binary formats	89
Installing a plugin	90
Using the plugins	90
Configuring a plugin	90
Upgrading a plugin	91
Using custom description (.desc) files	91
ZIP file support	92
Part III Server-side search engine (Using PHP, ASP, or CGI)	95
1 Files required.....	95
2 Installing the ASP.NET native control.....	96
Part IV Client-side search engine (Using JavaScript)	101
1 Files required.....	102
2 Limitations of Javascript.....	102
3 Using the CGI or PHP version without a web server	103

Part V Publishing your search engine on your website or CD-ROM	106
1 What to do after indexing.....	106
2 Uploading to your website.....	107
3 CD-ROM distribution.....	107
4 Additional notes for uploading CGI.....	107
5 Where is my search page?	107
Part VI How do I customize the look of my search page?	110
1 Customizing the search form	111
2 Customizing the search results.....	112
3 Customizing the recommended links.....	113
4 CSS class listing.....	113
5 How do I modify the search form on the search page?.....	116
6 How can I add a search form to my menus, main page, etc.?.....	117
7 Advanced template options.....	117
8 More how-to's.....	120
9 Changing your default editor	120
Part VII Advanced Options	123
1 International / foreign language support.....	123
European languages (French, German, Danish, Swedish, etc.)	123
Russian (Cyrillic)	124
Japanese	124
2 Translating the search page	124
3 Importing and Exporting additional start URLs.....	125
4 Editing the search script.....	126
5 Skipping sections of a page.....	126
6 Manually add words into the index.....	127
7 Alternative page titles and descriptions.....	127
8 Specifying a last-modified date for your web pages.....	128
9 Enable jump to match and highlight within document.....	128
10 Search Statistics PHP Script.....	129
11 Integrating Zoom in your own applications.....	132
12 The indexing process.....	133
13 Word delimitation when indexing	133
14 Word delimitation when searching.....	134
15 Other tips.....	134

Part VIII Troubleshooting and support	136
1 Frequently Asked Questions (FAQs).....	136
2 How to enter in my license key.....	136
3 Discussion forums.....	138
4 Known issues.....	138
5 Technical limitations.....	139
6 Notes for users upgrading from Zoom 4.x	139
7 Notes for users upgrading from Zoom 3.x	139
8 Contacting us.....	140
Part IX Appendix	142
1 What is PHP, ASP, CGI, or JavaScript?.....	142
What is PHP?	142
What is ASP (Classic ASP)?	142
What is ASP.NET?	142
What is CGI?	142
What is Javascript?	143
2 Base URL for CD-ROM distribution (absolute and relative paths).....	143
3 Where is the Program Data directory?.....	145
4 End User License Agreement (EULA).....	145
Index	149

Part

1

Overview

1 Overview

1.1 Introduction

Zoom Search Engine™ is a package for web developers to easily add a search function to their web site, CD-ROM, DVD or Intranet. It has the following advantages over other solutions:

- Allows you to create an internal search engine as part of your own website (without depending on any external web services to produce the results)
- Produces fast search results with a pre-indexed database that is optimized for speed with minimum overhead.
- Customizable look and feel of your search page, which you have complete control over.
- Easy to use and install
- No advertising
- No scripting knowledge required
- Integrates seamlessly with web pages built using FrontPage and Dreamweaver.
- Runs on practically all web servers and hosting plans because it supports multiple scripting platforms: [PHP](#), [ASP](#), [Javascript](#) or [CGI](#).
- Free Edition is available free of charge with a comprehensive feature set for small and personal websites.
- Standard, Professional, and Enterprise Editions are available for commercial development, and larger websites. See <http://www.wrensoft.com/zoom/> for more information.

1.2 What's New

Version 6.0 takes Zoom to the next level of easy to implement, website and CD searching with enterprise-level features. Many new features have been added and existing ones improved, based on feedback from our active user community. We have also worked hard to improve performance and flexibility, while maintaining the user friendly and intuitive design that is users have come to expect from previous versions of Zoom Search Engine.

Some of the new features in V6 include:

- **New User Interface:** We've given the UI an overhaul and in its place is a more intuitive and flexible design. The old configuration window is gone, in its place is the main "Configure" tab where you will find a list of buttons to the left to select different options. Note that you can scroll up and down through this list.
- **Log window with real-time filtering:** You can now filter index log messages on the fly which should be helpful when tracking down error messages, skip reasons, and broken links.
- Zoom will now open the most recently used ZCFG file by default.
- **New Search Ranking Algorithm:** The ranking algorithm and index structure have gone through some major improvements. This should produce much more relevant results than before and also allow you greater control of your search results. This includes preference given to words that are found closer together on the same page, and more. You can configure some of this on the "[Weightings](#)"

" section of the Configuration tab.

- **Faster exact phrase searches:** There can be an improvement from at least 5% to 60% in search time for exact phrase matches depending on the query and content.
- **Stemming:** This allows searches to match words which are similar or derivatives of each other. This addresses an often requested feature to match plural form of words to their singular forms. When this feature is enabled, a search for the word "boat" will match "boats" and "boating".
- **ASP.NET Server Control:** There is now a native ASP.NET platform option. This provides performance similar to the CGI version and is best suited to integrating with ASP.NET web sites. See "[Installing the ASP.NET native control](#)" for more information.
- **Office 2007 plugin:** You can now index and search Office 2007 file formats (e.g. DOCX, PPTX, XLSX)
- **Improved Vista support:** V6.0 is designed with Vista in mind, and features better compatibility with UAC (User Access Control, folder permissions, etc.).
- **Advanced Template options:** While your existing search templates from V5 will work in V6, we now also support advanced options so you can customize the appearance of your search results without having to modify the scripting. See "[Advanced template options](#)" for more information.
- **Custom Meta Fields:** specify arbitrary meta fields to be indexed and made search-able. For example, index and search on a real estate website, by "Number of bedrooms", "Suburb", "Price", "Property type", etc. This is a big feature as it effectively means you can build simple custom databases with a multi criteria search using Zoom without actually having a database. See "[Custom Meta Search Fields](#)" for more information.
- **Improved highlighting in context descriptions:** The highlighting in context descriptions has been improved such that words matching due to stemming, synonyms, or accent insensitivity options enabled can be highlighted.
- **Indexing ZIP files:** You can now index ZIP files and their contents. Each file stored within a ZIP file can be extracted/uncompressed and indexed independently. See "[ZIP file support](#)" for more information.
- **Automatic form login:** This will attempt to login to a website requiring cookie-based authentication with the given login details. See "[Authentication](#)" for more information.
- Wildcards in Skip Pages list: With V6, you can now use the "*" wildcard in the skip pages list to specify the URL's and folder paths you wish to exclude from your index.

A full list of new features can be found at our [What's New](http://www.wrensoft.com/zoom/whatsnew.html) web page at <http://www.wrensoft.com/zoom/whatsnew.html>.

1.3 Features

The following is a brief list of the basic features provided by Zoom,

- Easy to use and user friendly indexer application which runs from your very desktop (or Windows based web server) and allows you to configure all elements of the search engine.
- **Spider indexing mode** allows websites hosted on a web server to be scanned remotely, and index both dynamically generated web sites as well as static content. Supports multi-threading for super fast spider crawling.
- **Offline indexing mode** available for indexing local web pages, with only static content for speed and convenience.

- Plugin support* for indexing a [variety of file formats](#) (such as **DOC, PDF, XLS, PPT, SWF**, and lots more)
- Powerful search syntax allowing for **wildcard matches** (eg. "zo?m", "*zoom*"), **exclusion/negative searches**, **exact phrases***, **Boolean AND/OR**, and more.
- "Google-like" **context search results** shows the matched word and its surrounding context.
- **Highlight words found** in search results.
- Sort search results **by date or relevance**.
- Provide users with **spelling suggestions** when few results are found.
- You can specify [synonyms](#), variations of words, common misspellings, etc. and allow you to map them to an equivalent word in the index.
- [Weighting and boosting options](#) allow you to increase or decrease the importance of text (found in headings, or titles, etc.) and particular pages, giving you greater control over which page gets higher priority in the search results.
- **Flexible indexing options** allow you to select exactly what you want to index, from title of page, page content, to Meta author, and filename.
- **Customizable search page appearance** with HTML and CSS templates.
- [Categories](#) option allows you to group files together, and provide a drop-down or checkbox of categories to restrict your searches to.
- **Authentication support** for indexing secure websites in spider mode
- **Log user searches** performed on your website and analyse the statistics with our [Search Statistics Report](#) tool.
- **Built-in FTP uploading:** Zoom can upload your search files to the web server for you.
- **Built-in scheduling:** Schedule Zoom to automatically index or generate reports for your site.
- **International language support** (various charset support, Unicode/UTF-8, accent/diacritic options, localized search pages, translatable output, and more)
- **Advanced HTML document parsing**, indexing only the content to your website and avoiding various scripting code from being mixed up with the text of the document.
- Prevent indexing pages with identical content with CRC signatures.
- **Supports cookies and cache** used by Windows and Internet Explorer.
- **Supports HTML character entities** and numeric entities. Eg. &#acute; á
- Configuration files allows you to manage multiple search engines with ease.

*Not available in Free Edition

Advanced features include,

- **Incremental indexing:**
Update or add new pages to your index without needing to perform a full re-index. Available for spider mode indexing and PHP, ASP and CGI platforms only. See "[Incremental indexing](#)" for more information.
- **Icons and thumbnails:**
This feature allows you to configure images (such as icons and thumbnails) to appear alongside your search results. Combined with the image indexing option, you can implement image searching capabilities to your website. Note that Zoom does not generate thumbnails or icon images and you must specify the appropriate paths and filenames to these files. See the "[Icons and thumbnails](#)" chapter for more information.

- **Image indexing (JPG, GIF, PNG, TIFF):**
With the use of our new image plugin, you can now index and search image files. Not available in Free Edition. See "[Image indexing](#)" for more information.
- **Recommended links:**
Add pages and links to appear at the top of your search results when a user searches for certain keywords. This allows you to specify the best results for certain searches, or gives you the ability to setup listings for your sponsors if you so desire. See "[Recommended links](#)" for more information.
- **Content filtering:**
Selectively include or exclude pages from being indexed depending on keywords found (or not found) in the content of the file. See "[Content filtering](#)" for more information.
- **Broken link detection:**
Find broken links (links to mistyped URLs or missing pages returning 404 errors) on your website while indexing. See "[Broken link detection](#)" for more information.
- **Sitemap generator (Yahoo™ and Google™ Sitemap compatible):**
Generate a sitemap for your site at the same time as your index! This allows you to create sitemap files to submit to Internet-wide search engines such as Google and Yahoo to help them find all the pages on your site that their spiders may otherwise be incapable of finding. See "[Sitemaps](#)" for more information.
- **Indexing enormous sites:**
We have made some significant optimizing which has led to an overhaul of the index structure and design, consequently allowing us to index and search a much larger amount of data than before. Memory usage has been reduced by more than half in most cases, and this includes all the new features and functionalities which we have added! We have also implemented a number of new techniques which allow us to index much larger sets of data than your physical memory would allow. This includes the ability to flush and merge index data as it is progressively written to disk. The indexing capability of the engine now exceeds a million average web pages, and we have introduced the [Enterprise Edition](#) for users who have such enormous search requirements.
- **Charset by page:**
Zoom will now use the charset specified in the page's meta tag or the HTTP header sent by the server when indexing files. Previously, the Indexer would always expect content to be delivered in the same charset specified in the Zoom configuration window (one charset per session). This means that you can now index various web pages (or websites) which employ different character sets or encoding. The indexed content will then be converted to the encoding selected in the configuration window, and your search page will use the same encoding.
- **Improved categories:**
We have listened to our users feedback and given our category feature an overhaul with some much requested options. These include: allowing files to belong to multiple categories, being able to search for multiple categories, wildcard category match patterns, the option to disable the default/catch-all category, and the use of a ZOOMCATEGORY meta tag to associate a file with a category manually. See "[Categories](#)" for more information.
- **XML/RSS (OpenSearch compatible) output:**
Option to display search results in OpenSearch compatible XML/RSS output. Allows you to post-process the search results via your own scripts, or provide RSS feeds for end-users, as well as the myriad of OpenSearch solutions becoming available online. See "[XML/RSS output](#)" for more information.
- **Adding meta data to remote files:**
You can now override incorrect meta data on remotely hosted sites (which you may not be able to

change, or you do not wish to host .desc files on) by allowing Zoom to look for their corresponding .desc files in a local folder when spidering a remote website. See "[Scan options](#)" for more information.

- **Highlighting words in PDF files:**

You can enable this option to allow searched words to be highlighted within PDF documents when they are opened in Acrobat Reader 7.0 or later. See "[Configuring a plugin](#)" for more information.

- **Limit pages per start point:**

You can now restrict the number of pages to be indexed from a specific start point (and have different limits per start point). See "[Start spider URL](#)" for more information.

- **MP3 audio/music file indexing:**

New plugin for indexing and searching meta data within MP3 files such as Title, Artist, Album, Duration and more.

- AutoCAD file indexing (DWF)
- Indexing filenames of recognized binary files (EXE, WMV, ... etc.)
- And many more!

1.4 System Requirements

For indexing your site (using Zoom Indexer)

The Zoom Indexer needs to run on a Windows computer to scan your website and update the search engine index with information. It can run on any computer with the following specifications:

- Pentium or better PC with Windows 2000, XP, Server 2003, Vista, Server 2008, or Windows 7 (will run on both 32-bit and 64-bit operating systems).
- 64 megabytes of memory (RAM)
- At least 30 megabytes of free disk space
- For very large web sites containing tens of thousands of pages, more RAM (and disk space) is required.
- For spider mode indexing, you will require an established connection to the Internet.

For searching your site

To perform the online search queries made by visitors on your website, you will require one of the following:

- A. **A web server running PHP 4 or higher.** PHP is a server side scripting language supported by most UNIX / Linux based servers. For those unfamiliar with server side scripting, PHP is commonly installed on web servers so you might have access to it and not even know about it. Please consult your web host administrator. For more information about PHP, please refer to <http://www.php.net/>
- B. **A web server running ASP 3.0 or higher.** Microsoft's server side scripting platform with VBScript 5.5 and MDAC 2.8 or above. This requirement is supported by default in most Microsoft IIS installations.
- C. **A web server with permissions to execute CGI.** CGI (Common Gateway Interface) is an

alternative to PHP and ASP for server-side processing. You will require CGI executing permissions on your web server (check with your web host). Our CGI solution provides the best support for very large sites. You will also need to check that your web server runs on one of the following operating systems: Windows, Linux or BSD.

- D. If the above solutions do not cater to your requirements (eg. You want to run searches on a CD without a web server) or your web host simply do not provide these options, then you should use the Javascript version, in which case the only requirement is that your visitors must have Javascript enabled browsers.



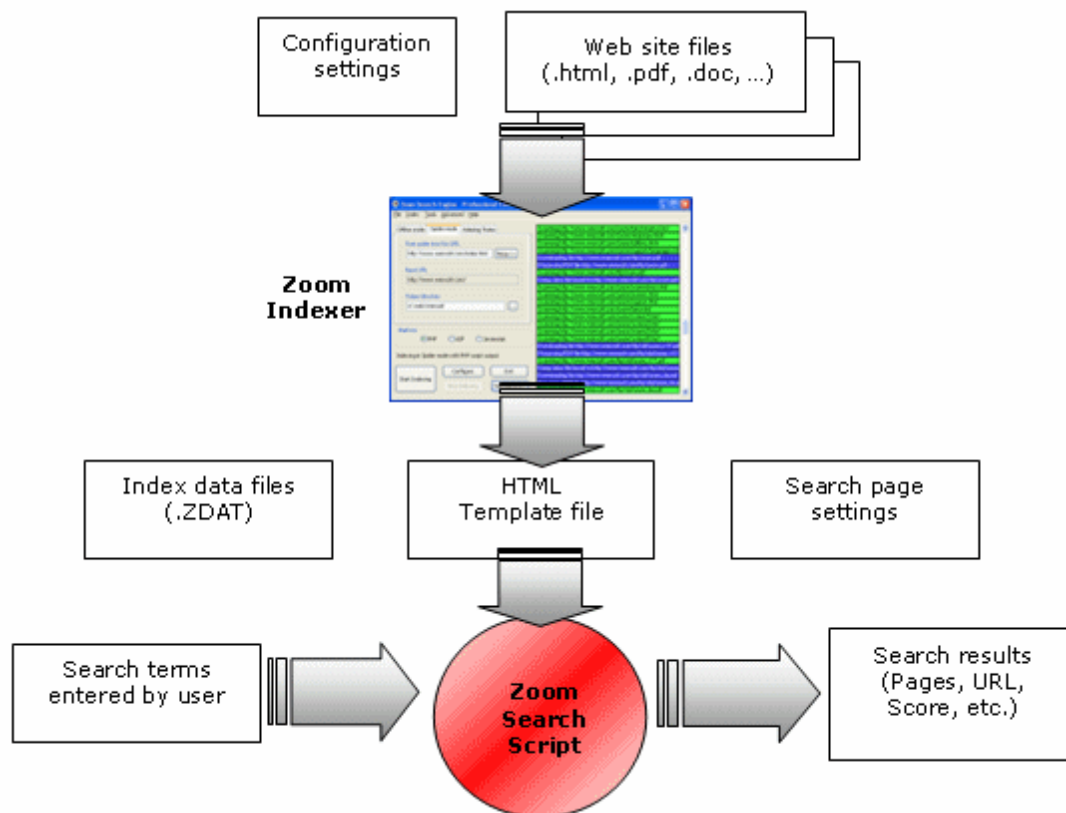
Confused?

If you are not familiar with some of the above terminology, refer to the appendix chapter [What is PHP, ASP, CGI](#) or [Javascript?](#)

1.5 How Zoom works

Below is an illustrated representation of how Zoom works.

As you can see, there are two main components to Zoom: the Zoom Indexer and the Zoom Search Script.



1.6 What should I do first?

The following can serve as a guide for first-time users, or if you are not sure what you should do first.

1. Install the Zoom Search Engine package on your Windows computer.
2. Determine which search script platform you should use. (see [Which search script platform should I use?](#))
3. Index your website with the Zoom Search Engine Indexer application. (see [Indexing your website](#))
4. Upload or copy the files in your output directory to your website or CD-ROM. (see [What to do after indexing](#))
5. You should now have a functioning search page. You can customise the appearance of the search page. (see [How do I customize the look of my search page?](#))
6. Advanced users can also optimise the effectiveness of their search results by examining the log file, adding key words on specific pages, and filtering out unwanted words, pages, or sections of pages. (see Advanced Options)

1.7 Which search script platform should I use?

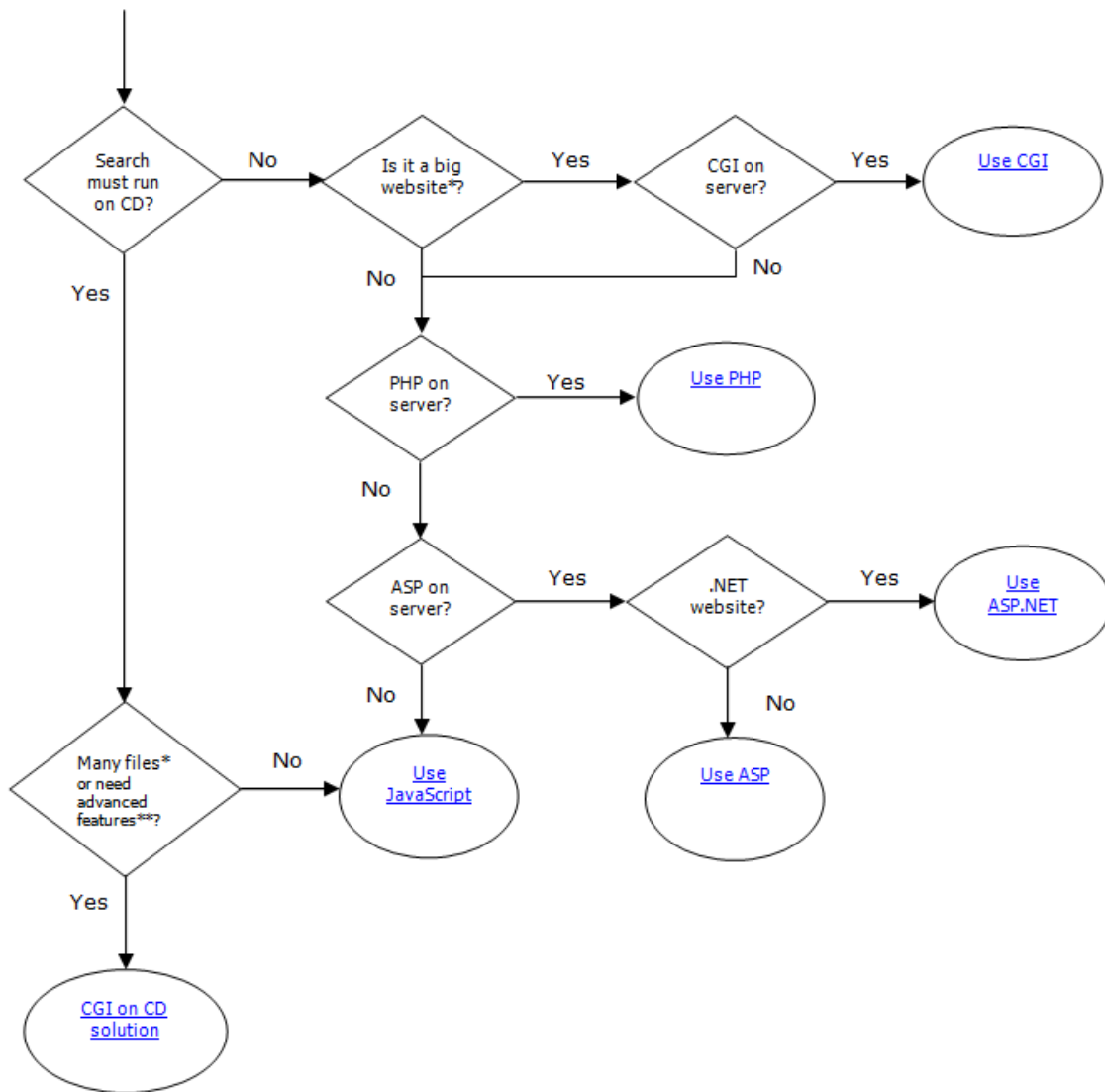
The search script is available in three different versions so that you can select what best suits your method of distribution. The following versions are available:

- **PHP**
- **ASP**
- **ASP.NET**
- **CGI**
- **Javascript**

The server-side versions (PHP, ASP, ASP.NET or CGI) provide the best performance as they operate on your own web server and is the recommended choice for an online website.

The client-side version (Javascript) lets you run the search engine from a CD-ROM or DVD distribution with the benefit of not requiring any external software besides the use of Javascript enabled web browsers. However, it is also more restricted in technical capability and efficiency than the server-side versions.

The following flow-chart may help you determine which script platform to use.



*A site with approximately 65,000+ pages or 80,000+ unique words would be considered a big website.

**Advanced features include "context description" and "exact phrases", both of which are not available to the JavaScript version due to limitations in the script platform.

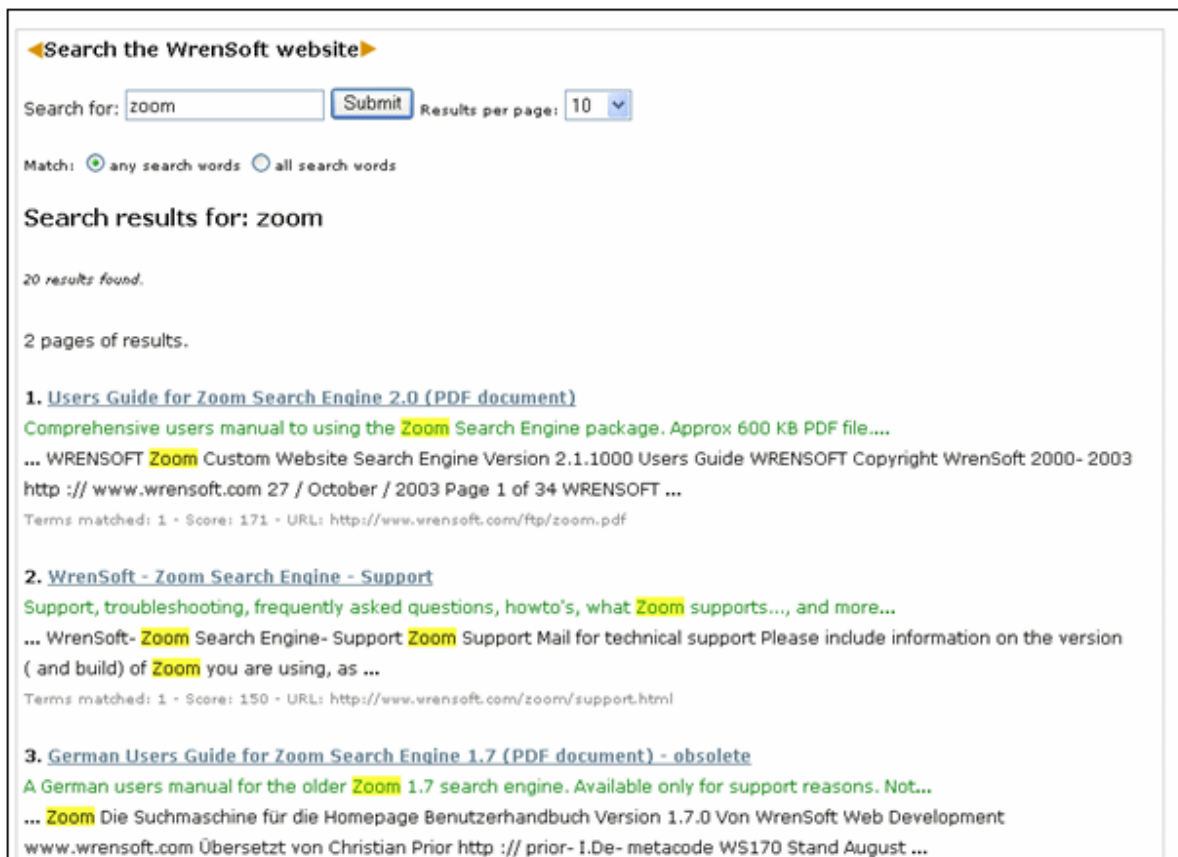


Confused?

If you are not familiar with some of the above terminology, refer to the appendix chapter [What is PHP](#), [ASP](#), [CGI](#) or [Javascript](#)?

1.8 What will the end result look like?

To give you an idea of what the resulting search page could look like, the following is a screenshot of a Zoom search page running on our very own website (you can see this online and run searches on it yourself at <http://www.wrensoft.com/search.php>).



1.9 Purchasing Zoom

If you wish to use Zoom for a larger website, or you are using it in a commercial project, require professional support, custom features, or simply like the software and wish to support its development, please consider purchasing one of the licenses available.

The free edition of Zoom is restricted to a typical size of a free, personal website (50 pages, 15,000 unique words, 100,000 bytes, and 150 characters respectively).

For larger websites, or commercial projects, we recommend the **Standard Edition** license which is capable of indexing larger website (up to 100 pages) and provide plugin support for Acrobat PDF, Word DOC, and a growing list of other file types.

For commercial developers, or owners of larger sites, we recommend the **Professional Edition** license which allows you to change these limits manually, up to a maximum of 200,000 pages and half a million (500,000) unique words.

For extremely large sites (of over a million pages) or a cross-site search engine spanning many different websites, we would recommend the **Enterprise Edition** which has no limit on the number of pages or unique words you can attempt to index (only limited by the amount of memory in your indexing computer).

For more information on the differences between these editions, please visit our webpage at: <http://www.wrensoft.com/zoom/editions.html>

1.10 Credits

Zoom Search Engine™ is developed by Wrensoft® Web Development, a subsidiary of PassMark™ Software. All scripts, binaries, and documentation included are Copyright © 2000-2010 Wrensoft.

The Zoom Search Engine software, or a component thereof, uses the following libraries:

Regular expression support is provided by the PCRE library package, which is open source software, written by Philip Hazel, and copyright by the University of Cambridge, England.

YACGI, copyright 1996-1997 by Andrew Girow (Andriy Zhyrov). Permission is granted to use YACGI in any application, commercial or non-commercial, at no cost.

UCData 2.9 by Mark Leisher, copyright 2005 Computing Research Labs, New Mexico State University. Permission is granted to use UCData without restriction subject to conditions specified.

Porter Stemming algorithm by Martin Porter. PHP implementation by Richard Heyes, published for use "free of charge for any purpose" on Martin Porter's website (<http://tartarus.org/~martin/PorterStemmer/>), Feb 2005. ASP implementation based on code by Christos Attikos, published at abovementioned URL, Jan 2005.

Snowball Stemmer, copyright 2001-2002 by Dr Martin Porter, and Richard Boulton. Permission is granted to use Snowball under the BSD license.

Part

2

Indexing your website

2 Indexing your website

2.1 Using the Zoom Search Engine Indexer

The Zoom Indexer is a Windows application that scans your entire website and indexes the content and information it finds on each page. It will then create all the files you need to upload to your website, for you to have a running search engine.

The Zoom Indexer is a part of the [Zoom Search Engine](#) package that provides powerful custom search functionality for your website or CD-ROM. For more information on the overall package, please see the "[Introduction](#)" and "[How Zoom works](#)" sections.

The Indexer has two main modes of operation:

- [Spider mode](#)
- [Offline mode](#)

It also has a choice of four different platform options (PHP, ASP, Javascript, or CGI).

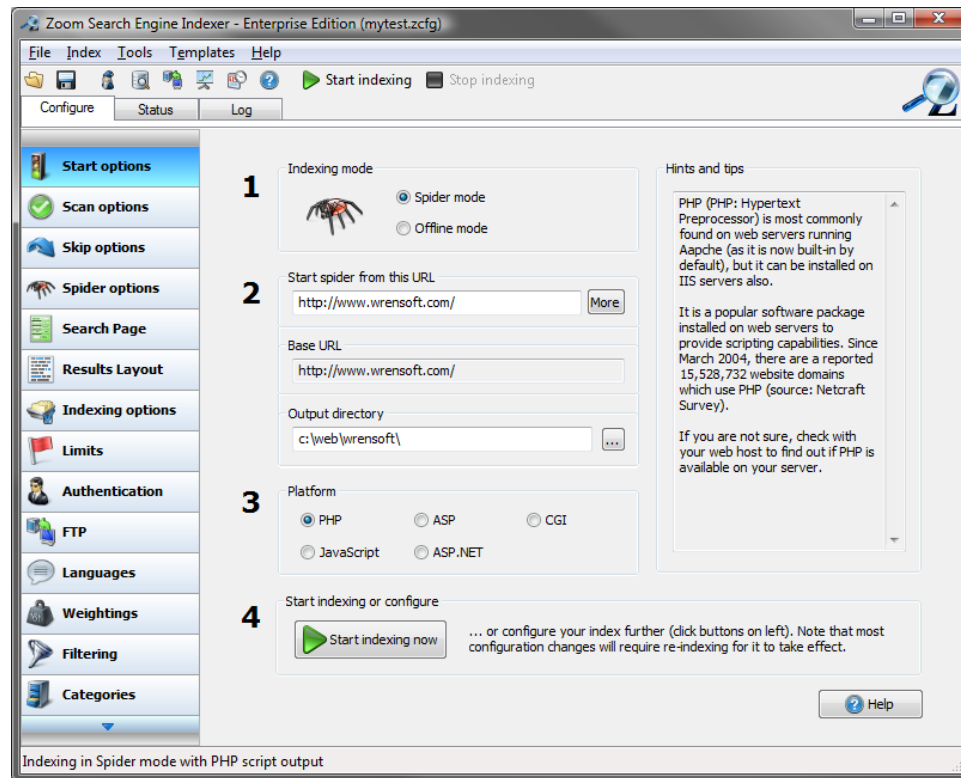
For more information on selecting a suitable platform, see [Which search script platform should I use?](#)



Getting started? First time users should follow the [Step-by-step wizard](#) to guide them through the main options. Click on the magic wand in the toolbar to begin. We would also recommend reading the rest of the Overview chapter, in particular the "[What should I do first?](#)" section.

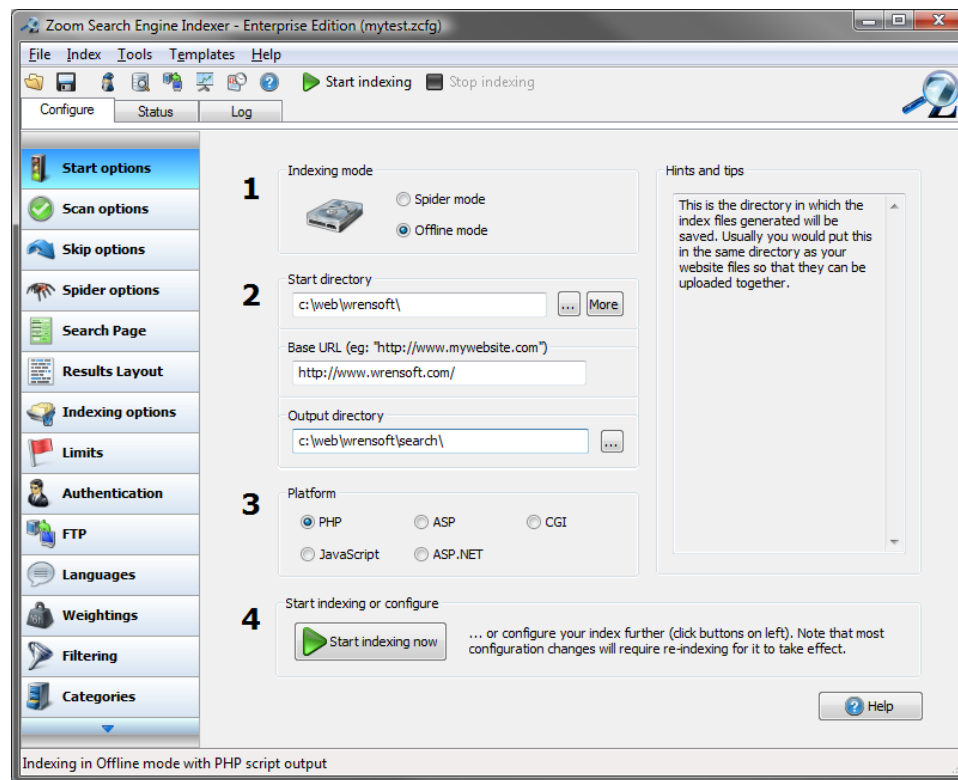
2.1.1 Spider mode

Spider mode indexes a remote copy of your website already uploaded and hosted on a web server. It does this via the use of a 'spider', which starts from a given start page, and follow the links it finds on each page. This allows the indexer to thoroughly index a website containing both static content (.htm and .html files which do not change) and dynamically generated content (such as websites with a PHP or ASP driven backend, message boards, etc.). It requires an established Internet connection.



2.1.2 Offline mode


Offline mode indexes a local copy of your website, stored on your hard disk. This is effective for static web pages and allows the user to index a website without uploading it to a web server, maximizing indexing speed and convenience. It can also be used for web pages that will be published on a disk or CD-ROM, where a web server will not be available and does not require an Internet connection.

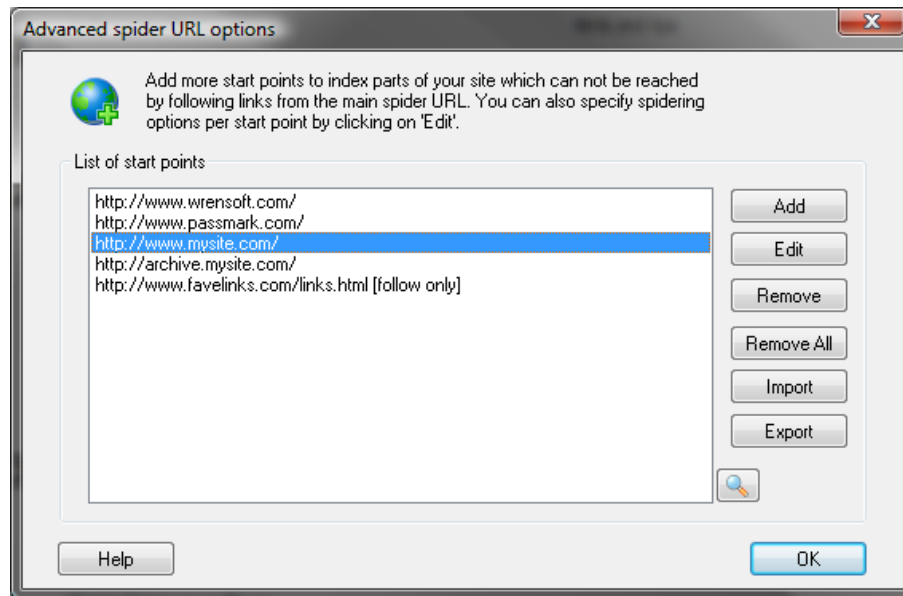


2.1.3 Start spider URL (spider mode only)

In [spider mode](#), you are required to specify the URL from which the indexer will start the spider scanning from. Typically, you would point this to the entrance page of your website, (such as index.html) so that it will be able to find links to other pages on your website by following the links it finds on each page (as a visitor would).

Also note that the spider indexing mode automatically skip links to external web sites, i.e. those that are outside of the base URL defined (see below). This is to prevent indexing pages outside of the specified website.

Advanced spider URL options: Clicking on the  button will bring up a window which allows you to add more spider URLs or specify advanced spider crawling options. This is particularly helpful when indexing across multiple websites or domains.

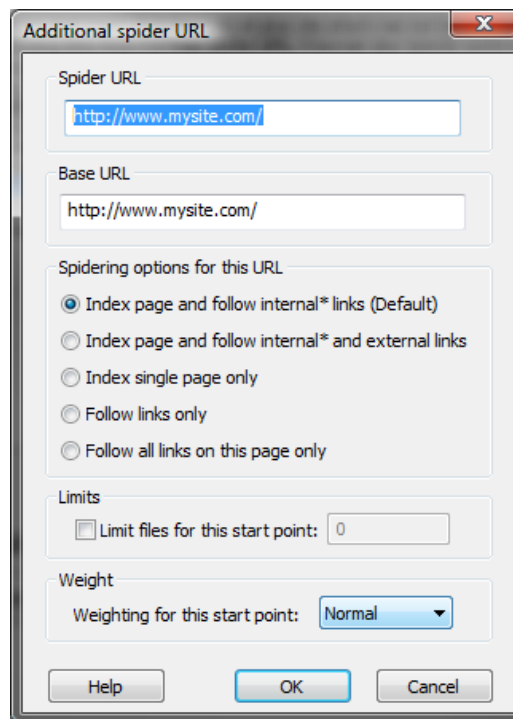



Spidering options

With each spider URL in this list you can specify the following options:

- Index page and follow internal links (default) – will index the contents of the page and follow any internal links found (URLs beginning with the base URL).
- Index page and follow internal and external links – will index the contents of the page and follow internal and external links (but only up to one level of external links – eg. it will scan each external page linked from an internal page, but will not index external pages linked from external pages).
- Index single page only – will only index the contents of the specified page, and not follow any links found.
- Follow links only – will only follow the links found on the specified page but will not index the content of the page itself. The spider will then index and follow the links found on the pages that are linked to from this page.
- Follow all links on this page only – will follow the links found on the specified page and index the linked pages, but not follow any further links. This indexes only one level of links, that is, only pages which are linked to this start point will be indexed.

You can also override the automatic base URL determined from this window, if necessary.



 **Tip:** You can specify multiple base URLs for each individual start point. This allows a start point to span across multiple domains or sub-domains. For more information, see "[Base URL](#)".

Limits files for this start point

You can limit the number of files to index from this particular start point by checking this option. You can specify a global limit for all start points on the "[Limits](#)" tab of the Configuration window. Note that when both the global and individual limit is set, both settings will apply, so which ever limit is first reached (ie: the lower limit of the two), will cause the indexer to stop indexing the current start point.

Weighting for this start point

This adjusts the score weighting for the pages indexed under this start point. This can be used to make pages found from a particular start point or domain to be ranked higher or considered more important than pages from other start points. See "[Weightings](#)" for more information.


Import and export start points

You can also Import and Export additional URLs from a text file using the Import and Export button. See "[Importing and Exporting additional start URLs](#)" for more information.

The number of start points you can have in this list are only limited by the system resources available. However, the total number of pages indexed would still be limited by the indexing limits (max. pages, max. unique words, etc.) specified on the "[Limits](#)" tab.

2.1.4 Start directory (offline mode only)

The start directory specifies the local directory for the offline scanning to begin in. All sub-directories under the start directory will also be scanned. In other words, you should point this at the folder in which you have created the files for your website.

Advanced start folder options: Clicking on the  button will bring up a window which allows you to add other start folders to include in this index. This allows you to index multiple folders (not already under the main start directory) and include them in the search index.

2.1.5 Base URL

This is the URL where your website will be published and uploaded to. For example, if your website will be published at `http://www.myisp.com/~me/index.html`, then `http://www.myisp.com/~me` will be the base URL of your website. This is used to determine the base location of each file on your website, so do not specify the filename of the main page (i.e. `index.html`, `home.htm`, etc.).

In **spider mode**, this is automatically determined for you based on the URL of the start page specified, but you can override it if necessary (click on the "More" button and select "Edit"). When specifying your own base URL in spider mode, you can also list more than one base URL by separating each with a semicolon (;). For example, to allow a spider to follow links from the start point "`http://www.mysite.com/index.html`", to pages hosted under the "`pages.mysite.com`" domain, you would have to specify a base URL of "`http://www.mysite.com/http://pages.mysite.com/`".

For CD-ROM/DVD distribution

For offline/local distribution, the base URL should be the relative path from where the search script will be. It does not have to be a HTTP address. For example, if the files you have indexed will be placed in the same directory as the `search.html` page on the CD, use `./` (which refers to the current directory) as your base URL. This is generally recommended over the use of absolute paths such as `/` or `C:\myfiles\`, which will raise compatibility issues between different operating systems (such as Macs). See [Appendix B](#) if you are unfamiliar with relative paths for more information.

For offline or Intranet use

Note that the base URL can not be a file path (such as `C:\Myfiles\`) or UNC address (eg. `\\MyServer\Files\`). Instead, you should use a valid file URL in the form of [file://MyServer/Files/](#).

You could also consider using relative paths for your base URL, especially if you need to be able to move the search files and host them on different servers. See [Appendix B](#) if you are unfamiliar with relative paths for more information.

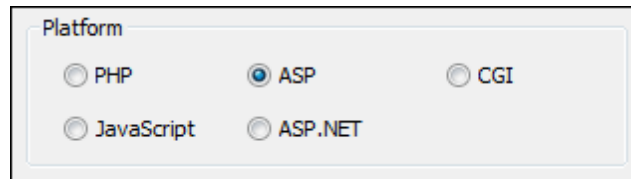
2.1.6 Output directory

This is the directory in which the index files generated will be saved. Usually you would put this in the same directory as your website files so that they can be uploaded together. Note that all the files created must be uploaded to the same directory.

See "[Publishing your search engine on your website or CD-ROM](#)" for information on what to do with the files created in this folder, after successfully indexing your website.

2.1.7 Platform

This setting selects the script platform, which you will host or distribute your search engine with. Select from either PHP, ASP, ASP.NET or CGI for server-side searches, or Javascript format for client-side searches.



The screenshot shows a 'Platform' selection window. It contains five radio buttons arranged in two rows. The first row has 'PHP', 'ASP' (which is selected), and 'CGI'. The second row has 'JavaScript' and 'ASP.NET'.

If you do not know which format best suits your method of distribution, see "[When to use a server-side search engine \(PHP, ASP, ASP.NET or CGI\)](#)" and "[When to use a client-side search engine \(Javascript\)](#)".



Note: Clicking on the CGI option will popup a window allowing you to select the target OS on which the CGI will run (this would be the OS of your web server).

See also:

[What is PHP?](#)

[What is ASP?](#)

[What is ASP.NET?](#)

[What is Javascript?](#)

[What is CGI?](#)

2.1.8 Start indexing

This starts the indexer using the current settings in the mode selected. When the process finishes, "Indexing completed" will appear in the log window on the right and the indexer will generate the data files in the output directory.

2.1.9 Stop indexing

This requests the indexing process to stop. Note that the indexer will attempt to finish indexing the file it was up to, and will generate functional index data files before stopping completely.

2.1.10 Indexing status

The Status tab provides an overview of the indexing process such as the number of words found, the number of links found, and the number of files found for each different file extension.

For more information on the current indexing progress, click on the "Log" tab.



Hint: Use the information available here to determine the size and limits of the site being indexed, and whether further configuration with the Scan options, Skip options, or Limits is necessary.

Errors and warnings

The errors and warnings list in the top right corner of the Status tab, is a summary of any errors, warnings or broken links that the Indexer has come across for this indexing session. Double-clicking on the "View" column will bring you to the "Log" tab, with all message types disabled except for the corresponding error you selected here. This allows you to track down specific error messages or broken link messages even when you have a very large Log.

The Last logged error message shows the most recent error or warning message that was written to the Log window.

Thread status

This bottom section of the Status tab gives a clear indication of what each thread is doing during indexing. You can only have one Indexer thread at any one time, but the number of download threads (labelled as "DL #1", etc.) can be adjusted for Spider Mode indexing on the "Spider options" panel of the

Configuration tab.

System information and memory status

This reports the resources used and/or available on your computer during indexing. Note that it will not update except only while indexing is under way. You should use this to determine the actual memory usage of your "[Limits](#)" settings, and note when your computer may be under equipped for an indexing task you have set.

2.1.11 Log

The "Log" tab provides detailed information on the indexing process. You should evaluate these messages to determine what files have been indexed, what files have not been indexed (and why), and any other possible issues to address during indexing.

You can now toggle the message types to be displayed on-the-fly. This will help you track down errors easier.

Note that some these settings may not be available if you have configured the Index Log mode (on the "[Index Log](#)" panel of the Configure tab) to "Basic".

Click the "Show all" button to quickly check all options or "Reset to default" to return the log settings to the default options.

<i>Indexing</i>	Messages relating to which files have been indexed and scanned successfully.
<i>Skipped</i>	Messages about files which have been skipped or not scanned.
<i>Filtered</i>	Messages about files which have been filtered out by the Content Filtering feature.
<i>Spidering</i>	Messages about the spidering process when indexing in spider mode (eg. when links are found and queued for downloading).
<i>Initialization</i>	Messages regarding initializing memory or index data for Zoom.
<i>Downloading</i>	Messages regarding files being downloaded (in spider mode only).
<i>Uploading</i>	Messages regarding files being uploaded with FTP .
<i>File I/O</i>	Messages regarding writing and reading of files such as the index data files.
<i>Information</i>	Messages regarding miscellaneous information.
<i>Thread info</i>	Messages regarding multi-threaded activity (in spider mode only)
<i>Error</i>	Messages regarding critical errors whilst indexing.
<i>Warning</i>	Messages regarding minor problems or issues that may need attention with indexing.
<i>Plugin</i>	Messages regarding the use of plugins or processing of external binary file formats such as PDF, DOC, PPT, etc.
<i>Summary</i>	End of indexing summary.
<i>Broken links</i>	If a link is followed and found to be missing or invalid, this option will report the original page where the broken link was found. See " Broken link detection " for more information.

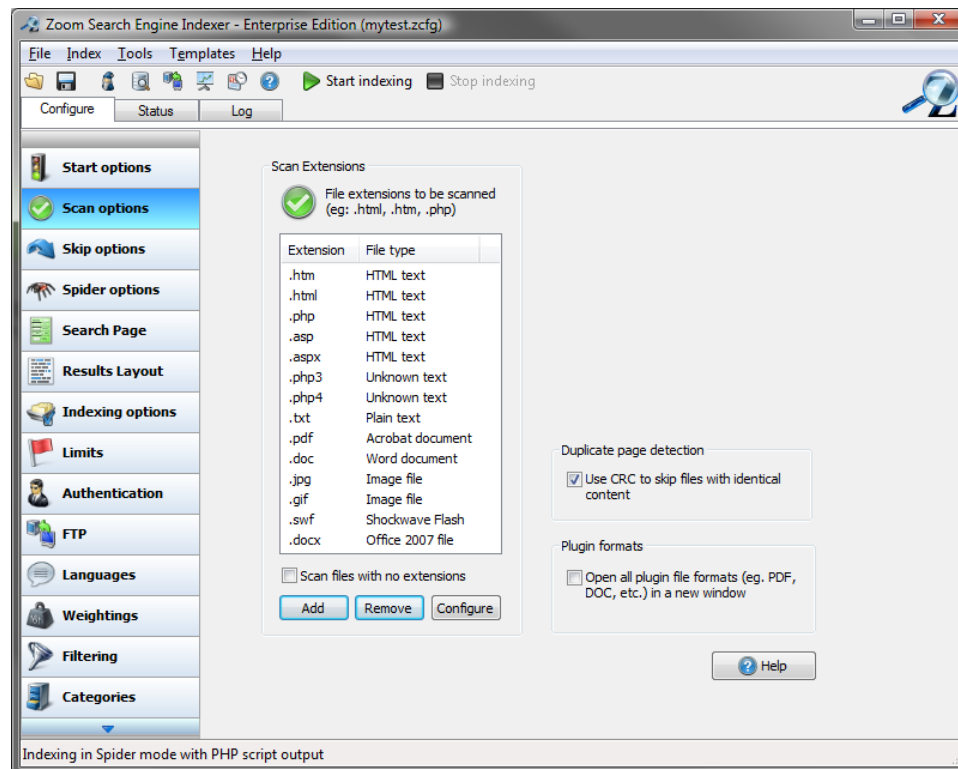
2.2 Configuring Zoom further

You can change everything from the indexing behaviour of the search engine, to the appearance and features of the search results page from the Configuration window. Click on the Configure tab and select the corresponding panel of options.

There are several groups of options presented which you can modify:

- [Start options](#)
- [Scan options](#)
- [Skip options](#)
- [Spider options](#)
- [Search page](#)
- [Results layout](#)
- [Indexing options](#)
- [Limits](#)
- [Authentication](#)
- [FTP](#)
- [Languages](#)
- [Weightings](#)
- [Filtering](#)
- [Categories](#)
- [Sitemaps](#)
- [Synonyms](#)
- [Recommended links](#)
- [Index log](#)
- [Advanced](#)

2.2.1 Scan options



Scan Extensions

This is a list of the filename extensions that will be scanned by the indexer. If a file has an extension that is not in this list then it will be skipped. Note that each file extension must begin with a "." (dot) character or it will be ignored. Up to 50 extensions can be defined in this list.

When you add a new extension, Zoom will automatically determine the File type that it should be indexed as. However, you can change this as required.

Checking the "Scan files with no extensions" options allows you to index files without any extension.

You can double-click on each file extension (or click on the "Configure" button) to edit settings for that particular file extension. This includes icon and thumbnail options for most file formats (see "[Icons and Thumbnails](#)").

If your edition of Zoom has plugin support (available in all editions except Free Edition), you could also install plugins and specify file extensions such as ".pdf" and ".doc". See "[Plugins](#)" for more information. You will also be able to configure the plugin settings from here by double-clicking on the related file extension (or clicking on the "Configure" button with the file extension selected).



Note: While you can specify the "File type" for each extension, the way a file is indexed can be dependent on the server's Content-Type response in Spider Mode. This behaviour is necessary and desirable because a server may have a PHP download script which returns a PDF file (for example), and in such a case, you would want Zoom to obey the web server and treat the file returned as a PDF document instead of a PHP page.

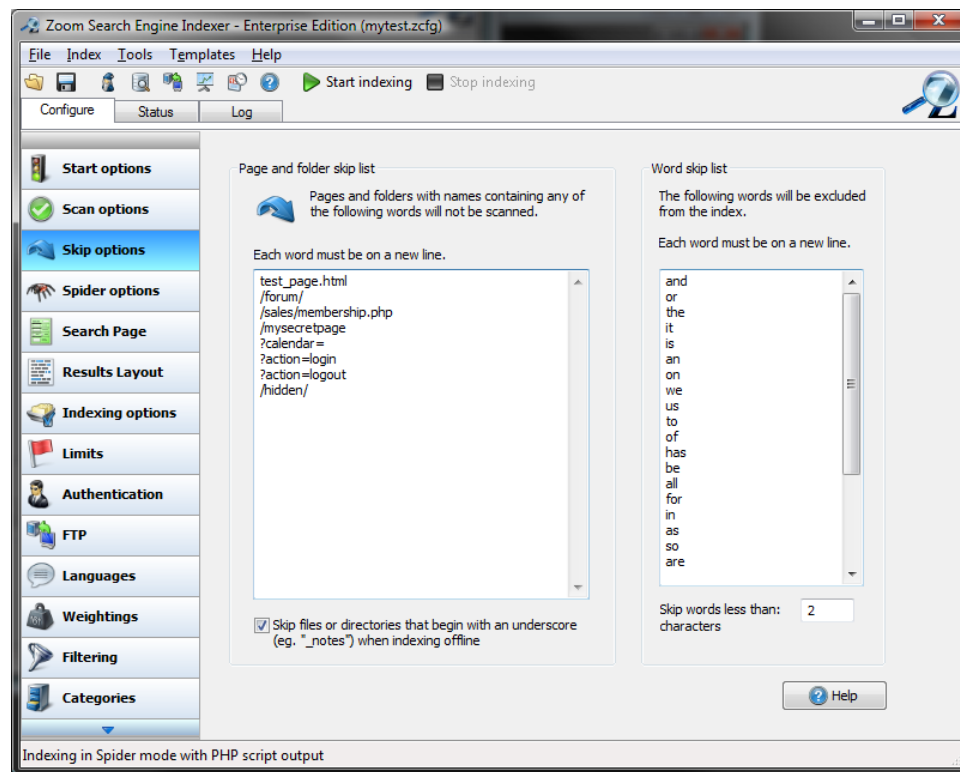
Duplicate page detection

Checking this option enables the use of CRC signatures to ensure that only pages with unique content are indexed. This is particularly useful for spider indexing websites with links to pages without a filename, for instance, to a directory (eg: <http://mywebsite.com/>). These links may otherwise be indexed twice if there is another link somewhere else on the website which points to the same place, but with the actual filename specified (such as <http://mywebsite.com/index.html>, <http://mywebsite.com/home.htm>, etc.). It is best to avoid this on your website and use a consistent linking method. However, you can also prevent this by turning on this option.

Open all plugin file formats in a new window

When this option is enabled, all plugin supported file formats (e.g. PDF, DOC, PPT files, etc.) will open in a new window when you click on them in the search results. Note that you will need to have plugins installed and the required file extension added to the Extensions list for such file types to be indexed. See "[Plugins](#)" for more information.

2.2.2 Skip options



Page and folder skip list

This is a list of pages and folders that will not be scanned during the indexing process. Note that filenames and paths are case sensitive. Typically you would want to filter pages that the user should never be able to get to directly via the search function. Note that if the path to a page partially or fully matches any entry in this list it will be filtered. For example, an entry of "\\private\\" will filter

"\private\file1.htm", "\private\file2.htm" and "photos\private\athome.htm".



Tip: You can also skip pages based on whether certain keywords are found or not found within the page content. See "[Content filtering](#)" for more information.

As of V6, you can now also use the asterisk wildcard (*) to match a sequence of characters. For example, an entry of "\private*\index.html" will skip all files named "index.html" in a sub-folder within "\private\".

Skip files or directories that begin with an underscore

When this is enabled, files or directories beginning with an underscore (such as those kept by Dreamweaver and FrontPage) will be ignored and skipped during Offline Mode indexing.

Word skip list

This is a list of words that will be filtered during the indexing process. Typically you would want to filter some small words that appear on every page such as, "and", "or", "the", etc. Depending on your site you may also want to skip other words. For example, if your site had a thousand pages about different aircrafts, you might want to remove the word "aircraft" from the index, as it would match every page on your site.



Tip: If a skip word begins with a '*' character then it will match and filter any word that contains this keyword. For example, a skip word entry of "*fun" would filter out the following words: "fun", "funny" and "nofun". Note that the '*' character must be at the beginning of the word for this to take effect.

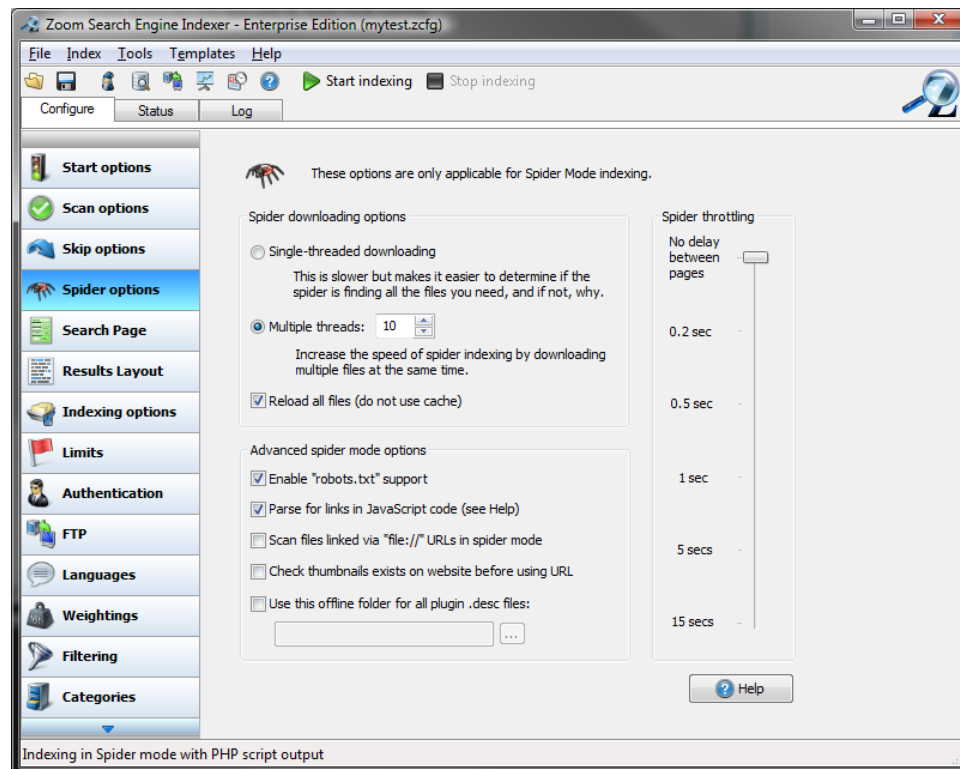
Skip words less than x characters

You can now specify the minimum number of characters that a word must be before it is indexed. It is defaulted to two characters so all single-character words are skipped.

Skipping sections of a page from being indexed

To exclude sections of a page (such as headers, footers, and navigation menus), please see "[Skipping sections of a page](#)".

2.2.3 Spider options



Spider downloading options

These are options which control the way in which Zoom will download files when it is in spider mode. Note that it does not apply for offline mode indexing.

Single-threaded downloading

This option uses only one dedicated thread for downloading files. This is typically the slower option but provides reasonable speed when indexing a site with a fast connection. It is also recommended when you are trying to follow the spider's crawling path, to determine if it is scanning the pages you are expecting.

Multiple threads

This option allows you to specify more than one dedicated thread for downloading files in spider mode. This is recommended to increase the speed of indexing. It allows Zoom to download multiple files in the background whilst indexing at the same time.

Reload all files (do not use cache)

Check this option to ensure that all files are downloaded from the site and that the cached copies of pages are not used.

Spider throttling

This option allows you to add or increase a delay between requests made to a web server when indexing in Spider Mode. This can be useful when you are crawling a web server which is under heavy load and you wish to minimize any additional load that can be placed on a server during the spidering process. Note that in most cases, the spider should not put much strain on a web server as it is limited by the bandwidth and processing capability of one single desktop computer. This option should only be necessary when indexing a server which is overloaded with an unreasonable number of websites or tasks, running on underpowered hardware. Using this option will significantly slow down your spider indexing process. For most other situations, we recommend setting this to "No delay between pages".

Enable "robots.txt" support

When this option is enabled, Zoom will look for "robots.txt" files when indexing a website in Spider Mode. The "robots.txt" file can specify instructions for spiders and other user-agents, on what pages should be excluded from indexing (similar in effect to the "[Skip pages list](#)") and also whether a crawl delay should be required (similar in effect to the "[Spider throttling](#)" option).

Zoom will download a "robots.txt" file (if available) for each start point, so this method allows you to have per start point skip pages and throttling settings. It is also a good idea when indexing third party websites, so that you can make sure your spider is obeying the webmaster's rules.

Note that Zoom will only locate a "robots.txt" file for each start point, at the root level of the domain being indexed. It will not parse "robots.txt" files which are located in sub-folders. This means you should have all your sub-folder robots settings located within the one "robots.txt" file, specifying your rules relative to the base URL.

For example, the following "robots.txt" file will block Zoom from indexing any files in a folder named "secret" and any files named "private.html". It will also force a delay of 5 seconds between requests to this start point.

```
# this is my robots.txt for http://www.mysite.com/ (this comment is ignored)
User-agent: ZoomSpider
Disallow: /secret/
Disallow: private.html
Crawl-delay: 5
```

For more information on the "robots.txt" file format, please refer to online resources such as <http://www.robotstxt.org/>. When this option is enabled, Zoom will also support the "robots" meta tag for "noindex" and "nofollow". This allows you to specify certain pages to be excluded from indexing (or crawled for links) by simply adding a tag such as the following within the page head:

```
<meta name="robots" content="noindex,nofollow">
```

Note that specifying "index" or "follow" values in the robots meta tag will have no effect as this is the default behaviour for all pages scanned.

Parse for links in JavaScript code

Some links on your web page may be embedded within JavaScript code (e.g. pop-up navigation menus). These links are generally considered to be search engine unfriendly, because a web spider is unable to execute the script and crawl the resulting links.

While Zoom will not execute JavaScript, this option asks Zoom to attempt to look for URLs within the JavaScript code and crawl/follow any valid links that it finds. This will typically find some of the links in your JavaScript but not necessarily all of them. It also increases the time it takes to index a web page. However, it can be a decent solution if you have many links within JavaScript code and do not have the time to fix your web page to be more search engine friendly.

You can find more information on indexing JavaScript links and better long-term solutions explained in our online FAQ here: http://www.wrensoft.com/zoom/support/faq_problems.html#javascriptmenus

Scan files linked via “file://” URLs in spider mode

This allows the spider mode to follow file:// style hypertext links. This can be useful for indexing an Intranet where you may have accessible files on the web server as well as the shared drives on the network.

Check thumbnails exist on website before using URL

This option applies for search result thumbnails configured as described in "[Icons and thumbnails](#)".

With this option enabled, Zoom will check each thumbnail URL and see if the image file exists on the web server (at the time of indexing) before deciding to use the determined URL for the thumbnail. This prevents "broken image" thumbnails.

Use offline folder for .desc files

This option allows you to specify custom description (.desc) files for your plugin supported files, by hosting them locally in an offline folder.

It allows you to override incorrect meta data on remotely hosted files (possibly on sites that you can not change, or where you do not wish to host the .desc files). For more information on .desc files, see "[Using custom descriptions \(.desc\) files](#)".

With this setup, you can now index external sites using Spider Mode, and the Indexer will look for the .desc files for any plugin supported file formats (such as .pdf, .doc, etc.) in the local directory. This allows you to specify custom .desc files without having to host them up on the remote web site.

The offline .desc files need to include the full domain name and URL path in its filename. This is usually everything after the "http://" or "https://" prefix. It must also end in ".desc" (see examples below).

However, since a number of characters possible in a URL are not valid as filenames, you must encode these characters in their hexadecimal form and precede them with a "%" sign. This is similar to the HTTP encoding required for URLs. The following is a list of the characters in URL which must be encoded.

Character	Encoded
\	%5C
/	%2F
:	%3A
*	%2A
?	%3F
"	%22
<	%3C
>	%3E
	%7C

For each of the above characters in a URL, substitute them with the Encoded form of the character when naming a .desc file for that URL.

Here are some examples of URLs and their corresponding .desc filenames,

Example 1,

URL:	<code>http://www.mysite.com/files/mydocument.pdf</code>
.desc filename:	<code>www.mysite.com%2Ffiles%2Fmydocument.pdf.desc</code>

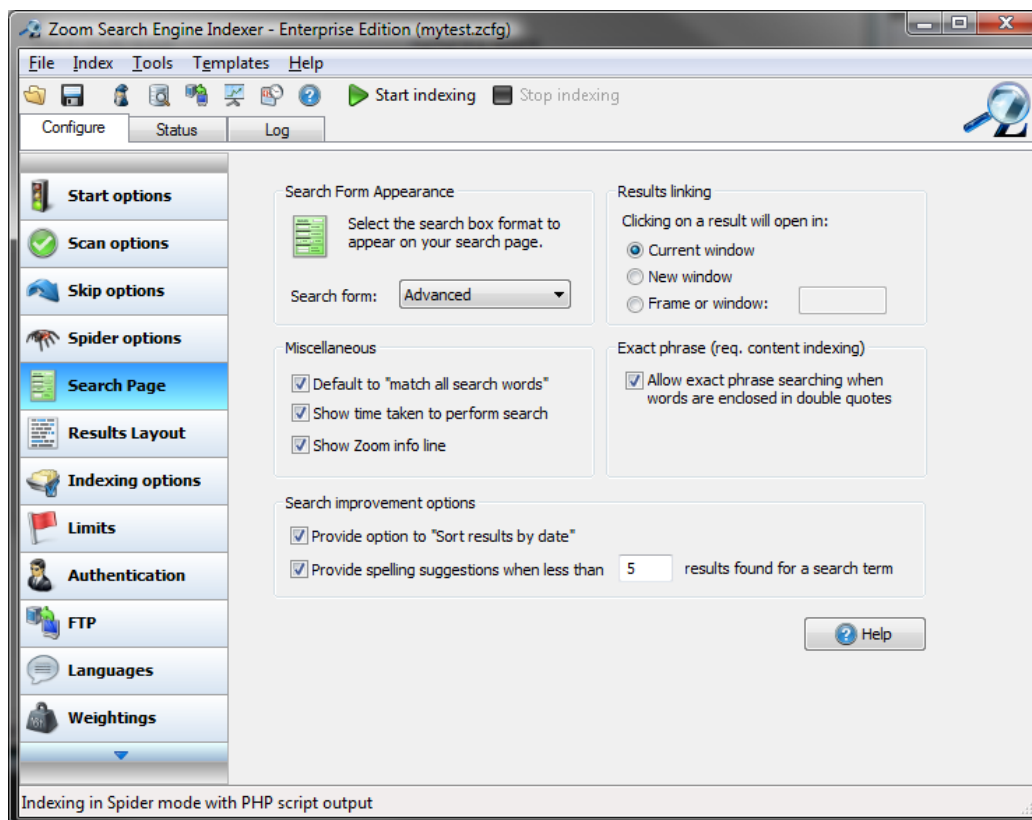
Example 2,

URL:	<code>http://www.mysite.com/download.php?fileid=123</code>
.desc filename:	<code>www.mysite.com%2Fdownload.php%3Ffileid=123.desc</code>

Of course the preferred solution would be to create documents with correct meta data in the first place. But when this hasn't been done, local .desc files can provide more accurate searches and better looking results.

2.2.4 Search page

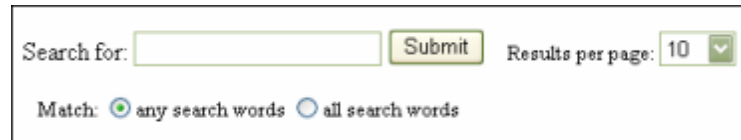
These are options specific to changing the appearance, or behaviour of the search page from which searches are entered and results are shown.



Search Form Appearance

The search form is the area of the search page where you can enter a search query, select parameters

and submit the search. Here you can choose from "Advanced" (with full options) to "Basic" (simple search box), to "Do not generate". With the latter option, the script will not generate the search box. This allows you to create your own search form in your search template HTML file (or provide the search interface elsewhere, such as in a separate frame). Note that the generated search form is useful because it can remember the entries of the previous search query.



If you wish to specify your own search form, please refer to "[How can I add a search form to my menus, main page, etc.?](#)"

Results linking

This option defines how clicking on a search result link will open the page. The options available are:

- The current window (default)
- A new window
- A frame or window with the specified name.

Note that this is effectively equivalent to a `target=` attribute in an anchor HREF link in HTML.

Default to "match all search words"

This will set the default searching method to "match all search words" instead of "match any search words".

Show time taken to perform searches

When enabled, the search page will measure the time it takes to perform each search process, and display a "Search took x seconds" line at the bottom of each search page.

Show Zoom info line

This displays a "Powered by Zoom Search Engine" line at the bottom of each search page.

Exact phrase

This option is not available in the Javascript version, or when context descriptions are disabled.

This feature allows users to search for words in a specific order as they appear on a page by surrounding the words in quotes (eg. a search for "wrensoft zoom" (with double quotes) will not match "zoom wrensoft" or "wrensoft something zoom").



Note: Exact phrase searches can be significantly exhaustive (and thus, slow) depending on how common the words in the search phrase are, on your site. Because of this, it is a user-defined option whether to allow users to enter exact phrase searching (or ignore them and treat them as multiple keywords, as in previous versions).

The Optimization switch on the [Limits](#) tab allows you to eliminate the chances of slow, exhaustive searches. Changing this setting will (among other things) decrease the chances of exhaustive searches. When a phrase is entered that is too slow to search thoroughly, a limited set of results are returned and a note is displayed advising the user to specify a more precise search phrase (usually by adding some less common words to the existing phrase).

Provide option to “Sort results by date”

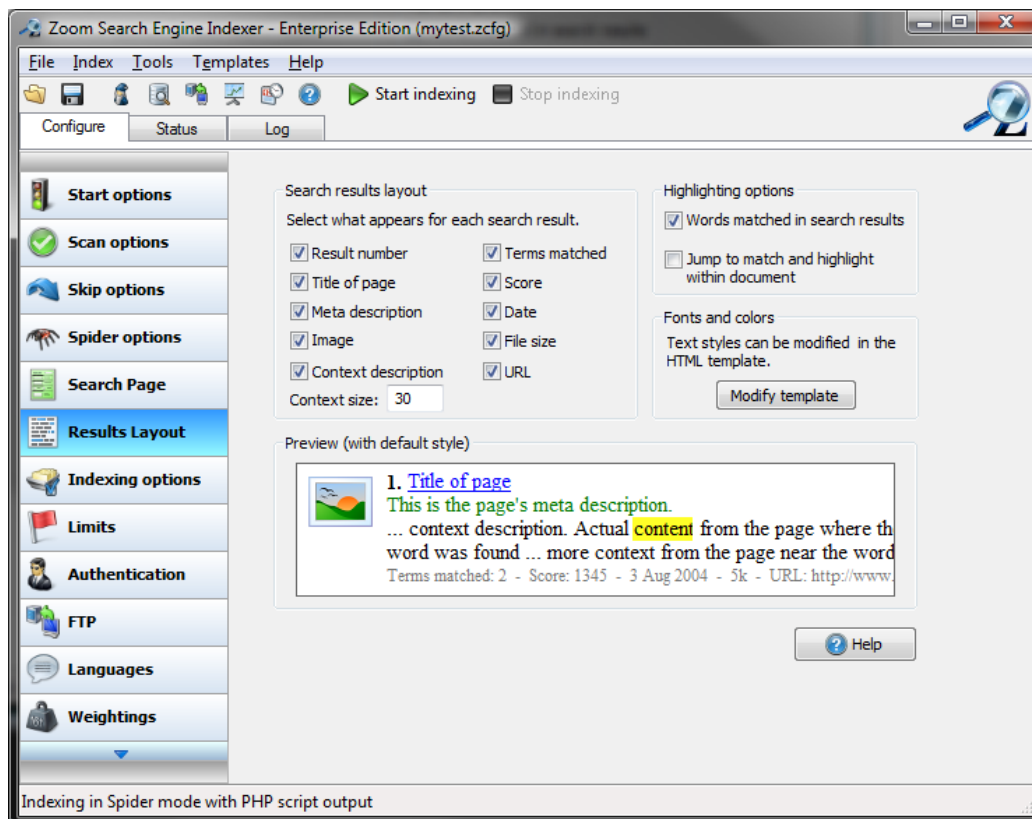
This option allows the user to sort their search results by date. The date is determined by the last-modified date and time of the file during indexing. For files which do not have a useful modified date or time (eg. dynamically generated web pages and server-side scripts), use a Meta tag to specify this information (see "[Specifying a last-modified date for your web pages](#)").

Provide spelling suggestions when less than x results found

This option is not available when the Javascript platform is selected.

This option provides users with alternate spelling suggestions for their search query when less than the specified number of results are found.

2.2.5 Results layout



Search results layout

This set of options allows you to modify the appearance of the search results. You can select or disable the elements that should be displayed per search result. Note the Preview pane at the bottom of the window gives you a representation of what your results may look like with your current selected settings.

Result number: Displays the rank number of the search result, with the most relevant page being numbered 1.

Title of page: Displays the link to the result with the page title as the link text. When disabled, the link text will simply be the URL.

Meta description: Displays the Meta description when available.

Image: Allows an icon or thumbnail to appear alongside search results. Images will only appear for files which have a ZOOMIMAGE tag specified or an icon or thumbnail configured for that file extension. See "[Icons and Thumbnails](#)" for more information.

Context description: Displays the surrounding words from the content of the page where the searched word was found. Note that this is not available when Page Content is not being indexed. It is also not available for the Javascript version due to technical constraints with the scripting language.

Context size: The number of words surrounding the searched word that should be displayed as part of the context description.

Terms matched: The number of user searched terms that were matched on this page.

Score: A score representing the relevance of the page found, depending on the number of words matched on the page, and where the word was found on that page (can be configured with Word boosting parameters set in the Indexing Options tab).

Date: Displays the date for the page. This is determined by the Last Modified date and time for that particular file. When this information is not available, no date is displayed. Note that pages that are dynamically generated by server-side scripts (e.g. PHP or ASP pages) often have no date associated with them unless they are scripted to return a date. See "[Specifying a last-modified date for your web pages](#)" for more information.

Highlighting options

Words matched in search results

This option enables the highlighting of keywords found in the search results. You can customize the appearance of this highlighting by changing the CSS of the search page template. See "Fonts and colours" below.

Jump to match and highlight within document

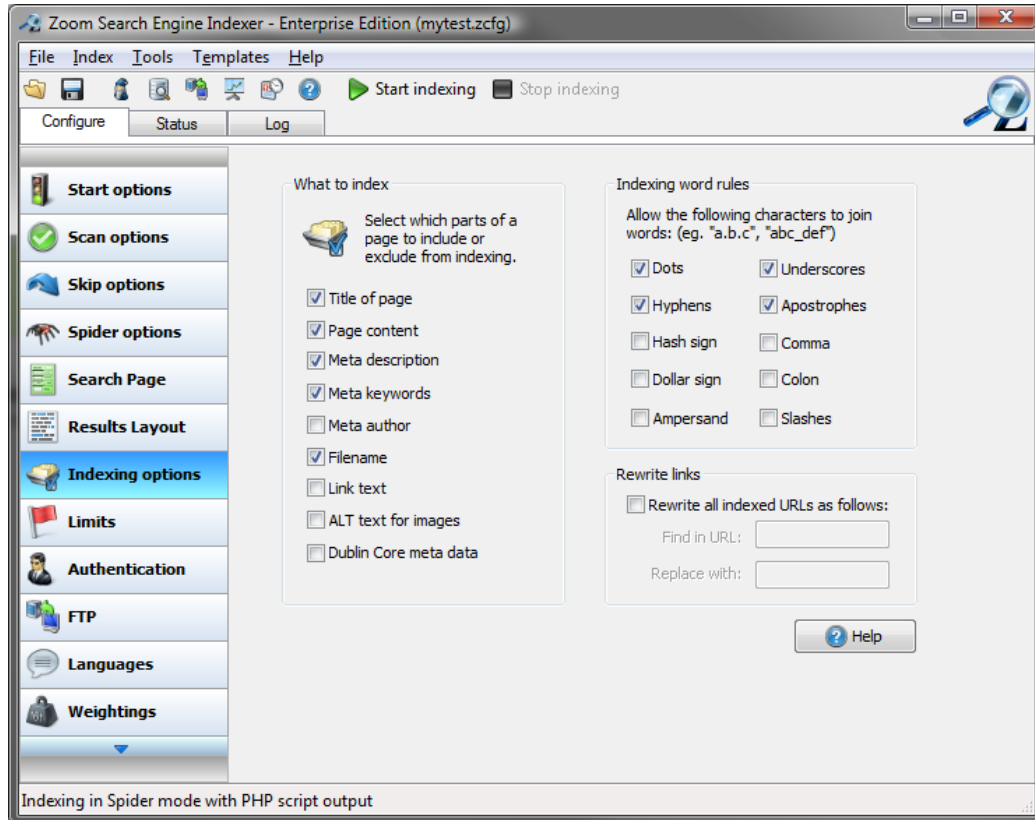
This option enables support for highlighting to occur on the actual page of your website when you click on a search result. However, this requires you to add some Javascript to each page of your site where you want this feature to take effect. It will highlight the word that the user was looking for in the search engine, as well as scroll the window down to the first appearance of the word. For instructions on how to setup this feature, see "[Enable jump to match and highlight within document](#)".

Fonts and colours

You can change the fonts, text colours, and text styles of your search results by editing the CSS styles defined at the top of the template HTML file. Clicking on the Edit HTML template button will open up the search template source file in your default HTML editor.

See "[How do I customize the look of my search page?](#)" for more information and a definition of the CSS styles you can use to change the appearance of the search results.

2.2.6 Indexing options



What to index

You can specify the parts of a page that should be included or excluded from indexing here. This includes the page title, content, and filename. Meta information can also be indexed such as meta descriptions, keywords, and author information. By excluding certain sections of pages, you can make index data files smaller and the indexing procedure faster, and less memory intensive. It may also help make your searches more accurate by including or excluding only the relevant sections of a page.

Dublin Core meta data can also be indexed. By enabling this option, Zoom will index DC.Title, DC.Subject, and DC.Identifier meta tags as described by the Dublin Core Metadata Initiative (DCMI).

Note that "Link text" and "ALT text for images" only affect the indexing of these elements for the target or destination file. That is, if a text link appears on "pageA.html" to "imageB.jpg", with the link text (or ALT text) "picture of my pets", then these words will be indexed for the file "imageB.jpg", and NOT for "pageA.html".

Indexing word rules

This allows you to specify which characters should be allowed to act as a join character between two words. Otherwise, these characters will act as separators of words (for example, if the 'dash/hyphen' character is a join character, words such as "web-based" will be indexed as one word. Otherwise, it

would be split into two words, "web" and "based"). Note that the character must be immediately preceded and followed by another valid character to be indexed.

A list of the characters available for this option:

Name	Character	Example words indexed
Dots	.	"F.B.I.", ".NET", "www.mysite.com", "32.10"
Hyphens	-	"web-site"
Underscores	_	"temporary_name"
Apostrophes	'	"John's", "it's"
Hash sign	#	"#3218B", "Serial#", "ID#"
Dollar sign	\$	"\$50"
Comma	,	"60,000"
Colon	:	"Ref:A", "Exhibit:A"
Ampersand	&	"A&B", "&var"
Slashes	/ \	"either/or", "12/5/2007", "\myfiles\pages.txt"

Rewrite links

This option allows you to rewrite the indexed URLs of the pages indexed. This can be useful if you are spidering a development version of your site on a test server (eg. <http://test.mycompany.com/>) and creating index files to go on the live server (eg. <http://www.mycompany.com/>). You would do this by specifying rewrite options to replace all instances of "<http://test.mycompany.com/>" in the indexed URLs with "<http://www.mycompany.com/>".

You could also use this option to change all the search result links to be relative rather than absolute by replacing the domain (eg. "<http://www.mysite.com/>") with a relative path (eg. [./](#)" or [../](#)"). We only recommend this for users who are very familiar with relative linking and understand that the linking would only work if the generated search files are placed in an appropriate folder on the server.



Note: Using the Rewrite Links option disables the ability to use incremental indexing on the produced set of index files. This means you will not be able to perform an incremental update, or add/remove pages from the index without re-indexing your site entirely. For more information on these features, see "[Incremental indexing](#)".



Note: Sitemaps are also affected by the Rewrite Links option. This means that your text and XML sitemaps will contain the URL as they would be after applying the rewrite link rules, and not the URLs as indexed. Be careful with XML sitemaps, since you need to specify a Sitemap Base URL for which the pages must fall within, or otherwise it will be filtered out. You will need to make sure that the Rewrite Link settings do not change the URL such that it does not satisfy the Sitemap Base URL any more. For more information, please see "[Sitemaps](#)".

2.2.7 Limits

These are the defined limits of the indexer, setting the maximum number of files to scan, maximum number of unique words to index, maximum file size scanned, and the number of characters used for the description of each file.

In the free edition of Zoom, these limits are restricted to a typical size of a free, personal website (50 pages, 15,000 unique words, 100,000 bytes, and 150 characters respectively).

For larger websites, or commercial projects, we recommend the **Standard Edition** license which is capable of indexing larger website (up to 100 pages) and provide plugin support for PDF, DOC, and other file types.

For commercial developers, or owners of larger sites, we recommend the **Professional Edition** license which allows you to change these limits manually, up to a maximum of 200,000 pages and half a million (500,000) unique words.

For extremely large sites (of over a million pages) or a cross-site search engine spanning many different websites, we would recommend the **Enterprise Edition** which has no limit on the number of pages or unique words you can attempt to index (only limited by the amount of memory in your indexing computer).

For more information on the differences between these editions, please visit our webpage at: <http://www.wrensoft.com/zoom/editions.html>

Max files to scan

This specifies the number of web pages or files that can be scanned and included in the index.

Max unique words

This specifies the number of unique words (NOT total words) that can be indexed. Note that the English dictionary generally has around 50,000 unique words but this does not include people's names, brand names, conversational slang, etc. Note also that unique words only count the "base word", so when stemming is enabled, "learned", "learning" and "learnt" would only be counted as 1 unique word ("learn").

Max. file size scanned

This is the maximum size in KB (kilobytes) of a file that can be scanned. This is not the total size of all files indexed, just the size of the largest file to index. Also note that it is specified in KB, which means 1024 KB = 1 MB.

The notice message that appears when you specify over 10 MB is only there to warn users who accidentally over-specify this amount, due to confusing KB with bytes.

Max Description length

This is the number of characters used for the static description of each file (either the Meta description or an extracted portion of the content) in the search results listing. By increasing this amount, you can include a larger portion of text as the page description in your search results.

Truncate titles longer than x

When this option is enabled, Zoom will truncate long page titles before indexing them. This prevents pages with very long titles from stretching out the layout of the search results unnecessarily.

Limit files per start point

This allows you to specify a limit for each start point of the indexer, before it stops, and moves on to the next start point. It can not be greater than the maximum pages to scan limit specified earlier. You can also specify a different limit for each individual start point from the "[Advanced spider URL options](#)" window. Note that when both the global and individual limit is set, both settings will apply, so which ever limit is first reached (ie: the lower limit of the two), will cause the indexer to stop indexing the current start point.

Limit words per file

This allows you to specify the maximum number of words to index from each file. Once this limit is reached, the indexer will move on to indexing the next file. This can be useful if you are indexing a very large archive of content, and only consider the first 100 words on a page to be useful. Another example is when you are indexing PDF documents, which may contain many pages. Using this feature you can limit the indexing to the words on the first page (with an approximation of 600 words per page for example).

Optimization

This slider bar allows you to control the behaviour of the search script or CGI when searching large sites. It allows you to give preference to either faster searches - at the cost of accuracy and potentially omitting some search results, or more accurate search results - at the expense of slower searching. This is particularly influential on exact phrase matching, and searches which may return a huge number of results (eg. over 1000).

For the PHP and ASP versions, this control currently only reduces the accuracy of exact phrase searches (which means that it may miss some phrases) and only gains some speed when performing a phrase search.

For the CGI version, it controls the maximum number of matches for a search term that is considered before asking the user to specify a more specific or less common search query. Since the CGI version is already very fast and efficient, it is really only worth changing if you are searching 100,000+ pages or more. With the default settings, we have tested Zoom to search over a million pages in less than a second (this can vary depending on the content indexed).

2.2.8 Authentication

This tab allows you to enter login and password details to spider a secure site that requires authentication.

There are two common implementations of authentication on websites:

- (1) HTTP authentication
- (2) Cookie/session based authentication

You will need to identify the type of authentication you require.

HTTP authentication

HTTP authentication usually appears as a special login window (when you access the page in your browser) and is a standardised method of authenticating over HTTP, implemented by the web server.



Example 1. A typical website with HTTP authentication accessed via Internet Explorer

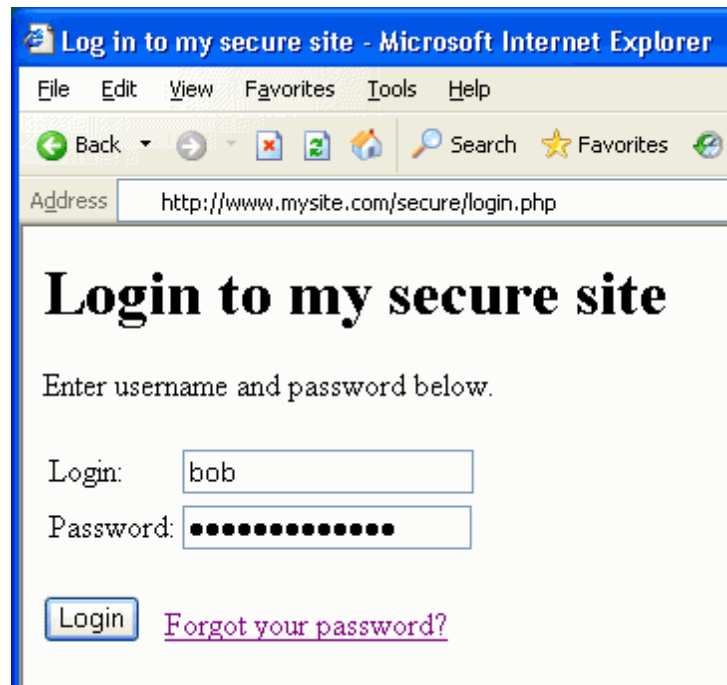
If your website uses HTTP authentication, you can simply enter your login information into Zoom (under the "Authentication" tab of the Configuration window) and the spider will automatically login when required and index the protected parts of your website.



Note: Authentication information is saved in the ZCFG file when you select Save configuration from the file menu. The password is obfuscated, but not heavily encrypted. For sites with security sensitive information, we recommend creating a special user account for indexing on the web server where possible, so you can disable this user account after indexing.

Cookie or session based authentication

Cookie-based authentication however, usually appears as a form on a page, and is implemented by server-side scripts (such as PHP or ASP or Cold Fusion).



Example 2. A typical website with cookie-based (or session-based) authentication

Zoom now provide a way to automatically login such pages. To do so you will need to provide the following information and settings:

- **Use cookies from Windows and IE:** This option enables cookie support in Zoom. You will need to check this option to access cookie-based authentication websites. Note that Zoom uses Windows' internal cookie cache (as part of WinInet) which means that it shares cookies with Internet Explorer.
- **Automatic login on following page (URL):** Here, you should specify the URL to the page containing the login form. Using the example above (Example 2 screenshot), this would be "http://www.mysite.com/secure/login.php". On this page, the HTML for the form may look like the following:

```
<form action="?op=login" method="POST">
Login: <input name="username" size="15"><br>
Password: <input type="password" name="pass" size="8"><br>
<input type="hidden" name="secret" value="handshake">
<input type="submit" value="Login">
</form>
```

It is important to look at the HTML for the login form because you will need the name for the login variable and the password variable in the next steps.

- **Login variable name:** This is the name of the login input text box. That is, it is the part after "name=" for the input tag where you will enter your login. In the above HTML example, this would be "username".
- **Your login:** This is the actual login you would be typing into the text box normally. In the above example, this would be "bob".
- **Password variable name:** This is the name of the password input text box. It would be the part after "name=" for the input tag where you enter your password. In the above HTML example, this would be "pass".

- **Your password:** This is the actual password you would be typing into the text box.

Note that the automatic login process will submit these values to the action= URL specified for the form. It will also pass along any hidden variables within that form as they are often also required by the login process.



Remember: If you are using one of the above methods to allow the spider to login to your cookie or session-based authenticated site, you need to make sure that the spider does not follow a link to the "logout" page, subsequently logging itself out of your website. You can prevent this by simply specifying the logout page in the "[Skip pages and folder list](#)" (in the Configuration window, under the "Skip options" tab), eg. "logout.asp" or "&logout=1", etc.

When automatic login will not work

Automatic login may not work on some sites or forums with anti-spider/anti-bot mechanisms that prevent exactly this type of automatic logins (they are usually put in place to avoid spam bots). In such cases, you will need to manually login with Internet Explorer. While making sure you have "Use cookies from Windows and IE" enabled in Zoom. With this feature enabled, you can login to your website using IE, and then (without logging out first), start indexing in spider mode. The spider would be authenticated with the required cookie due to the shared cookie cache.

2.2.9 FTP

Clicking on this tab will bring up your settings for uploading files to your web server via FTP. See the dedicated "[FTP](#)" section for more information.

2.2.10 Languages

Zoom provides options to cater to searching sites of different languages and encodings. For additional information, please see "[International / foreign language support](#)".

Encoding and character sets

This will be the encoding and character set (charset) of the resultant search page and index files. Zoom will scan files and convert the content from the character set specified on the web pages to the encoding selected here. Files which do not have a charset specified will be assumed to be the character set specified here.

Change your settings regarding foreign language support and character encoding. Firstly, if your website uses Unicode UTF-8, you must enable the "Use Unicode" option. Otherwise, specify the encoding (also known as charset) used. "windows-1252" is the most common option for English, French, German, and a number of other Latin based languages.



Tip: If you are indexing a single site (or several sites) where all the pages use the same charset, you should simply set this to be the same charset which is used on your website. However, if you are indexing web pages of varying character sets, we would recommend setting this to UTF-8 so that you can have a universal search page which will work with all the different languages. The content will be converted to Unicode before being indexed.

International searching options

Enable accent/diacritic/ligature insensitivity: This will map all occurrences of accented characters to their non-accented equivalent (eg. ó, ò, ô, etc. will all be treated as "o"). With this enabled, a user can enter the search word "cliché" and it will find all occurrences of the word on your website spelt as either "cliché" or "cliche". You can now specifically enable or disable this feature for accents (ó, ò, ô, õ, etc.), umlauts (ä, ë, ï, ö, ü), and ligatures (å, ø, æ).



Tip: Note that when this feature is enabled, and highlighting is also enabled, some words found in the context description may not be highlighted.

Use digraphs for umlauts: When this option is enabled together with umlaut insensitivity, characters like "ö" will be considered the same as "oe" as opposed to "o". Similarly, "ä"="ae", "ü"="ue", etc.

Support single-case languages (eg. asian languages): This should only be used if you are using a language where there is no case-difference and problems can occur when the script or indexer attempts to convert case (such as some East Asian languages).

Substring matches for all searches: This will mean the script will consider search words that occurs within another word to be considered a match. (Eg. a search for the word "hot" will match "hotcake", "shotgun", etc.) This may be useful for East Asian languages where words are not distinguished by spaces and you always want to search for a single character within a set of words.

Strip Arabic diacritic marks in words: This option will strip diacritical marks from Arabic words which are typically never searched for, and is only used to represent accurate pronunciation of the word.

Stemming

This option is not available in the JavaScript version.

When this feature is enabled, search results will match similar words or words which are derivatives of each other (e.g. plurals). For example, searching for the word "fish" will return pages containing the words "fishes", "fishing", etc.

The CGI version features an improved stemming feature which can be configured for languages other than English. PHP or ASP only supports English stemming. Stemming is not available when single-case matching (i.e. "Support single-case languages") is enabled.



Note: If your website contains multi-lingual content (that is, your index will need to allow the searching of text in more than one language), then it may be best to disable stemming. The stemming algorithms are language specific, and will not work well when it is imposed on text of a different language.

Search page language

You can modify the text that appears on the search page and search results, by customizing the Zoom Language Files (.ZLANG files). Almost every bit of text on the search page can be modified or translated, including "Search results for..." and "x results found", etc. This allows you to translate the search page to

the language of your choice, without having to modify the search script.

Zoom also comes with a few pre-translated language files which both serve as examples, and allow you to create French or German search pages straight out of the box, by selecting it from the drop-down menu.

For more information on how to create your own translations, please refer to "[Translating the search page](#)".

2.2.11 Weightings

Word weighting

These drop-downs allow you to increase or decrease the importance of words found, depending on whether it was found within the following:

- Title of the page
- Meta description of the page
- Headings of the page (ie. within <H1> and <H2> tags)
- Filename
- Meta keywords or ZOOMWORDS tag of the page
- Link text and ALT text for images

The boosting can be specified between 10 levels, ranging from -5 Deboost to Normal, to +5 Boost. For example, if the Title is set to have a word weighting of "+2 Boost", then each words which appear within the page title would be treated as though have appeared twice on that page.

This makes it possible for you to give priority to pages (placing it higher up in the results), when the word is found in certain sections of the page.

Page boosting

You can also boosts ALL words found on specific pages, by use of the ZOOMPAGEBOOST meta tag, eg.

```
<meta name="ZOOMPAGEBOOST" content="5">
```

Putting this on the most important page on your site would help make it appear higher up in the search results. You can probably use less than 5 to do the same thing on a small site. Similarly, a negative value would decrease the weight of words on that page. ZOOMPAGEBOOST values can also range from -5 to 5.

Word position

By changing this adjustment, you are altering the preference given to the following:

- Multi-word matches that are found closer together on the same page.
- Pages where words are found throughout the whole page (and not limited to a section of the page).
- Occurrences of words found at the top of the page.

Content density

This is an automatic weighting adjustment that is made by the indexer, based on the word density of the page. With "Standard adjustment", the weighting of words found in a large file (such as a 50+ page PDF document) will be lowered so as to prevent such files from swamping the results and always considered the most relevant. This will effectively give preference to small and medium sized documents. "Strong adjustment" provides an even greater level of scaling, and "No adjustment" would disable this feature so that all files are treated equally.

Give preference to short URLs

This is an automatic adjustment made by the Indexer, based on the length of the URL of the page being indexed. With "Standard adjustment", the weighting of a page may be slightly lowered if the URL is considered lengthy. Similarly, a page with a short URL may receive an increase in weighting. This will effectively give preference to "home pages" of the site, as opposed to those which are several folders deep. "Strong adjustment" provides an even greater level of scaling, and "No adjustment" would disable this feature so that all files are treated equally regardless of their URL length.

2.2.12 Content filtering

Content Filtering allows you to filter out an entire page based on words found within the page's content. A list of filter words can be entered, prefixed with a "+" or a "-". You can specify positive filters (keywords beginning with a "+" character) which means that only pages with these words will be indexed, or you can specify negative filters (keywords beginning with a "-" character) meaning that pages containing these words will NOT be indexed.

This can be useful for two reasons:

1. It helps if you want to create a specialised 'vertical' search engine. For example you could create a search engine about pets. In this case the word filter list might look like:
+dog
+cat
+bird
+mouse
+hamster
+pet
2. You might want to avoid indexing some types of content. For example if you were building a religious search engine or a search engine for children, you might want to use negative filters:
-adult
-casino
-sex

You can also filter out pages based on the URL by using the [Skip options](#) by using the skip pages list. This method however, requires that you know the pages' URLs in advance, and manually determine which page has wanted or unwanted content. This is not always possible when indexing external sites, so content filtering solves this problem.



Tip: Note that content based filtering will be less efficient than URL based filtering because each page must be downloaded before it can be filtered. With URL based filtering (using the "[Skip pages](#)" list), the page can be discarded before it is downloaded, thus speeding up the indexing process. So URL filtering should still be used when possible.

Content filtering is applied to the HTML source code of the page being scanned. This means that you can filter by HTML tags, for example:


```
-<meta name="robots" content="noindex">
```

However, note that this means that support for filtering by exact phrases is limited because the words may be broken up by HTML tags such as line breaks, comments, etc. (For example, it will not match if we are trying to filter out pages containing the phrase "adults only", and the page actually contains the HTML "adults
only").

2.2.13 Categories

Clicking on this tab will bring up the Categories manager. See the following "[Categories](#)" section for more information.

2.2.14 Sitemaps

This option allows you to generate a plain text sitemap (compatible with Yahoo Sitemaps) and a XML sitemap (compatible with Google Sitemaps) for your website at the same time as having Zoom index your website.

Sitemaps can be useful for submitting your websites to Internet-wide search engines such as Google and Yahoo, helping their spiders index your website more quickly and increasing your web presence. You may also find it useful for maintenance or other development purposes. The sitemap includes information such as the last modified date and the relative priority of the page within the website. Note that you should only create sitemaps for individual websites (and not sitemaps that span multiple websites).



Tip: Note that you can use Zoom's Incremental Indexing to update your search index, as well as your sitemap, in one go without having to re-spider your entire website.

Text file URL list (Yahoo Sitemap compatible)

By selecting this option, Zoom will create a file named "urllist.txt" in your output directory at the end of indexing. This is a simple text file containing a full list of the web pages that have been indexed, compliant with the Yahoo Sitemap requirements, and can be submitted to Yahoo for search engine submissions and other search engines accepting a similar sitemap format. It could also, of course, be used for your personal development or maintenance purposes, to keep track of the active files on your website.

XML sitemap (Google Sitemap compatible)

By selecting this option, Zoom will create a XML sitemap which is compatible with the Google Sitemap specifications. The files will be created in your output directory at the end of indexing.

As required by the Google Sitemap protocol, each XML file can only contain a maximum of 50,000 URLs. Multiple XML sitemap files are created when a website exceeds this number of pages (named "sitemap.xml", "sitemap2.xml", "sitemap3.xml" and so forth). A sitemap index file is also created which contains a list of these individual sitemap files ("sitemap_index.xml").

For more information on the Google Sitemap protocol, you can visit the official web page here: <https://www.google.com/webmasters/sitemaps/docs/en/protocol.html>

Sitemap Base URL

Note that a Sitemap Base URL is required to generate a XML sitemap. This is necessary for two reasons:

1. To determine the URL of your split sitemap files (e.g. "sitemap2.xml", etc.) should it be necessary to generate a sitemap index ("sitemap_index.xml") as explained above.
2. To exclude any indexed URLs which are under a different domain (or base URL) to the location of the sitemap file, as required by Google's Sitemap specifications.

With the default option "Include only URLs within the Sitemap Base URL" selected, the latter requirement is enforced to obey Google requirements for sitemap files to only contain URLs which are below the base URL of the sitemap file. For example, you can not normally have a URL such as "http://www.myotherdomain.com/mypage.html" included in a sitemap file located at "http://www.mysite.com/sitemap.xml".

However, new Google rules make an exception for this when your multiple domains/sites have been verified with Google. In this case, you can submit a sitemap file containing URLs to multiple domains (known as "cross submits"). To create a sitemap for "cross submits" in Zoom, you should select the option to "Include all indexed URLs". Find out more about verifying your sites with Google here: <http://www.google.com/support/webmasters/bin/answer.py?answer=35179>



Note: When "Include only URLs within the Sitemap Base URL" is selected, Zoom will automatically exclude any URLs from your XML sitemap if it is not within the specified Sitemap Base URL, even though that URL may have been indexed (or it may be a recommended link). This means that your XML sitemap may not contain all the URLs indexed if this is setup incorrectly or if you are indexing with multiple base URLs. See "Uploading your XML sitemap files" below for more information.

By default, each XML sitemap file will contain the URL and the last modified date of the pages that were found on your website. There are also the following options available:

Use ZOOMPAGEBOOST values for Priority field in sitemap (XML only)

By selecting this option, the XML sitemap file(s) will also contain a priority value (from 0.0 to 1.0) which indicates the importance of a page, relative to the rest of your website. It will determine this value based on the ZOOMPAGEBOOST meta tag which is also used for prioritizing pages within Zoom's index (see "[Weightings \(Page Boosting\)](#)" for more information).

The default priority is 0.5, for any page where this attribute is not specified. These values will correspond to the ZOOMPAGEBOOST value (where a -5 page boost is equivalent to 0.0 priority, 0 page boost is 0.5 priority, and +5 page boost is 1.0 priority).

Uploading your sitemap files

The sitemap files can also be automatically uploaded at the end of indexing, along with your search files. You will need to enable this option on the Sitemaps tab and the "Automatically upload files at the end of indexing" option on the [FTP](#) tab of the Configuration window.

You could also upload your sitemap files manually using a third party FTP client if you prefer.

Notes on uploading Google XML Sitemap files

You may need to specify a different folder or path on the server for your sitemap files, as the Google Sitemap protocol requires the files to be located at the base URL of the files which have been scanned. This means that if your sitemap files contain URLs such as:

```
http://www.mysite.com/index.html
http://www.mysite.com/story/page1.html
http://www.mysite.com/news/archive/test.html
```

Then your sitemap files must be hosted at the common base URL of the above files, and that is the root directory at <http://www.mysite.com/>

However, if your sitemap only contains URLs such as:

```
http://www.mysite.com/news/index.html
http://www.mysite.com/news/page1.html
http://www.mysite.com/news/archive/test.html
```

Then you should have your sitemap files hosted at: <http://www.mysite.com/news/> and this should be your Sitemap

Note: the folder/path specified in this window should either be a path which is relative to your home directory (eg. "public_html/news/") or an absolute path on the server (eg. "/usr/home/myaccount/public_html/news/"). The folder path should NOT be a URL and should NOT start with "http://".

For more information on the Google Sitemap protocol requirements of the location of your sitemap files, please visit their webpage here:

<https://www.google.com/webmasters/sitemaps/docs/en/protocol.html#sitemapLocation>

Tip: Submitting your XML sitemap to search engines

If you anticipate that your website will not grow any larger than 50,000 pages, you can simply submit the URL to your "sitemap.xml" file to Google or other search engines using the Google Sitemap protocol format. However, if you anticipate your website growing larger than 50,000 pages at some point, we would recommend submitting the URL to your "sitemap_index.xml" page instead. This would allow your submission to stay valid even when your website grows beyond this limit, and you need to have more individual sitemap files in order to list your entire website.

2.2.15 Index log

You can configure the type of messages that are displayed in the main status log window during an indexing session here. This allows you to configure Zoom to only log certain messages types as necessary, which can be useful if you wish to minimize the messages produced so it's easier to filter through.

Log modes

There are two main logging modes.

"Basic" omits the following message types: Skipped, Filtered, Spidering, Thread info, Initialization. When this mode is selected, these messages will not be recorded and you will not be able to toggle them on or off from the "[Log](#)" window.

"Detailed" enables and logs all message types. Note that you can still filter them on-the-fly on the "[Log](#)" tab window.

Log HTML warnings

When this option is disabled, warning messages regarding invalid HTML code will be suppressed (e.g. "Suspected invalid HTML ...", etc.). Note that HTML warning messages are important to help you find

broken HTML code, and also to be alerted to why certain pages may not have been indexed correctly. If you cannot find the cause of the "invalid HTML" warning messages, you should try running the web page in question through a HTML validator such as <http://validator.w3.org/>

Log to file

You can ask Zoom to write the contents of the index log window to a file, allowing you to archive or view the log at a later date with any text editor.

To enable this, check the "Save index log to file" option and specify the filename and path of the file which the index log will be written to.

The "Debug mode" option is used to help resolve crash issues and other situations where you will need the log file to be written to disk per line entry. It will also enable additional debugging information to be displayed in the index log.

2.2.16 Advanced

Do not show wizard on startup

Stop the wizard from being displayed when Zoom is started.

Beep at the end of indexing

Play the system alert sound when indexing completes.

Search logging

This option is not available when the Javascript platform is selected. This is due to the fact that Javascript is a client-side scripting language and logging of user searches is not possible.

This enables search word logging so that searches made on your website will be recorded into a server-side log file. Note that you will have to specify the filename and path of the location of the log file on the server-side relative to the location where you will place the search script.



Note: You will need to be able to change server-side file permissions to allow write access to the log file you specify. We recommend this option to more advanced users familiar with server side file permissions only. Note that many cheaper web-hosting services do not allow you to change your file permissions, so consult your web hosting company regarding this beforehand. This can normally be done via your FTP client or through a Unix shell account by using the "chmod" command. For more help and information on setting up your server-side permissions for logging, please refer to our online FAQ: <http://www.wrensoft.com/zoom/support/logging.html>

Logged search results contain, the date (YYYY-MM-DD), the time, the IP address of the user, the search query, and the search options used. The following is an example of a log file record (on one line).

```
2004-03-05, 09:52:08, 127.0.0.1, "test search query", Matches = 6, OR, PerPage = 10, PageNum = 0, No cats, Time = 0.084, Rec = 0
```

This file can be analysed independently or with the statistics report generator (see "[Search Statistics](#)")

[Report](#)" for more information). Since the log file is in CSV format, you could also import the file directly into Excel.



Note to CGI users: if you are hosting on an Apache server (in a "cgi-bin" directory), remember that most Apache servers do not allow text files to be accessed via HTTP from the cgi-bin directory. This means that if you store your log file in "/cgi-bin/logs/", you will not be able to access this file via HTTP using your browser or Zoom's Statistics Report tool – instead, you will only be able to download the file with FTP and analyse it locally. Alternatively, make sure you specify a log file path outside of the "cgi-bin" directory.

Custom script source path

If you wish to use a copy of the search script (`search.asp`, `search.php` or `search.js`) that is located in a different folder besides the one in the Zoom installation directory, then you can change the path here. This is only necessary if you are making modifications to the search script source code, and you wish to maintain multiple versions to be used with different sites.

Embedding script

This option allows you to specify the URL used to link back to the search script. By default, it should be left blank.

However, if you wish to embed the search script in your own server-side scripts (eg. you have included `search.php` into your own PHP file, or included the CGI within a PHP script, etc.) you can then specify the filename of your embedded script here so that it will link to the correct file.

Spider User-Agent (Enterprise Edition only)

Users of the Enterprise Edition of Zoom can modify the User-Agent text that is used to identify the spider when it crawls web servers. By default, the spider identifies itself as "ZoomSpider - wrensoft.com [ZSEBOT]".

If you specify a name here, e.g. "mysearchengine.com", your spider will then identify itself as "mysearchengine.com [ZSEBOT]". This allows you to include your website name or other contact information, should you wish to allow third-party websites to identify you, or be able to contact you if required. It can also be helpful if you wish to add specialized behaviour to your website, so that, for example, you could allow your personal spider exclusive access to certain pages on your site, and prohibit access for any other user-agents.

Disable charset enforcing on search script

This option disables the search script's attempt at specifying the charset via a HTTP header. This is done to ensure that the corresponding character set used for the encoding of the index data, is also used for the search page. However, if you are an advanced user trying to embed the search script within your own scripted pages, this extra HTTP header may not be required (as you may already have HTTP output before the search script is executed). In this case, you can safely disable this feature here.

XML/RSS output (CGI only)

The CGI version can now return search results in XML/RSS format. Selecting this checkbox will enable the XML output as default, and clicking "Configure" will allow you to specify some XML specific settings.

See "[XML/RSS output](#)" for more information.

2.2.17 Synonyms

This feature allows you to specify synonyms, variations of words, common misspellings, etc. and allow you to map them to an equivalent word in the index. This can be very helpful if many users on your site are searching for names which are different to the ones you used on your website, but have the same meaning.

For example, the word "question" could be made equivalent to "inquiry,enquiry,query,questions", so that when a user searches for any of the aforementioned words, they would get the same results.

A synonym definition has two fields:

1. **Word:** This is the word that the synonyms will be mapped to. It must be a word that actually appears in the content of your website.
2. **Synonyms:** This is a list of words separated by commas that will be considered equivalent to the indexed word. When a user searches for any of these words, they will get the same search results as if they searched for the indexed word. All occurrences of the words in this synonym list will also appear as a search result when you search for the indexed word.

You will need to re-index your website for any new synonyms to take effect.

Note: When a word is synonymous with another word, it is implied that they are equivalent both ways. For example, if an entry for "cat" has a synonym "feline", then "feline" is also considered the same as "cat". You do not need to specify another synonym entry to define this.

You can also Import and Export your list of synonyms to and from a text file. This should be a text file containing a synonym entry per line, in the form of:

```
word=synonym1,synonym2,synonym3, etc...
```

You can create this text file in any text editor (such as Notepad). An example text file which can be imported into the Synonyms list would be as follows:

```
question=inquiry,enquiry,query
problem=trouble,issue,bug,help
rat=mouse,mice
dog=puppies,puppy,canine,dawg
```

Note that you can only have single word synonyms, and not synonyms for phrases.

Zoom will expect this text file to be in the same encoding or charset as specified on the "[Languages](#)" panel. So if "Unicode (UTF-8)" is selected here, the file would be expected to be in UTF-8 encoding. Only the encodings available on the Languages panel are supported.

2.2.18 Recommended links

This feature allows you to specify a list of keywords or phrases that you can associate with specific URLs. When a user does a search for a matching word or phrase, these recommended results will then appear before the rest of the search results like so:

Search results for: help

75 results found

8 pages of results.

Recommended links

[Zoom Search Engine Users Guide](#)
Essential manual to installing and using Zoom. Illustrated and detailed PDF document.
<http://www.wrensoft.com/zoom/usersguide.html>

[Technical Support and Help for Zoom](#)
Support pages for Zoom containing a comprehensive FAQ section, forums, and more.
<http://www.wrensoft.com/zoom/support/>

Sorted by relevance / [Sort by date](#)

1. [Zoom Search Engine - Users Guide](#)
Download the free PDF Users Guide for Zoom Search Engine
... . It provides illustrated, and easy-to-read instructions for all of the features and options available in Zoom, to help you get your search engine up and running, and fine-tune its performance and options to best suit your site ...
Terms matched: 1 - Score: 39 - 18 May 2006 - URL: <http://www.wrensoft.com/zoom/usersguide.html>

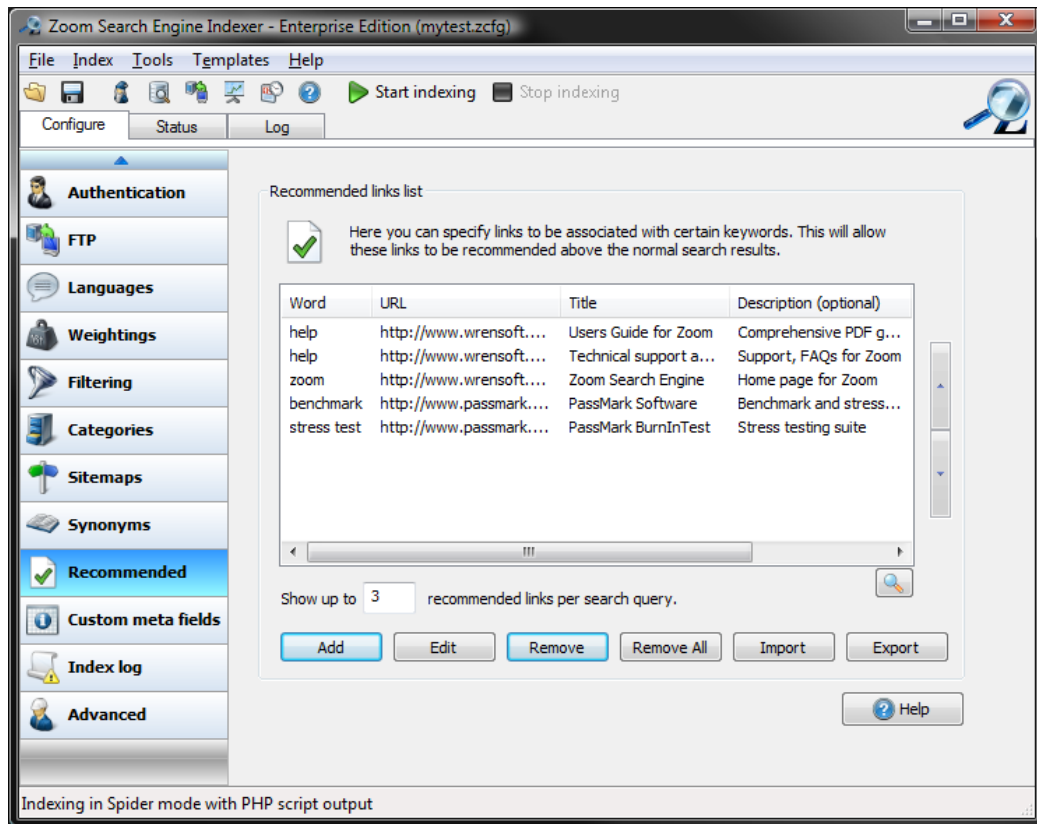
These recommended links will not only appear above all other results, but they can also be customised to stand out from the other search results. By using new style sheets classes it will, for example, be possible to make these results appear in bold or be highlighted with a different background colour (as seen above). You can also configure them to appear in the same manner as normal results (by appearing before any other search results).

The Recommended Links feature allows a wealth of new possibilities. It means you can effectively override the current #1 'organic' result with a result of your own choosing.

Some examples of use could include,

- Directing people to the right product home page based on product number.
- Selling search results positions on your web site. Doing this would provide new income streams for sites that choose to do this. How much if the #1 search position worth on your sites search engine?
- Directing people to your most profitable products

You can manage your list of recommended links from the Configuration window (under the "Recommended links" tab).



You can specify the maximum number of recommended links to display per search, with 3 links being the default maximum. It is also possible to have multiple recommendations for the same word and control the priority by adjusting the order of the rows in the list above.

Note that each recommended link must have a keyword (or key phrase), URL, and a title. The description field is optional and limited to 300 characters per recommended link. The keyword for the recommended link may contain wildcards (e.g. "*" or "?") so if you create a recommended link for the keyword "drive*" it will appear when searching for "driver", "drives", etc.

You can also import and export a large collection of recommended links from/to a text file by clicking on the "Import" and "Export" button respectively.



Tip: Recommended links will match against the full search query, so a search for 'text box' (ie. without quotation marks) will also return the recommended links entered for the word/phrase: text box. Note that some characters (such as brackets and other punctuation characters) and spaces may be stripped or trimmed from your keyword or phrase. This is so that the phrase will match the search query (which will also have such characters stripped or trimmed).

The following are some examples of recommended link entries, and an explanation of search queries that will yield its return:

Example 1.

Word:	help
URL:	http://www.wrensoft.com/zoom/support/

Title:	Technical support and help for Zoom
Description:	Support, FAQs, and more

The above recommended link will appear for any search queries containing the single word help. However, it will not return for a search containing an exact phrase (assuming [exact phrase functionality](#) is enabled, and the query uses double quote characters) such as "help file".

Example 2.

Word:	help file
URL:	http://www.wrensoft.com/zoom/usersguide.html
Title:	Users Guide
Description:	Comprehensive users guide to Zoom

The above recommended link will appear for a search query containing the exact phrase "help file", but it will not return in a search for the single word help. It will however, also return for the two word search query help file (ie. without quotation marks).

Importing and Exporting Recommended Links

If you have a large list of recommended links to add or maintain, you can import them from a text file created by any text editor. The file needs to be pipe separated (the pipe is the vertical '|' character), with a recommended link entry per line. For example:

```
Keyword1|http://www.url1.com/|Page title|This is the page description
Keyword2|http://www.url2.com/|Another Page title|This is the page description for the 2nd page
```

You can also export an existing list of recommended links to a text file of this same format. This way you can maintain them outside of Zoom if need be, and/or import them back in to other ZCFG files.

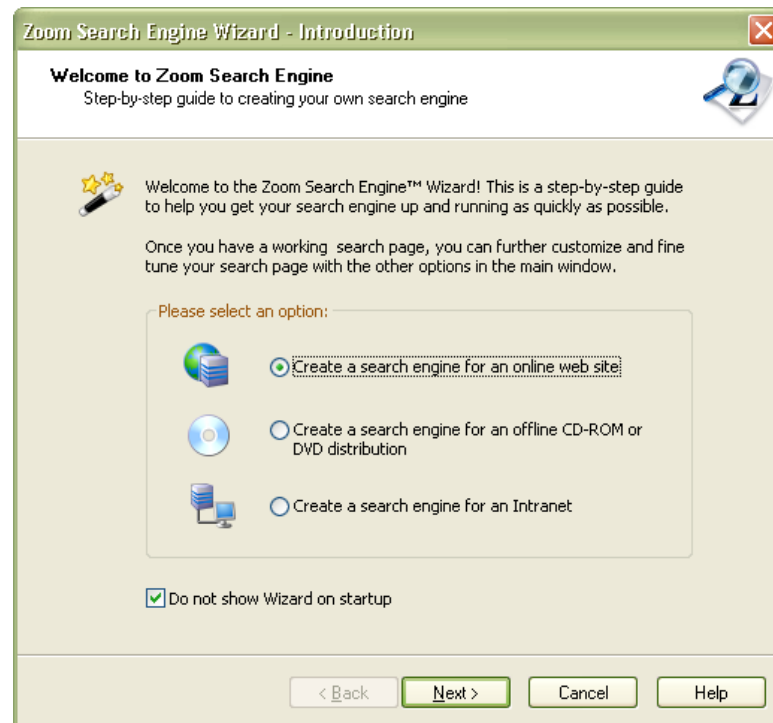
2.2.19 Custom Meta Search Fields

Clicking on this tab will bring up the Custom Meta Fields manager. See the following "[Custom Meta Search Fields](#)" section for more information.

2.3 Step-by-step Wizard

When you start Zoom Indexer for the first time, you will be greeted with a Wizard which can guide you through the basics of setting up a search engine. This aims to help new users familiarize themselves with the indexing procedure. It also provides an alternative for veteran users to get something up and running quickly.

The Wizard will guide you through the main choices involved in setting up the indexer for your website/ intranet, or CD/DVD distribution. Click on "Help" if you are stuck at any point or are unsure of what to do. You can also "Cancel" at any point and make your own configuration changes directly, as instructed in the following chapters. You can re-start the Wizard by clicking on the Magic Wand in the toolbar or by clicking on the "Tools" menu and selecting "Run Wizard".



If you do not want the Wizard to be displayed when you next start Zoom, tick the Do not show Wizard on startup check box.

For an introductory overview of Zoom, please see the "[Overview](#)" chapter.

2.3.1 Select file types step

Here you can select the file types you wish to include in your search functions. Note that you can specify exact file extensions and more options in the Configuration window under the "[Scan options](#)" tab.

Also, some file formats are only accessible with plugins installed. For more information on plugins and the file extensions supported, see "[Plugins](#)".

2.3.2 Output folder and limit step

Zoom will create a set of index files, which will be required for your search function. Specify a folder here for Zoom to create these files in. This can be a temporary folder of your convenience, or a folder under your local copy of your website.

For users of the Professional Edition (or above), you can also specify a limit for the maximum number of pages to index. More limits can be changed in the Configuration window, under the "[Limits](#)" tab.

Start indexing immediately

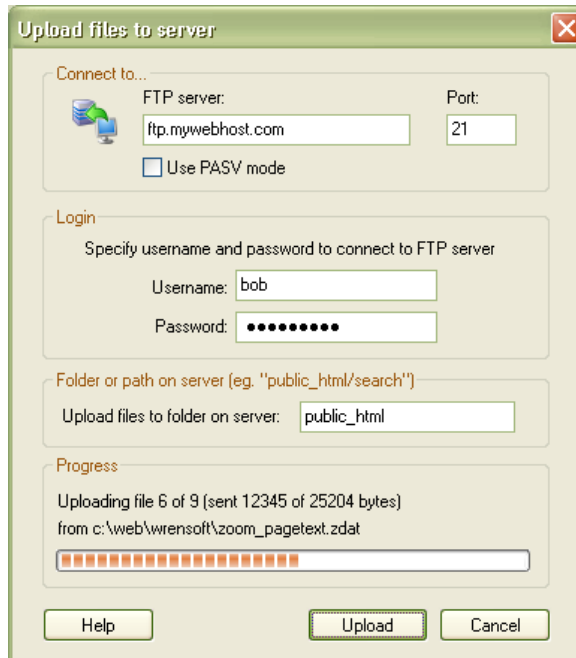
This is the default option and will attempt to index your site as soon as you click Finish for the wizard.

Do not start indexing

Select this option if you have more configuration changes or settings that you would like to make, and would like to start indexing at your own convenience.

2.4 FTP

Zoom provides built-in FTP capabilities so that it can upload the files required for your search function directly to your web site.



FTP server

Specify the FTP site address of your server, eg. "ftp.myserver.com". You can also specify the port number (default is 21 for FTP).

You can specify PASV (Passive FTP) mode if you have trouble connecting to your FTP server because of your firewall.

Username and Password

Specify the login information to your FTP account.

Folder or path on server

This should be the relative or absolute path on your FTP server where the files are to be uploaded to. Note that the folder must already exist. Check with your web host (or use a full browsing capable FTP client) to confirm the correct path if you are not sure. Typically, it may be something like "public_html/search" or "www/search" if you wish to put it in a "search" folder underneath the main site address. If you wish to upload the files to the default current directory upon logging in, specify "./" as the path here.

Automatically upload files at the end of indexing

This option is only accessible from the Configuration window (under the "FTP" tab).

When enabled, Zoom will automatically attempt to upload your files every time you finish indexing. We recommend you only use this when you have confirmed that your FTP settings are correct, and that the

files are uploaded to the correct folder or path.

This option is particularly useful for scheduled indexing, allowing you to configure Zoom to update your website without any user interaction, and on an automated, regular basis.

Do not upload search template

This option is only accessible from the Configuration window (under the "FTP" tab).

When checked, Zoom will not upload the search template file, assuming that you already have one on the server. This can be useful if you have some server-side changes that need to be made (or the template file is dynamically generated) and you do not wish Zoom to overwrite the file on the server.

Upload with .tmp filenames and rename when completed

This option is only accessible from the Configuration window (under the "FTP" tab).

When checked, Zoom will upload your search files with a ".tmp" extension so as not to overwrite your existing index files during the upload process. It will then only rename the files from the ".tmp" extension to its proper file extensions (and thus overwriting our existing index) when all files have been successfully uploaded. This minimizes the time which your search function would be unavailable while the old index files are being updated.

Set execute file permission after uploading

This option is only accessible from the Configuration window (under the "FTP" tab).

This option will set execute file permission on "search.cgi" via FTP, after the files have been uploaded. Execute permissions are required for the CGI search option only. Some servers (such as Windows-based FTP servers) do not allow you to change file permissions remotely via FTP. For such servers, we would recommend disabling this option and setting the file permissions manually. See ["Additional notes for uploading CGI"](#) for more details and instructions on how to do this.

2.5 Categories

This allows you to specify categories in your search engine, allowing your users to select from a drop-down to restrict searches to certain sections of your site. Each search result will also be tagged to indicate the category that they belong to. Searching with the "All" category will return results belonging to any of the categories.

Search this site

Enter one or more keywords to search for using the Zoom Search Engine.
Note that '*' and '?' wildcards are supported.

Search for: Results per page:

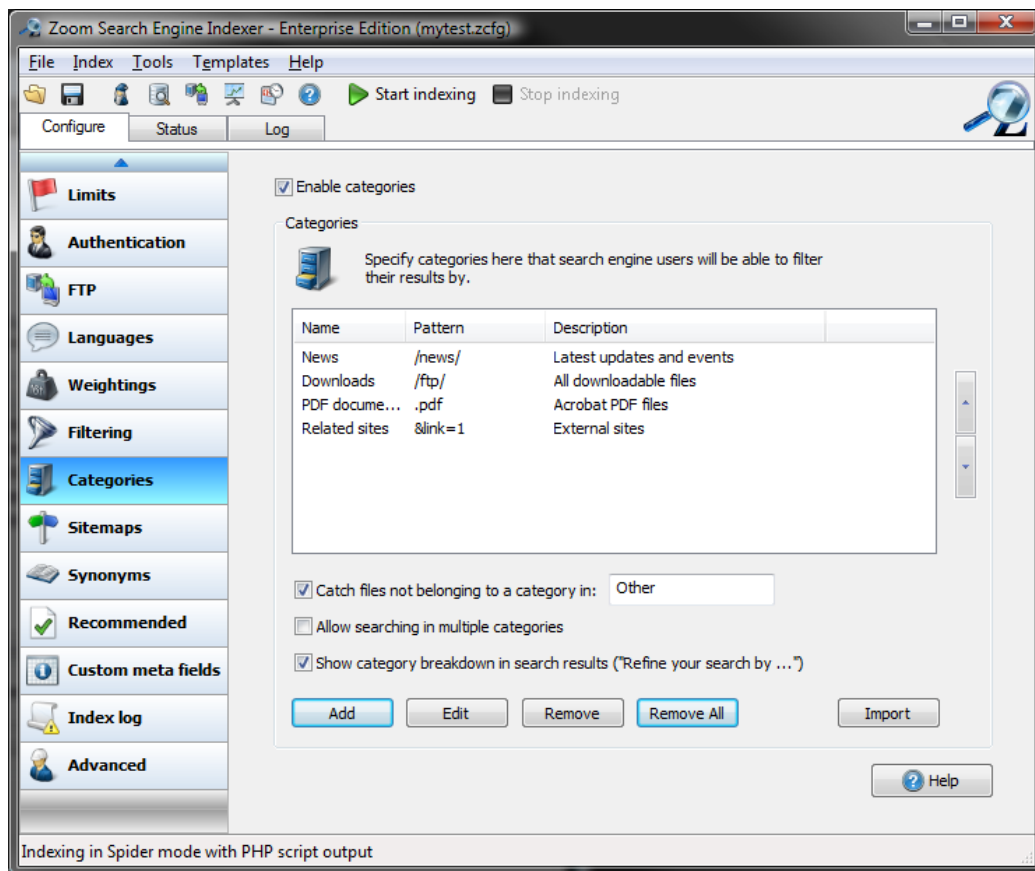
Category:

 All
 News
 Downloads
 PDF documents
 Related sites

Match: any search words all search words

Search powered by [Zoom Search Engine](#)

To enable categories support, open the Configuration window and select the "Categories" tab. This will bring up the categories management list, where you can add, remove, and edit the categories.



Adding, editing and removing categories

When creating or editing a category, you have the following fields to specify:

1. **Category name:** The name of the category must be unique. These category names will be listed in the drop-down box of the search page, or as check boxes if you have "Allow searching in multiple categories" selected.
2. **Match pattern:** The pattern is used to determine what pages belong to the category. It is matched against the pages' full path or URL. This means that should the pattern appear anywhere on a page's URL, it will be qualified for that category. Note that this includes the base URL or domain name of each page, so that you can index multiple domains, and have each defined as a separate category.

For example, a pattern of "test" will collect the following pages:

http://www.mysite.com/test.html
http://www.mysite.com/test/index.html
http://www.test.com/
... etc.

Tip: You can use patterns such as ".pdf" to create categories based on file extensions (eg. "PDF documents", "PowerPoint presentations", "Flash movies", etc.)

Category patterns are case insensitive. You can also use wildcard characters ("*" and "?") in your pattern. This means that you can create a pattern such as "/news/updates_*.pdf" which will get all PDF files, inside the news folder, with a filename starting with "updates_".

You can specify multiple patterns for each category, separated by a semi-colon character. For example, a category named "Downloads" may contain a variety of file formats with a pattern like ".pdf;.doc;.xls;.ppt;.exe" (without the quotes). Note that with multiple patterns, a file will only need to match any one of the listed patterns for it to qualify for that category.

For files which you can not match by a common pattern in the filename or URL, see "[Specifying category per file](#)".

3. **Description:** This is a short description of the category that will only be used internally in the indexer for your own convenience in the future. It is not a required field.
4. You can toggle the option of whether files caught by this category can belong to any other category. By default, this is off, which means that a file can belong to multiple categories (so long as it matches the category patterns). When this option is enabled for a category, a file caught by this category can no longer qualify for any other category.

Note that when using this option, you should be careful with the order of your categories. If your category patterns are too vague, or ordered incorrectly, you may have files grouped away in an earlier category, before it gets a chance to be matched against a later category that you intended it to fall under.

You can use the up/down buttons to the right of the list to re-arrange the order of the categories. The order of the categories determines two things:

- The order of which they will appear in the drop-down list of the search page, or (if you have "Allow searching in multiple categories" selected) the order of the check boxes on the search page.
- The process of selecting which category a page belongs to. This is important if you have any category using the "Files belonging to this category can not belong to any other category" option. This means that if a page is deemed to qualify for an earlier category (higher up in the list), it will not qualify for any other categories below it in the list.



Note: Remember that you will need to re-index for your category changes to take effect.

Catch files not belonging to a category

This option allows you to toggle the use of a "catch-all" or "default" category, which will contain any files that did not fall in one of the specified categories in the list. You can use this to create a category for "Miscellaneous files" or "Other documents" by checking this option and specifying the category name in the text box.

Allow searching in multiple categories

This option will allow your users to restrict their searches for files belonging to multiple categories. If you are allowing Zoom to generate its own search form, this will change the category drop-down box to a checkbox selection list as seen below.

Search this site

Enter one or more keywords to search for using the Zoom Search Engine.
Note that '*' and '?' wildcards are supported.

Search for: Results per page:

Category: All News Downloads PDF documents Related sites

Match: any search words all search words

Search powered by [Zoom Search Engine](#)

As always, you can customize the appearance of this category list via CSS, making it appear vertically or in any other layout. See "[How do I customize the look of my search page?](#)" and the "[CSS class listing](#)" for zoom_categories for more information.

Show category breakdown in search results ("Refine your search by ...")

This enables the displaying of a summary of the categories which your search results belong to. It allows your end users to see how many results belong to each category, and provides links to further refine their search by that category.

Search results for: carbon in all categories

14 results found.

Refine your search by category: [News \(7\)](#) [Downloads \(2\)](#) [PDF documents \(1\)](#)

2 pages of results.

Specifying category per file

For files which have no easy way to be categorized by a pattern in the filename or URL, you can specify a category for these files on a file-by-file basis. You do this by adding a ZOOMCATEGORY meta tag within your file, like so:

```
<meta name="ZOOMCATEGORY" content="News" >
```

Specifying the name of the category in the content part of the meta tag. This above example will categorize the file in the "News" category (overriding the URL/filename matching method).

You can specify multiple ZOOMCATEGORY tags within the same file if you wish to have the page belonging to more than one category. For example:

```
<meta name="ZOOMCATEGORY" content="News" >
<meta name="ZOOMCATEGORY" content="Old website archive" >
```

Importing categories

You can click on the "Import" button to import a list of categories from a CSV or text file. The comma separated text file should be in the following format:

```
<name>, <pattern>, <description>
```

Note that since each field is to be comma separated, you will need to enclose a field in quotation marks if you wish to sp

```
News, /news/, This would be the news section
Articles, /articles_, Article webpages
"Links, resources, and other things", /links/, links to external sites
French, /fr_
```

As you can see in the above example, you can omit the description if one is not required. But a name and pattern is always required. You can specify a category as being mutually exclusive (so that the file belonging to the category can not belong to any other category) by adding a fourth field on the line, like so:

```
Sales, /sales/,my sales related files,EXCLUSIVE
```

Note that importing categories will append to the existing list of categories already loaded. Note also that categories must have a unique name, and if a category with the same name already exists, it will be ignored on import.

2.6 Custom Meta Search Fields

You can specify custom meta fields that you want Zoom to index and make searchable. This is typically useful for online shops and web sites where a database-like search criteria is necessary.

For example, you may have a real estate website with the following meta tags in a HTML file:

```
<meta name="NUMROOMS" content="1">
<meta name="PRICE" content="300000">
<meta name="SUBURB" content="Sydney">
<meta name="AGENT" content="Bob McGuinn">
```

Note that these tags must be in the <head> ... </head> section of the page for them to be valid meta tags according to HTML standards.

By configuring Zoom's Custom Meta Fields feature, you can index this information (along with the actual page content) and make this searchable as seen below.

Search this site

Enter one or more keywords to search for using the Zoom Search Engine.
Note that '*' and '?' wildcards are supported.

Search for: Results per page:

Match: any search words all search words

Min number of rooms:

Max price \$:

Suburb

- All
- Sydney
- Randwick
- Surry Hills
- Maroubra

They will also appear alongside each search result which has the meta fields specified.

To index these tags, you would create entries such as the following in the Zoom "Custom Meta Fields" configuration window:

Meta name: NUMROOMS

Type: Numeric

Show in Search Results as: Number of rooms

Search criteria name: Min. number of rooms

Search criteria method: Greater than or equal to

The screenshot shows a dialog box titled "Add/Edit Custom Meta Field". It contains the following fields and options:

- Meta name:** A text box containing "NUMROOMS". Below it is the text: "This is the name of the meta tag which will be picked up."
- Type:** A dropdown menu with "Numeric" selected. Below it is the text: "Specify the type of data this field contains."
- Show in Search Results as:** A text box containing "Number of rooms". Below it is the text: "This is the label that will appear in your search results for this value"
- Search criteria name:** A text box containing "Min. number of rooms". Below it is the text: "This is the label for the search criteria to appear on the search form."
- Search criteria method:** A dropdown menu with "Greater than or equal to" selected. Below it is the text: "This specifies the way the above criteria will match by."

At the bottom of the dialog are three buttons: "OK", "Cancel", and "Help".

Data Types

Each custom meta field needs to be specified as a certain data type. This affects the way in which the value will be indexed, and also how it can be searched for by the end user.

The following data types are available:

- **Numeric**

This is a numeric integer value which you can match by: exact match, or less than/greater than comparison.

- **Text**

This is a text value suited to names, product codes, etc. The search criteria available are "Exact match" and "Partial (substring) text matching".

- **Drop-down text**

This is a pre-determined set of text values as specified in the "Dropdown values" list. The values in

your meta tags must match the ones in the list for them to be recognized. On the search page, the end user will be provided with a dropdown box to select one of these possible values to search by.

- **Multi-select**

This is similar to the Drop-down text option, however it allows for multiple selections, while Drop-down text only allows for one value to be selected and specified per page. Multi-select also allows you to specify multiple meta values on the same page. For example, a book item may belong to multiple genres, and have the following meta values:

```
<meta name="GENRE" content="Sci-fi">
<meta name="GENRE" content="Horror">
```

As with the drop-down text option, you must have these values specified in the list for them to be recognized and supported.

The multi-select box on the search form is a standard HTML multi-select box. It requires the end user to hold down the CTRL button to select multiple options.

- **Money**

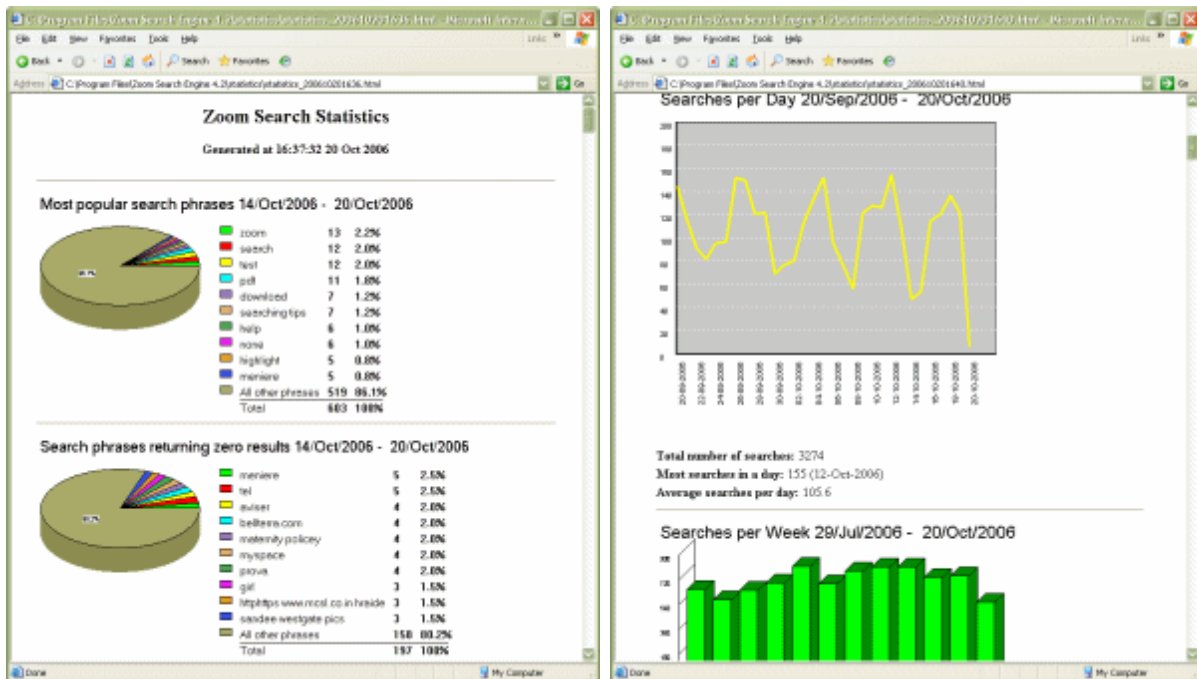
A money meta field allows you to specified a monetary value. Clicking on the "Configure" button next to the data type allows you to specify the Currency unit which will be displayed next to the value in the search results. You can also configure whether you want decimals/cents to be displayed and indexed. The meta values on your page should look something like this:

```
<meta name="PRICE" content="9.99">
```

You can customize the appearance of the meta fields via CSS. This includes the way the meta fields are displayed alongside each search result, as well as the search form from which you specify the meta field search criteria. A class listing of the CSS styles can be found in "[CSS class listing](#)".

2.7 Search Statistics Report

The statistics tool analyses the search logs generated with the Search Logging feature and produces statistic reports featuring pie charts, and graphs, offering insight into the searching patterns of your site's visitors. These reports are generated as HTML web pages and GIF images, allowing you to put them online and make your search engine statistics available to your visitors.



First, you must have enabled the Search Logging option in the Configuration window (under "Advanced"). You must have already uploaded and re-indexed your site after making this change, and your site must be successfully logging user searches to a log file.

You can then download this log file, and specify the location of this file in the search statistics window. You can also simply enter the online URL by which this log file is accessible, and it will download the file for you (eg. "<http://www.mysite.com/logs/searchwords.log>")



Note to CGI users: most Apache web servers do not allow text files to be accessed via HTTP from the cgi-bin directory. So if you store your log file in "/cgi-bin/logs/", you will not be able to access this file via an online URL. Instead, you must download the file with FTP and load the local copy. Alternatively, make sure you specify a log file path outside of the "cgi-bin" directory.

Once you have specified a valid Zoom search log file, you can select one of the many following report types to generate. Each will be a new section in the report.

- **Top 10 Search phrases**

This provides a pie chart breakdown of the top 10 search phrases made on your website. This gives

you a good indication of what people are looking for on your website.

- **Top 10 "No result" phrases**

This provides a pie chart breakdown of the top 10 search phrases on your website which found "no results". This is useful for determining what your visitors are searching for on your website but not managing to find. You can use this information to provide content that better cater your website to your visitors (or determine what meta keywords to add to allow your users to find what they are looking for).

- **Searches per day (over a 1 to 31 day period)**

This provides a bar or line chart representing the number of searches that are made on your website per day.

- **Searches per week (over a 1 to 12 day period)**

This provides a bar or line chart representing the number of searches made on your website per week.

- **Searches per month (over a 1 to 6 month period)**

This provides a bar or line chart representing the number of searches made on your website per month.

- **List the top x searched words (sorted by popularity)**

This option appends a HTML table of search words (sorted by popularity). You can specify the number of searched words you would like to list here from 10, to 50,000.

Search Statistics

Enter the location of your log file. This can be a downloaded copy on your hard disk or an online http:// URL.

http://www.mysite.com/logs/searchwords.log

Report types

- Top 10 Search phrases
Last 7 Days 20/10/2006 To 26/10/2006
- Top 10 No Result phrases
Last 7 Days 20/10/2006 To 26/10/2006
- Searches per day (over a 1-31 day period)
Date Range 1/09/2006 To 20/10/2006
 Bar Line
- Searches per week (over a 1 to 12 days)
Last 12 Week 4/08/2006
 Bar Line
- Searches per month (over a 1 to 6 months)
Last 3 Months 1/08/2006
 Bar Line
- List the top 500 searched words (sorted by popularity)

October 2006

Mon	Tue	Wed	Thu	Fri	Sat	Sun
25	26	27	28	29	30	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

Today: 26/10/2006

Generate Options Close Help

Each selection has a drop down list of options to simplify setting the date range, with preset ranges such as 7 days, 4 weeks, 3 months etc. There is also a "Date Range" option that will allow the selection of a specific date range using the date picker controls on the dialog.



Note: Your report options will only be remembered after you have successfully generated a report. If it failed whilst trying to generate the report, and you click "Close" on the "Search Statistics" window, you will lose your report options. Once Zoom has successfully generated the report, you should save it along with your indexing configuration by clicking on the "File" menu, and selecting "Save configuration".

The statistics report generated is a HTML web page with GIF images. These files are located in the "statistics" folder where Zoom is installed (unless configured otherwise, see "Options" below). You can simply upload the contents of this folder to your web server to publish your website's search statistics.

Options

By clicking on the Options button in the main Statistics window, you can specify a different folder to save the reports to.

You can also check the "Append date and time to filenames" option, which will add a number based on the date and time that the report was generated, to the filename of all report files. This ensures that new reports generated do not overwrite existing ones in the same directory.

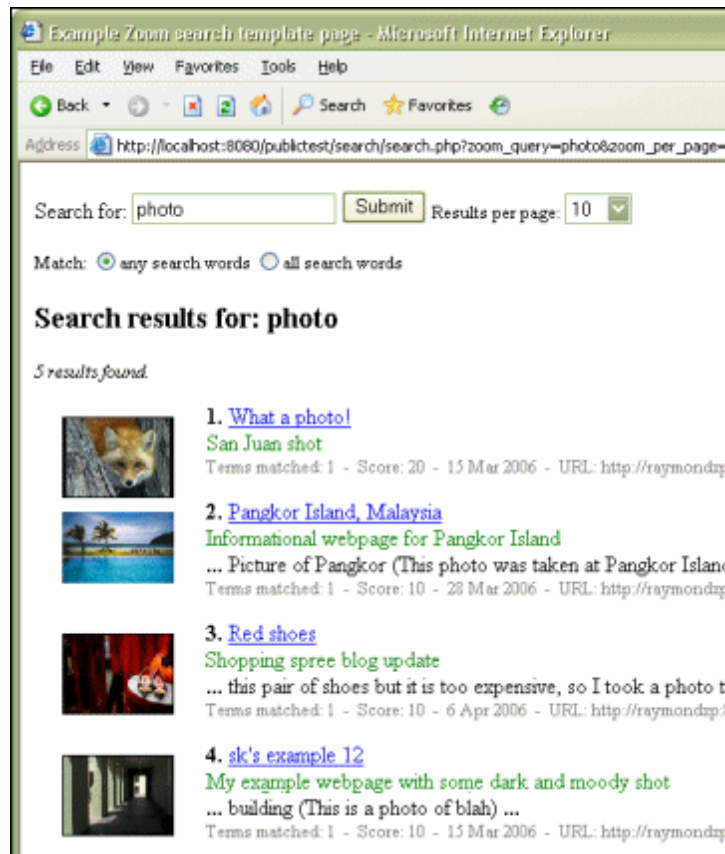
Live Statistics Reporting

If you wish to make your statistics available on your website, you will need to upload the report generated by the above tool.

However, if you have PHP support on your web server, we provide an additional script that you can use to generate reports on-the-fly, without the need to download the log file to your computer. For more information, see "[Search Statistics PHP Script](#)".

2.8 Icons and thumbnails

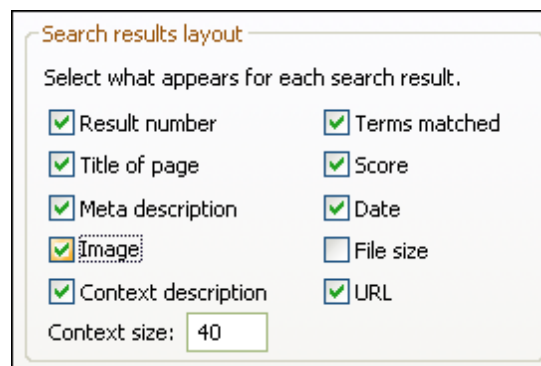
You can now associate an image with each file or web page in your index. These images will be displayed alongside the links in your search results as shown below:



Zoom provides you with a number of ways to associate images with your search results. You can also customize the appearance of these images (in width/height or position on the search results page) to a great extent using the CSS classes available.

Enable images

To allow images to appear in your search results, you will first need to enable "Image" on the "[Results Layout](#)" tab of the Configuration window.



Once this is enabled, Zoom will be able to show images alongside your search results. However, you will still need to setup the association of images with your files before any images will appear. There are 3 different ways to associate an image with a page:

- [Associating an image with a particular page](#)

- [Associating icon images with a file type](#)
- [Associating thumbnail images with files](#)

2.8.1 Associating an image with a particular page

If you wish to associate a single image with a particular page, you can use the ZOOMIMAGE meta tag to tell Zoom the filename and location of the image file to be displayed for this page. You can do this by inserting a meta tag like the following on your web page:

```
<meta name="ZOOMIMAGE" content="images/redshoes.jpg">
```

This will associate the image "redshoes.jpg" in the images folder with the page that this tag is inserted in. This option would be particularly useful for websites where you may have an individual page per product, and you would like a picture of the product to appear along their corresponding pages (or any other website where you may want a single image representing a page).



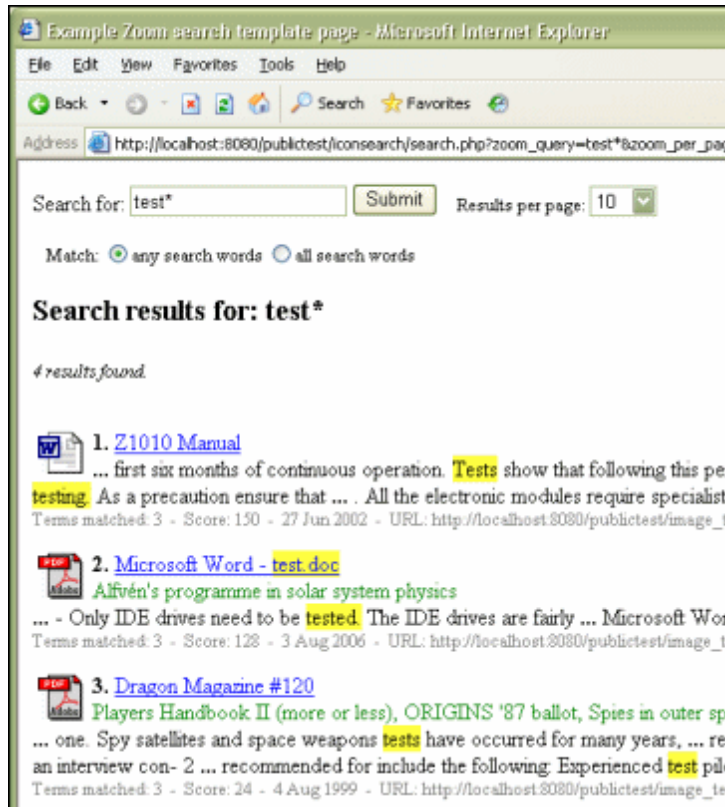
Note: The path to the image files can be relative or absolute (ie. full http:// URLs). If you are using a relative path, note that the path would be relative to the page that the meta tag was found on. Zoom will automatically resolve this relative path and determine the full absolute path for your image.

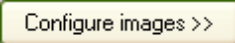
If you wish to specify a ZOOMIMAGE meta tag for PDF or other plugin supported file formats, you can do so by creating .DESC files. For more information, see "[Using custom description \(.desc\) files](#)".

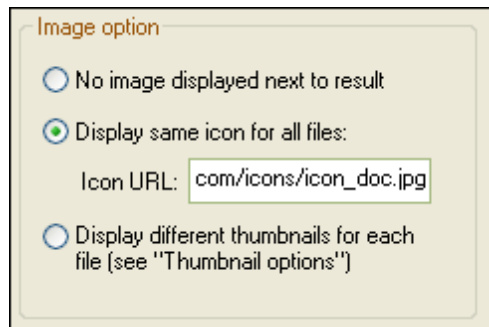
Note that the ZOOMIMAGE meta tag has the highest priority over the different image association methods, which means that you can use it to override your thumbnail or icon settings. For example, if you have all PDF files associated with a PDF icon (as described in the next section, "[Associating icon image with a file type](#)"), you can still specify a different image for one particular PDF file, by way of a ZOOMIMAGE tag and a .DESC file as described above.

2.8.2 Associating icon images with a file type


You can associate an icon image with a certain file type so that whenever these files appear in your search results, an icon representing their file type is displayed alongside them.



To do this, double-click on the file extension you wish to configure on the "Scan Options" tab of the Configuration window. The options available for that file type will appear in a window. Click on the  button. Here you will see the image and thumbnail configuration window for this file type.




Select the "Display same icon for all files" option, and specify the full URL to the image file you wish to use as an icon for this file type. The icon URL must be a full URL that points to the image file as it is hosted on your website (eg. "http://www.mysite.com/images/icons/icon_pdf.jpg"). Note that you should NOT enter a local file path here.

 **Note:** Icon file formats (eg. ".ico") can not be used here. Only image file formats which are supported by your web browser can be used (eg. ".jpg", ".gif", etc.)

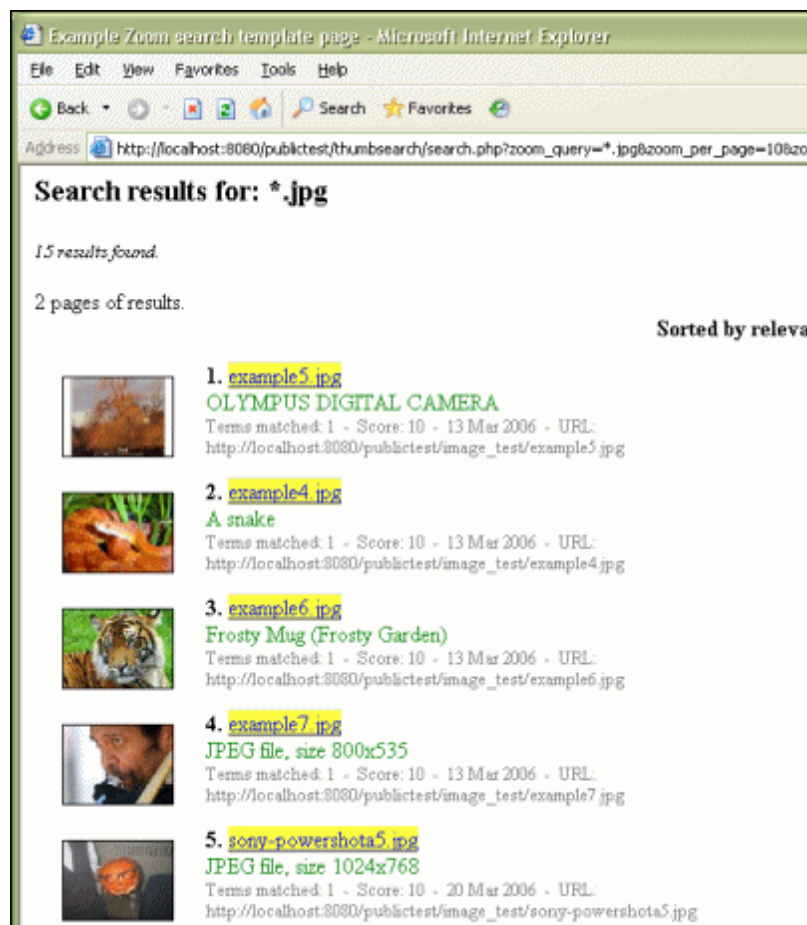
Repeat the above instructions for each file type that you wish to associate an icon/image with.

2.8.3 Associating thumbnail images with files

You can associate thumbnail images for your files so that a small preview of each file can be displayed alongside the links in your search results.

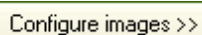
 **Note that Zoom does not generate thumbnail images for you.** You will need to have pre-generated thumbnail images hosted on your website. You can then configure Zoom as follows, telling it where to look for these thumbnails and how the files are named accordingly. There are third party applications available which can batch-process a large number of files and generate thumbnails for many image and document formats.

Using this feature, you can configure Zoom to display thumbnails for your image files (indexed using the image plugin as described in "[Image indexing](#)"), or even your web pages and PDF documents (so long as you have thumbnail images created for these files).



Example 1. Displaying thumbnails for your searchable image files.

To do this, double-click on the file extension you wish to configure on the "Scan Options" tab of the Configuration window. The options available for that file type will appear in a window. Click on the

 button.

Here you will see the image and thumbnail configuration window for this file type.

Image option

No image displayed next to result

Display same icon for all files:
Icon URL:

Display different thumbnails for each file (see "Thumbnail options")

Select the "Display different thumbnails for each file" option. You should now refer to the right side of the window which will be enabled, entitled "Thumbnail options". These options allow you to tell Zoom where to look for your thumbnail images for each URL of this specific file extension.

Thumbnail options

Associate files with thumbnails found at the following URL folder:

Look for thumbnails with the same filename as the original file but with:

before the name and/or,

after the name

Thumbnails will have the following file extension:


Based on the above settings, when Zoom indexes a file at the following URL:

It will look for its thumbnail here:

Note the example URLs given at the bottom of the window are useful in helping you make sure you have this properly configured. They will change according to your settings, and should give you a good idea of what the end result will be.


The first text box specifies the folder path for which the thumbnail images will be located. The default is "./" which refers to the current directory (of where the indexed file is located). You can specify a relative directory such as "./thumbs" which will tell Zoom that the thumbnail images will be located in a subdirectory named "thumbs", within the same folder that the indexed file is located. As always with relative paths, you can specify the previous directory with "../", etc.

You can also specify an absolute path here. For example, you could specify a path of "http://www.mysite.com/images/" which would tell Zoom that ALL of your thumbnails will be located in the same folder (regardless of where your actual indexed file is located). This method does mean however, that you will not be able to have different thumbnails for files with the same name but located in different folders.

 **Tip:** You can also specify a path starting with "/" such as "/images/", which again, will ensure that all of your thumbnails will be located in the same folder by referring to the root directory. This would be a good alternative to the absolute URL if you are hosting your files on a CD/DVD-ROM, and you need all your files in one folder.

You can also alter the filename by specifying text that precedes or follows the original filename. This would particularly suit scenarios where you may have thumbnails in the same folder of the document/files being indexed, so a file such as "myphoto.jpg" would have a corresponding thumbnail in "thumb_myphoto.jpg", etc. You can also specify the file extension of your thumbnail images here.

Again, it is important to refer to the example URLs at the bottom of the window, to check if your settings are configured to look for the thumbnails in the correct places.

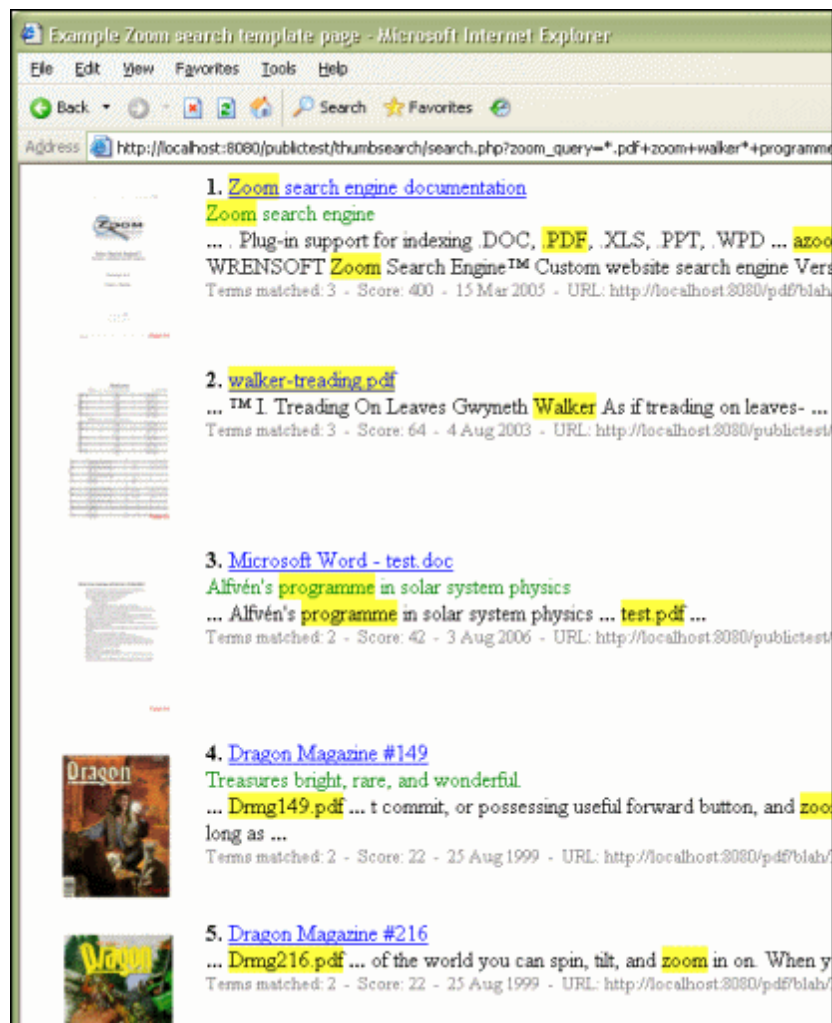
 **Note:** If file sizes are not important to you (eg. you are hosting on an Intranet or on an offline medium such as a CD-ROM), you could allow Zoom to use the original images as thumbnails of themselves. You do this by leaving the default thumbnail settings of looking in the same folder as the original image file indexed, and leaving the filename settings blank so that the filename it looks for is the same as the original. This means that it will simply use the original image (resized by the browser based on your CSS) and display them next to your search results.

This is not recommended for online websites, because it often means the page will be very slow to load as the original images may be very large. Also, web browsers do not do a particularly good job at resizing images, so these "thumbnails" may appear blocky. In most cases, we would recommend creating proper thumbnails of the original and following the instructions above to use smaller images for thumbnails rather than the original images.

Thumbnail for non-image file formats

As mentioned above, it is possible to setup thumbnails for your non-image file formats such as PDF documents, Word documents, or even your HTML web pages and video file formats. So long as you have these thumbnails generated (using a third party application) and they are hosted on your website with a filename that refers to the original file or document, then you can follow the same instructions as above to setup thumbnails for your document and other file formats.

Below is an example of a setup with thumbnails generated for PDF and DOC files.



Example 2. Displaying thumbnails for documents.

2.8.4 Customizing the appearance of your icons or thumbnails

The size and appearance of your icon or thumbnail images are defined by CSS just as the rest of your search results page appearance.

See "[How do I customize the look of my search page?](#)" for more information on how to apply CSS changes.

The CSS classes that largely control the appearance of your images is the "result_image" class.

The following are some examples of common changes that you may want to apply.

Changing the dimensions of your images

You can change the width/height of your images to be displayed by changing the "img" subclass of your "result_image" CSS definition. The following example would set your image sizes to a fixed width of 100 pixels (whilst keeping the aspect ratio by letting the browser automatically determine the height):

```
.result_image img { margin: 10px; width: 100px; border: 0px; }
```

The following would set your images to a fixed size of 100x100 pixels (which may affect aspect ratio):

```
.result_image img { margin: 10px; width: 100px; height: 100px; border: 0px; }
```



Tip: If you wish to use the original dimensions of your images or thumbnails, you can simply omit the width and height attributes in the above CSS classes.

Changing the position of your images

You can change the position or location of your images relative to your search results. By default, they are positioned to the left of your links, but you could just as easily switch them over to the right.

```
.result_image { align: right; }  
.result_image img { margin: 10px; width: 100px; border: 0px; float: right; }
```

There are many other possibilities with some clever CSS changes, so be sure to experiment!

2.9 Image indexing

Image files (JPG, GIF, PNG and TIFF) can be searched for and indexed by Zoom using our ImageInfo plugin. Note that plugin support is not available in the Free Edition.

With the image plugin installed and enabled, you will be able to search for image files by:

- Meta information such as title, description, author, etc. which is stored within the image files.
- Technical data stored in the image file (eg. camera make or model, shutter speed, flash setting, etc.) - see "[Searching for technical information](#)"
- Filename of the image files (provided you have Filename enabled in "[Indexing options](#)")
- Text that is used in any direct links to the image file
- ALT text used to describe the image on the web page (provided you have ALT text enabled in "[Indexing options](#)")

Enabling image search

To enable image search, you will need to install the image plugin and add the necessary file extensions to your "[Scan Options](#)" in the Configuration window. For more information, please refer to "[Installing a plugin](#)" and "[Using the plugins](#)".

Enabling thumbnails

You can also setup thumbnail previews of each image to be displayed alongside each search result. See "[Icons and thumbnails](#)" for more information on how to achieve this.

Improving image search accuracy

As mentioned above, the image searching capability in Zoom depends on the accuracy of the text used to describe the image, and the meta information which has been assigned to the image.

To improve your image searching accuracy, you should make sure that you have meaningful text in the abovementioned places. For example, the following would be a poor use of the ALT text to describe the

image:

```

```

This would mean the image would only be found by searches for the word "screenshot" which is hardly helpful or representative. A more meaningful use of ALT text would be:

```

```

This would allow the image to be found by searches of the word "pangkor", "beach" and "malaysia".

Similarly, if you have text links to your images like so:

```
<a href="pangkor_beach.jpg">This</a> is a picture of pangkor beach.
```

Zoom would only associate the word "this" with the image, which is again, less than useful. A more meaningful link would be like so:

```
<a href="pangkor_beach.jpg">A picture of pangkor beach</a>
```

But of course, the best way to ensure good image searching is to populate the meta information within your image files with correct and meaningful information.

Editing meta information for my images

There are several ways you can edit the meta information in your image files. The easiest way is to do so from Windows Explorer, by right-clicking on the image and selecting "Properties". Here, under the "Summary" tab, you can specify the "Title", "Subject", "Author", and add "Keywords" and "Comments".

There are also third party applications which allow you to specify more information (or add meta data to a number of files in one go). Note that with these other applications, there are additional fields available, but Zoom will only index the abovementioned fields.

For more information and examples on editing meta information, please see our FAQ page here: http://www.wrensoft.com/zoom/support/faq_plugins_image.html

Skipping small image files

By enabling image indexing, you may end up indexing a large number of tiny images used for positioning or as layout tools (eg. little arrow images, text images, boxes and backgrounds). To avoid this, you can configure Zoom to only index images which are larger than a certain file size, effectively skipping small images which are unlikely to be meaningful.

To do this, double-click on the image file extension (on the "[Scan options](#)" tab of the Configuration window) and change the value in the setting labelled "Only index images larger than x kilobytes". The default value is 5 kilobytes.

Alternatively, you can skip image files by filename or path via the "[Skip options](#)" tab of the Configuration window, as you would to skip any other files from indexing.

Searching for technical information

Some image file formats contain various technical information regarding the image's dimensions, camera type, etc. which you can search by. The following is a table listing the various technical data available for each of the image formats supported.

	Image file format				Example
	JPEG	PNG	TIFF	GIF	
Meta information					
Author	Y	Y	Y	N	John Smith
Description	Y	Y	Y	N	Photo of a beach
Title	Y	Y	Y	N	Pangkor Island, Malaysia
Keywords/Comments	Y	Y	Y	N	beach, sun, sand, vacation
Technical information					
Width	Y	Y	Y	Y	800
Height	Y	Y	Y	Y	600
Make	Y	Y	Y	N	SONY
Model	Y	N	Y	N	Cybershot
ExposureProgram	Y	N	Y	N	manual control
FNumber	Y	N	Y	N	F2.8
Aperture	Y	N	Y	N	F2.8
ShutterSpeed	Y	N	Y	N	2.5
ExposureTime	Y	N	Y	N	1/100
Flash	Y	N	Y	N	fired
DateTaken	Y	Y	Y	N	2005-11-07 13-48-01

These fields of information are indexed in the format of <type>:<value>

In order to search for a specific field, you should enable Colon character as a word join character in the "[Indexing Options](#)" tab of the configuration window. This would allow you to search for words containing a colon character as required for these fields.

You can then search for specific fields on your search page, for example, to search for images with a width of 500 and a shutter speed of 2.5, you would enter the following into your search box:

```
width:500 shutterSpeed:2.5
```

You can combine these searches with your normal search terms, as well as using wildcard characters ("*" and "?") to

2.10 Incremental indexing

Zoom allows you to perform incremental indexing so that you can update, add pages, and make changes to an existing set of index files without having to perform a full re-index. This can be particularly useful for large sites (or an index of a large number of sites) where there may be minor incremental changes that require updates to the index, but would be too time consuming (or bandwidth consuming) to re-index the entire site using Spider mode on a regular basis.

Requirements

1. Incremental indexing is only available for the PHP, ASP, ASP.NET, and CGI platforms. It is not available for the Javascript version since it is incapable of indexing a large enough set of files where incremental indexing would be beneficial.
2. In order to use incremental indexing, you must NOT have modified your indexing configuration since

the last index was made. The ZCFG file must contain the exact same settings, and the index files must still be in the output folder specified.

3. "[Updating an existing index](#)" is only available in [Spider Mode](#). It is not available in [Offline Mode](#). Offline mode indexing does not use any Internet traffic and it is many times quicker than spider mode, so this should not be necessary and we would recommend a full re-index for offline mode users.

Incremental indexing options

The following options are available from the "Index" menu, by selecting "Incremental indexing":

[Updating an existing index](#)

[Add start points to existing index](#)

[Add list of new or updated pages to existing index](#)

[View or delete pages from existing index](#)

[Command-line parameters for incremental indexing](#)

2.10.1 Updating an existing index

This option will look through the list of pages found in your existing index and check if they have since been modified. It will then perform a partial index of only the pages that have changed (and potentially index any new pages that you have added links to).



Note: With each subsequent update, the index gets larger and less efficient. We recommend performing a full re-index regularly where possible (perhaps once a week, or once a month, depending on how often you perform a partial index).

It is also important to note that the ability for Zoom to determine whether a file was modified or not is dependent entirely on the last-modified date retrieved and the file size. If these attributes are inaccurate or do not represent the changes to the file, then it will not be able to accurately find the files which have been changed.

If you are unable to perform an incremental update, please see "[Incremental indexing requirements](#)".

2.10.2 Add start points to existing index

This option allows you to add and index a list of start points (usually a new website, or a part of a new website) to an existing index. This can be useful if you manage a list of websites as start points and you wish to add new start points to the index on a regular basis.

It will index the new start point, append this data to the existing index (without having to re-index the existing start points). After a successful incremental index, it will also automatically save the configuration with your newly added start points (so that on your next full re-index, the new start points will be included).

If you are unable to perform an incremental update, please see "[Incremental indexing requirements](#)".

2.10.3 Add list of new or updated pages

This feature allows you to specify a list of new pages which are to be indexed and added to the existing index.

If you specify a page here which already exists in the index, Zoom will assume that this page has been updated/modified (without checking its file size or last-modified date), and it will remove the old data for this page, and add the new one.



Note: For each use of this function requiring a page update, the index gets larger and less efficient. We recommend performing a full re-index regularly where possible (perhaps once a week, or once a month, depending on how often you perform a partial index). We also recommend using this function with a list of pages that need to be updated in one go, rather than attempting to add pages to be updated one at a time.

Pages in the list which are new (ie. not in the existing index) and only needs to be added will not impair the index's performance and compression. Only pages which need to be updated will affect the efficiency and performance of the index.

If you are unable to perform an incremental update, please see "[Incremental indexing requirements](#)".

Base URL (Offline mode only)

If you are using this feature in Offline mode, you will need to specify a base URL for the pages to be indexed. This will have the same effect as the base URL you normally need to specify for the start directory (see "[Base URL](#)" for more details). Note that all of the pages specified in the list will have the same base URL. If you wish to add pages which will have different base URLs, you will need to add them separately (by repeating this step with the list of pages for the other base URL).

2.10.4 View or delete pages from existing index

This option allows you to browse the list of pages which exist in your current index. It also allows you to mark certain pages for deletion - removing them from the searchable content.



Note: Deleting pages using this function will NOT decrease the size of your index files nor decrease the total number of files indexed (as reported by the "[Indexing status](#)"). This means that if you have reached the [Limits](#) specified for your current configuration, you will not be able to use this feature to remove some pages and index some more. This feature will only remove the pages from the searchable content so that they will not be found when performing searches on the search page.

As with the update incremental features, it is recommended to perform a full re-index when possible, to obtain a more tightly compressed and efficient set of index files.

Note that the list of pages shown here is limited to 1,000 URLs only. In order to locate a URL in an index which contains more than 1,000 URLs, you need to enter part of the URL filename or path in the text box, and click on the "Filter" button. This will allow you to narrow down the number of URLs displayed so that you can find the URL you're looking for within the list.

If you are unable to perform an incremental update, please see "[Incremental indexing requirements](#)".

2.10.5 Command-line parameters for incremental indexing

We've added a list of command-line parameters to Zoom that will allow you to call upon the above incremental indexing features via the command-line. This will allow developers to call Zoom to perform these operations via external scripts or applications (eg. you could have a server-side script which calls upon Zoom to add a new start point to an existing index when a user submits them via a webpage).

See "[Command-line parameters](#)" for more information.

2.11 XML/RSS output (CGI only)

This option is only available for the CGI platform.

Zoom can now return search results in an OpenSearch compatible XML/RSS format. This allows for a number of possibilities including:

- The ability for developers to post-process or reformat the search results using their own scripts or applications.
- The possibility of providing RSS feeds for end-users to monitor the results for a subject they are interested in.
- The ability to integrate your search engine with the increasing number of OpenSearch platforms and solutions out there. This includes the ability to submit your search engine to an aggregator such as a9.com, or as a search provider to IE7/Firefox browsers. See "OpenSearch compatibility".



Note: The XML/RSS output does not contain a search form so you will need to have a search form created on another page or script which passes the query to the CGI as HTTP GET parameters. The XML output is not designed to be an end-user interface. It was designed to be made available for parsing by scripting languages, 3rd party software and aggregators.

You can enable this as default behaviour by selecting the checkbox Use XML/RSS output on the "[Advanced](#)" tab of the Configuration window.

Alternatively, you can switch between HTML and XML/RSS output via a URL parameter of "zoom_xml".

So the following URL would display search results for "dog" in HTML:

```
search.cgi?zoom_query=dog&zoom_xml=0
```

And the following URL would display search results for "dog" in XML/RSS:

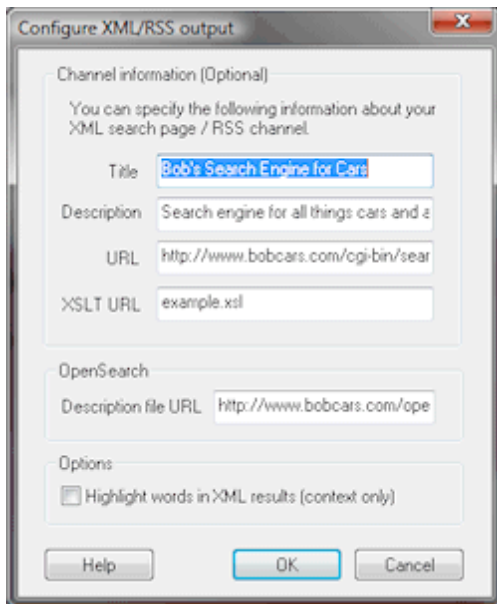
```
search.cgi?zoom_query=dog&zoom_xml=1
```

OpenSearch compatibility

With Zoom's XML/RSS output format being compatible with OpenSearch, you will be able to integrate your search engine with a number of OpenSearch platforms. This includes the ability to submit your search engine to an aggregator such as a9.com, or as a search provider to IE7/Firefox browsers. For more information on OpenSearch, please visit the websites: <http://opensearch.a9.com/> and <http://www.opensearch.org/>

To configure your XML output for use with OpenSearch aggregators, you should specify channel information and an OpenSearch description file. To do this, click on the "[Advanced](#)" tab of the

Configuration window, and click on the "Configure" button under the "XML/RSS output" section. This will bring up the XML/RSS output options.



Here you can specify the **Title**, **Description**, and **URL** of your search engine. This is what the OpenSearch aggregator will use to identify your search engine.

You can also specify the URL to your **OpenSearch description file** here. This is a XML file which you host on your website, containing information on how the aggregator can send queries to your search engine.

The **XSLT URL** allows you to specify a XSL (Extensible Stylesheet Language) Transformation file to transform the XML output to another required XML format or document. You can find out more about this here:

<http://www.w3.org/TR/xslt>

For more information on the official OpenSearch specifications for details on the description file and Query Syntax:

http://www.opensearch.org/Specifications/OpenSearch/1.1#OpenSearch_description_document

The following is an example of a description file for a CGI search engine created with Zoom:

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/-/spec/opensearch/1.1/">
  <ShortName>Bob's Search Engine for Cars</ShortName>
  <Description>Search engine for all things cars and automobile</Description>
  <Contact>bob@bobcars.com</Contact>
  <Url type="application/rss+xml"
    template="http://www.bobcars.com/cgi-bin/search.cgi?zoom_query={searchTerms}
    &zoom_page={startPage?}&zoom_per_page={count?}&zoom_and={searchAll?}
    &zoom_cat&#91;&#93;={category1?}&zoom_cat&#91;&#93;={category2?}" />
</OpenSearchDescription>
```

Highlighting within XML results (context only)

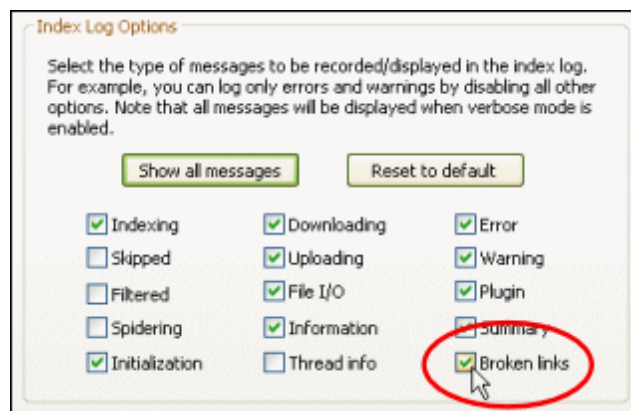
This option enables highlighting within the context description returned in the XML results. This is disabled by default to return a more natural XML response (so that <zoom:context> contains text rather

than XML). By enabling this option, highlighting will be indicated by <zoom:highlight> tags within the <zoom:context> tag content.

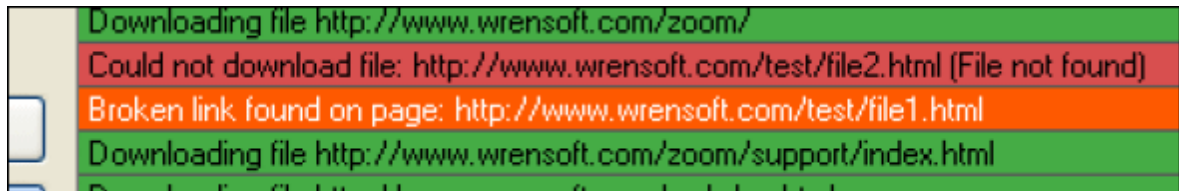
2.12 Broken link detection

When indexing your website with Zoom (using "[Spider Mode](#)"), it can also act as an effective broken link detector. Zoom's spider mode will follow every link on a page and try to download each page and index them. Any broken links (links to missing pages, or links with typos that do not lead to a valid page, resulting in a 404 page not found error) will be found by Zoom during this process.

To help you determine the actual page containing the broken link (as opposed to simply reporting the invalid URL which returns a 404 error), you can enable Broken links on the "[Index Log](#)" tab of the Configuration window.



With this option enabled, broken links will be reported during the indexing process, as seen below:



2.13 Configuration (ZCFG) files

The settings of the indexer, including the starting directory, URL, output directory, and configuration settings can be saved as a configuration file. This provides you with the convenience of simply loading the configuration file the next time you wish to re-index your website. By default, the application loads the 'zoom.zcfg' configuration file in its current directory, but this can be changed by the optional command line parameters mentioned in the following section.

If you develop multiple websites, you can also keep a separate configuration file for each website and load up the one you wish to re-index. Simply double clicking the zoom configuration file from Explorer will automatically start up Zoom with the file loaded, but if you wish to automate the indexing process, refer to the following command line parameters available.

2.14 Command line parameters

The Zoom Indexer accepts command line parameters if you wish to use it with an automated scheduling program or just to create a convenient shortcut in Windows. The usage definition is:

```
ZoomIndexer [auto-run options] <config-file> [incremental options]
```

Auto-run options

- s Auto-run in *spider mode*, start indexing as soon as the application loads. This option requires that a config-file be specified also.
- o Auto-run in *offline mode*, and start indexing as soon as the application loads. This option requires that a config-file be specified also.
- r Auto-run *Statistics Report Generator*. This will automatically create the search statistics page based on the report options saved in the config file specified.
- c Console mode (Enterprise Edition only). When using this mode, Zoom will run without a GUI, and output all log messages to stdout. See "[Integrating Zoom in your own applications](#)" for more information.

<config-file> The configuration file to load on start up

Auto-run options are mutually exclusive and only one can be selected at any one time. For example, the following is not valid: "ZoomIndexer -s -o zoom.zcfg". If you need to schedule an indexing session and also generate a report, then you will have to index two separate tasks.

Incremental indexing options

The following command-line options allow you to call upon the "[Incremental indexing](#)" features along with the auto-run options. This allows developers to call Zoom to perform these operations via external scripts or applications (eg. you could have a server-side script which calls Zoom to add a new start point to an existing index when a user submits them via a web page).

- update** This will perform an incremental update (as described above) on the specified ZCFG file. You must also specify the index mode (offline or spider) and the config file like so:

```
ZoomIndexer.exe -s zoom.zcfg -update
```

- addpage** This will add a specific page to the existing index specified by the config file and index mode. eg.

```
ZoomIndexer.exe -s zoom.zcfg -addpage http://www.mywebsite.com/newpage.html
```

Note that if you are using offline mode, you will need to specify a base URL following the addpage URL with a semi-colon (";") character, eg.

```
ZoomIndexer.exe -o zoom.zcfg -addpage C:\mywebsite\newpage.html;http://mywebsite.com/
```

- addpages** This is the same as -addpage but allows you to specify a text file containing a list of new pages (rather than calling it for one page only). eg.

```
ZoomIndexer.exe -s zoom.zcfg -addpages newpages.txt
```

Note that if you are using offline mode, you will need to specify a base URL following the filename with a semi-colon (";") character, eg.

```
ZoomIndexer.exe -o zoom.zcfg -addpages newpages.txt;http://mywebsite.com/
```

-addstartpt This option will perform an incremental add start point operation on the specified config file and index mode. eg.

```
ZoomIndexer.exe -s zoom.zcfg -addstartpt http://www.mynewsite.com/
```

Offline mode will expect a base URL following the start directory, separated by a semi-colon (";") character, eg.

```
ZoomIndexer.exe -o zoom.zcfg -addstartpt C:\mynewwebsite\;http://www.mynewwebsite.com/
```

-addstartpts This is the same as -addstartpt but allows you to specify a text file containing a list of start points.

In spider mode, the format of this text file is the same as the "Import start points" feature, which allows you to specify spidering options such as "index and follow" or "index only", etc. As well as allowing you to specify a Limit of the number of pages to index for each start point. See the chapter on "[Importing and exporting additional URLs](#)" in the Users Guide for more information.

-deletepage This parameter will delete the specified page from the index as configured by the ZCFG file given and the index mode specified. eg.

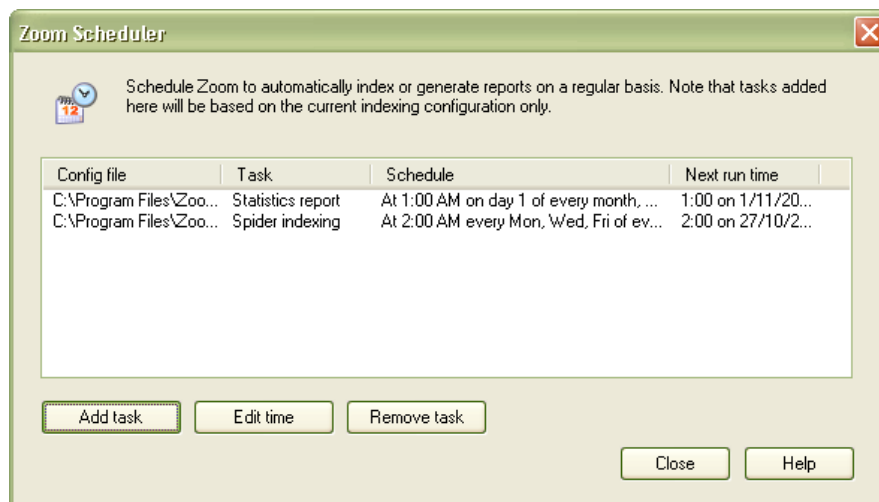
```
ZoomIndexer.exe -s zoom.zcfg -deletepage http://www.mywebsite.com/oldnews.html
```

2.15 Scheduling and automatic indexing

You can setup Zoom to automatically index your site, upload the files, or generate statistical reports on a regular basis with the built-in Zoom Scheduler.

You will first need to make sure that your current indexing configurations are correct, and that you have these settings saved in a (.zcfg) configuration file. Make sure that the saved settings in this file will successfully index your site to your liking.

If you have checked that your saved configuration is suitable for scheduling to run unsupervised, you can then click on the scheduler button on the toolbar, or from the "Tools" menu, and select "Schedule automatic indexing".



In the Zoom Scheduler window, you can add, modify, and remove scheduled tasks for various configuration files. Note that you can only schedule a task for the currently selected configuration file, but you can modify or remove existing scheduled tasks for other configuration files.

Add task

Click on "Add", and select the task option of either running in Spider mode, Offline mode, or to generate reports. You should schedule separate tasks to index and generate report, they can not be achieved with one single scheduled task.

Note that you MUST specify a username and password for the user account that will be running the scheduled task. On Windows XP, you can leave this empty, and it will use the currently logged in user. However, on most other versions of Windows, you will require both Username and Passwords to be filled in – and you must use an account that has a password specified. Otherwise, Windows will refuse to run the scheduled task.

You can then click on the "Set schedule" button and create a schedule for the task (by clicking on "New"). A large variety of schedule combinations can be configured here, with options for Daily, Weekly, Monthly, etc. You can also click on the Advanced button for even more flexible scheduling options.

Edit time

To modify the schedule for a task, simply select the task from the list and click on "Change schedule".

Remove task

To stop or remove a scheduled task, select the task from the list and click on "Remove task".

Troubleshooting scheduled task

If your scheduled task does not appear to be running correctly, check the Windows Scheduled Tasks window for more information and an error log (from the Control Panel, click on "Scheduled Tasks"). If Zoom did run, but the configuration is causing issues with indexing, you should analyse a log of what happened in its last scheduled execution. To do so, you must make sure you have enabled logging to file from the "Index log" tab of the Configuration window, for the ZCFG file scheduled (see "[Index log](#)" for more information).

2.16 Using the 64-bit Indexer

The 64-bit Indexer is only available in the Enterprise Edition.

You can find the 64-bit version of the Zoom Indexer under the Start menu, in Zoom's Program folder.



Note: In most cases besides those listed below, you should run the normal 32-bit Indexer on your computer. Note that 32-bit software are fully supported on 64-bit operating systems and you do not have to use the 64-bit version just because you have 64-bit hardware or a 64-bit OS.

You should only need to use the 64-bit Indexer if you meet one of the following conditions:

1. You are indexing an enormous number of web pages or websites, and your index files are exceeding 2GB in size. The normal 32-bit Zoom Indexer will warn you when this happens.
2. You are indexing an enormous number of web pages or websites, and you require more than 4GB of memory **and** you have more than 4GB of memory installed on your computer.

Note that there is no advantage to using the 64-bit Indexer unless one of the above criteria are met.

System Requirements for Indexing

- You must have a 64-bit CPU (e.g. Intel Core Duo)
- You must be running a 64-bit Windows Operating System (e.g. Windows Vista 64-bit)
- If your [Limits](#) require more than 4GB of memory, then you will naturally need more than 4GB of RAM installed on your computer.

Server Requirements for the CGI

You do not need a 64-bit web server to host the CGI and index files generated by the 64-bit version of Zoom. These files will run on a 32-bit web server.

However, if your web server is running older versions of Linux and BSD, you may find that it does not offer Large File Support (LFS). This means that it can not open or access any single file which is larger than 2GB in size. If you are indexing up towards a million pages or content which feature over a million unique words, then there is a good chance of hitting this limit. The CGI itself is designed to work with files larger than 2GB in size on any version of Linux and BSD which provide LFS. If you are not sure, check with your web host what version of Linux or BSD they have installed on your server and whether it features large file support. Upgrading the server to a newer version of Linux or BSD which feature LFS will address this problem.

2.17 Plugins (Indexing PDF, DOC and other document formats)

Searching and indexing non-HTML (text) files such as Word documents and PDF files require the use of external plugins available (as a free download) only for the Standard and Professional editions of Zoom. You can download the plugins from our website here:

<http://www.wrensoft.com/zoom/plugins.html>

See also:

[File formats supported](#)
[Installing a plugin](#)
[Using the plugins](#)
[Configuring a plugin](#)
[Upgrading a plugin](#)
[Using custom description \(.desc\) files](#)

2.17.1 File formats supported

We currently provide plugin support to allow for indexing and searching of the following file formats:

Extension	File type
.PDF	Acrobat document
.DOC .DOT	Word document
.PPT	PowerPoint presentations
.XLS	Excel spreadsheets
.WPD	WordPerfect document
.SWF	Shockwave Flash animation
.RTF	Rich Text Format document
.DJVU	DjVu documents
.JPG .JPEG .GIF .PNG .TIFF .TIF	Image files (see " Image indexing ")
.MP3	MP3 audio file
.DWF	AutoCAD DWF file
.ZIP	ZIP compressed archive (See " ZIP file support ")

Please visit our Plugins website for download links and a more up to date list of supported file formats:
<http://www.wrensoft.com/zoom/plugins.html>

2.17.2 Unsupported binary formats

There are some binary file formats which are recognized by Zoom but there is no plugin available to extract content or meta data. For example, ".exe" and ".zip" files do not contain much meaningful searchable meta data, so a plugin would not serve much purpose. Other file formats may contain meta data but we simply do not have a plugin available to extract this at this point.

For the file extensions in the following list, you can still add them on the "[Scan options](#)" tab, and provided you have "Filename" enabled for indexing (see "[Indexing options](#)"), Zoom will index these files so that you can search for them by filename (eg. a search for "*.exe" will return all executable files on your website).

You will also have the option to use description (.desc) files for these file types, so that you can specify titles, meta descriptions and meta keywords to these files on a file-by-file basis. See "[Using custom description \(.desc\) files](#)" for more information.

Extension	File type
.EXE	Executable file

.WMV	Windows Media Video
.MPG	MPEG video
.AVI	AVI video
.TAR	Compressed archive
.GZ	
.CAB	
.RAR	
.DGN	Vector image format

2.17.3 Installing a plugin

Download the plugin required from the Zoom website.

To install an individual plugin, unzip the file downloaded to the "plugins" folder found within the Zoom Program Data directory (see "[Where is the Program Data directory?](#)").

Once you have done this, you will need to restart Zoom Indexer. When you start Zoom Indexer, click on the "Log" tab where you should see a message in the window that indicates the plugin has been found and the file extension is now enabled. You can now add the file extension to your "Scan Options" panel on the "Configure" tab and index this file format.

If there is no indication that the plugin was found on the "Log" tab of the Indexer upon startup of the application, then your plugins were installed in the wrong folder. Double check the above instructions and try again.

2.17.4 Using the plugins

Once the plugin is installed, all you have to do is open the Configure dialog window, and add the file extension supported to the extension list. For example, ".doc" for Word document support. Click OK and start indexing.

2.17.5 Configuring a plugin

You can configure the behaviour of a plugin by double-clicking on the file extension (or selecting the extension and clicking the "Configure" button) in the "[Scan options](#)" tab of the Configuration window.

Each plugin may vary as to what options are available. The following are a list of options that are available for some plugins:

Use description (.desc) files

Enabling this option allows Zoom to look for custom description (.desc) files for this file extension/plugin. See "[Using custom description \(.desc\) files](#)" for more information.

Image and thumbnail options

Clicking on "Configure images" allow you to setup icons or thumbnails for this file extension/plugin type. See "[Icons and thumbnails](#)" for more information.

Retrieve internal meta information

This option extracts Meta information (where available) from the plugin supported files. For example, this

may include Author, Subject, and Keywords from PDF files.

Retrieve technical data when available

This option will extract technical data (when available) from this file type. For example, this may include camera brand, level of exposure, and other data which may be available within an image file format.

Scan Method (PDF only)

This option allows you to utilize alternative methods of extracting the text content from PDF files. Due to the technical limitations of the PDF file format, the textual content stored within a PDF file can be ambiguous in its order of presentation. For example, text may be split up in several columns, but this may not be defined within the PDF file itself as to when a sentence ends and when it wraps around. It is only structured visually.

For some PDF files (it depends on how they were created), the default scan method ("presentation layout") may not be the best at preserving the order of text as intended, and in such situations, you should try the other two methods available: "raw formatting order", and "text layer".

Use password (PDF only)

This allows you to specify the owner password to be used for extracting content from PDF documents which have password security enabled.

Highlight and locate within PDF documents (PDF only)

This feature will allow searched words to appear highlighted within Acrobat Reader when you click on a PDF document in the list of search results. Acrobat Reader will also scroll and locate the first occurrence of the word. This feature will only work for Acrobat Reader 7.0 or later.

Note that this feature is dependent on the capabilities of the Acrobat Reader application which does not currently support exact phrase matching and substring matching. This means that it may highlight some words which were not the ones specifically found or matched by Zoom in such cases.

2.17.6 Upgrading a plugin

As new versions of the plugins become available on the website, simply copy them into the same "plugins" directory that they were installed to (see "[Installing a plugin](#)"), making sure to allow it to overwrite the old plugin files.

2.17.7 Using custom description (.desc) files

As many external binary documents do not contain useful title and description information, Zoom allows you to specify custom Meta information for any plugin supported files. The option can be enabled in the indexer configuration dialog and must be enabled for each plugin (see "[Configuring a plugin](#)"). Once it is enabled, the indexer will look for .desc files for this plugin supported file extension.

For example, if you have a file called "mydocument.doc", you can create a text file called "mydocument.doc.desc" in the same directory with the following contents:

```
<title>This is my document custom title</title>
<meta name="description" content="This is my document's custom description">
```

Zoom will then index the words found within "mydocument.doc", but use the title and description information found in "mydocument.doc.desc" – so that you will see your custom title and description in your search results.

You can also specify other meta information in .desc files, including extra keywords with the ZOOMWORDS or KEYWORDS fields (see "[Manually add words to the index](#)"); ZOOMPAGEBOOST options (see "[Weightings](#)"), as well as Last-Modified date information (see "[Specifying a last-modified date for your web pages](#)").

2.17.8 ZIP file support

ZIP file support is built into the Indexer. This means that it is different to the other file extensions listed on the "[File formats supported](#)" page, as it does not require a plugin to be installed. However, like the plugin support, it is not available in the Free Edition and you will need a registered edition of Zoom to index ZIP files.

There are two available methods of indexing ZIP files. You can configure between them by double-clicking on the ".zip" file extension on the "Scan Options" tab of the Configuration panel.

Extract and index all files inside ZIP archive

When this method is selected, ZIP files will be uncompressed and each file stored within the ZIP archive will be individually indexed. This means that if you have one ZIP file containing 20 HTML files, Zoom will index 20 files, with each of the HTML pages indexed as separate pages.

Note that while search results will treat each of these files separately, there is no universally acceptable way to directly link to a file within a ZIP. So what you will see in your search results is a list of various pages (with their matching page titles, etc.) but they may all in fact link to the same ZIP file if these pages were all extracted from one ZIP file.

Note that ZIP files found within a ZIP archive will not be indexed.

Below is an example of a ZIP file search result with the "extract and index all files inside ZIP archive" option enabled. In this example, all 3 files returned are in fact from one single ZIP file and they all link to the same result with "#<filename>" at the end of the URL.

Search results for: benchmark software in all categories

2 results found containing all search terms. 1 result found containing some search terms.

1. [PassMark PerformanceTest - PC benchmark software](#)

Benchmark the speed of your PC computer hardware, then compare the result to other machines. Inc...
... PerformanceTest Documentation Fast, easy to use, PC speed testing and benchmarking. PassMark Pe
Hardware Store Support Forums Press room About Us Home **Software** PerformanceTest PassMark Pe
22 Sep 2008 - URL: <http://mysite/backup.zip#pt.htm>

2. [Antivirus & Internet Security products Performance Benchmarking](#)

Benchmark Testing on IS and AV products
... O F T W A R E Antivirus& Internet Security Performance Benchmarking Document: Antivirus& Intern
Pty. Ltd. www.passmark.com ACN: 099 321 392 **Benchmark** Report Antivirus& Internet Security Perf
6 Jun 2008 - URL: <http://mysite/backup.zip#Antivirus-Performance-Testing-Ed2.pdf>

3. [PassMark WirelessMon - Wireless 802.11 WiFi monitoring software](#)

PC utility software to monitor 802.11 wireless networks, access points, speed, channel & signal ...
... International, Home **Software** Hardware Store Support Forums Press room About Us Home **Software**
Purchase Standard Versions: V3.0 build 1002 Latest release date: 21 July 2008 What's new in this releas
15 Aug 2008 - URL: <http://mysite/backup.zip#wirelessmonitor.htm>

Only index filename of ZIP archive

By selecting this indexing method, your ZIP files will not be uncompressed, and their contents will be ignored. Only the filename will be indexed, and any meta data provided via a .desc file will be used (if the "Use description (.desc) files" option is selected).

Part

3

**Server-side search engine
(Using PHP, ASP, or CGI)**

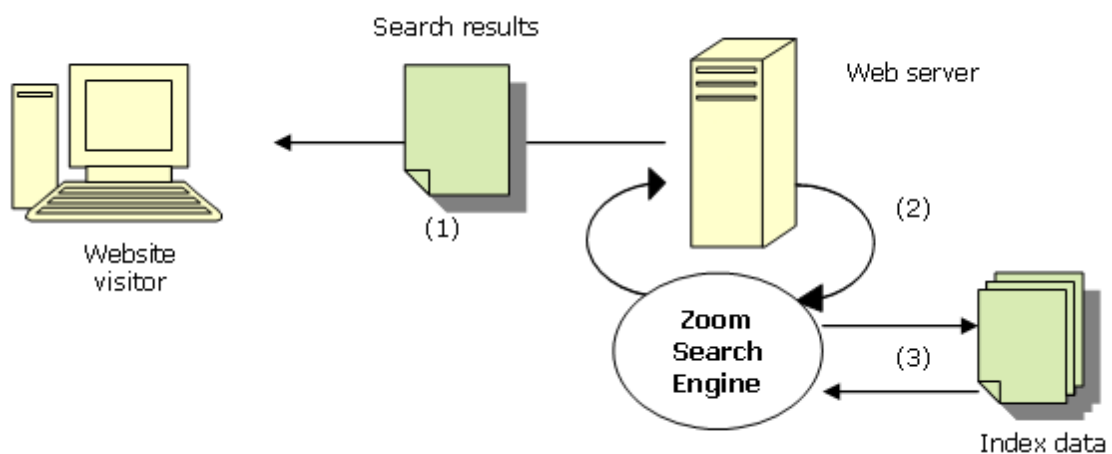
3 Server-side search engine (Using PHP, ASP, or CGI)

When to use

- When your web server supports PHP or ASP, or has CGI execute permission
- When you do not intend to run the search engine without a web server.
- When you have a large set of data (eg. over 20,000 pages)

How it works

- Requires PHP or ASP support on your web server (or CGI execute permission)
- Processes search queries on the web server
- Performance does not depend on the users' web browser or computer
- Index data is stored on the server, minimal Internet traffic.



(1) Only the search results need to be sent to the user.

(2) The search query is processed on the server.

(3) The index data files (containing the keywords, and other search data) stay on the web server and does not need to be downloaded.

See also:

[Files required](#)

[Using the CGI or PHP version without a web server](#)

3.1 Files required

The following is a list of files required to install a server-side search engine on your website. They will be created in the [Output directory](#) (that you specify before indexing) by the Zoom Indexer, after it has successfully scanned and indexed your website.

Filename	Purpose	Notes
search.php <i>OR</i> search.asp <i>OR</i> search.cgi	Search script / search page	This is the main search page. Open this file in your browser to access the search page.
settings.php <i>OR</i> settings.asp <i>OR</i> settings.zdat	Settings file	This file defines the settings for the search engine.
search_template.html	Search template	This is a template file that defines the appearance of the search page. See " How do I customize my search page? " for more information. If this file does not already exist in the output directory, a default template is created.
zoom_dictionary.zdat zoom_wordmap.zdat zoom_pageinfo.zdat zoom_pagedata.zdat	Search index data files	Data files
zoom_pagetext.zdat	Context descriptions data file	Data file This file is only created in the output directory if context descriptions are enabled.
zoom_recommended.zdat	Recommended links file (<i>Optional</i>)	Data file This file is only created in the output directory if you have specified Recommended Links.
zoom_spelling.zdat	Spelling suggestions file (<i>Optional</i>)	Data file This file is only created in the output directory if spelling suggestion is enabled.



Note: The ASP.NET option (not Classic ASP) is a special case, and requires the installation of the ASP.NET Server Control (see the next chapter for more information). The "search.aspx" file, which is the search page will not be found in the Output directory as the files above. It will be installed with the ASP.NET Server Control directly on the server.

3.2 Installing the ASP.NET native control

Note that this procedure is not necessary for the PHP, Classic ASP or CGI versions.

If you are using the ASP.NET native control for your server-side search, there is a different installation procedure required. You will need to install the ASP.NET control directly on the server itself.



Note: You will need the .NET Framework (version 2.0 or higher) and ASP.NET (version 2.0 or higher) installed on your server. The latest versions are available for download from the Microsoft website.

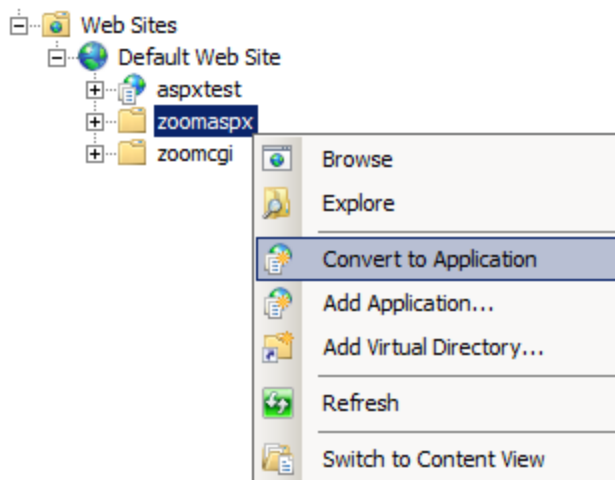
First, you will need to download the MSI installation package from our website here:
<http://www.wrensoft.com/zoom/aspdotnet.html>

You will need to run and install this package on the server where you will be hosting the ASP.NET search from.

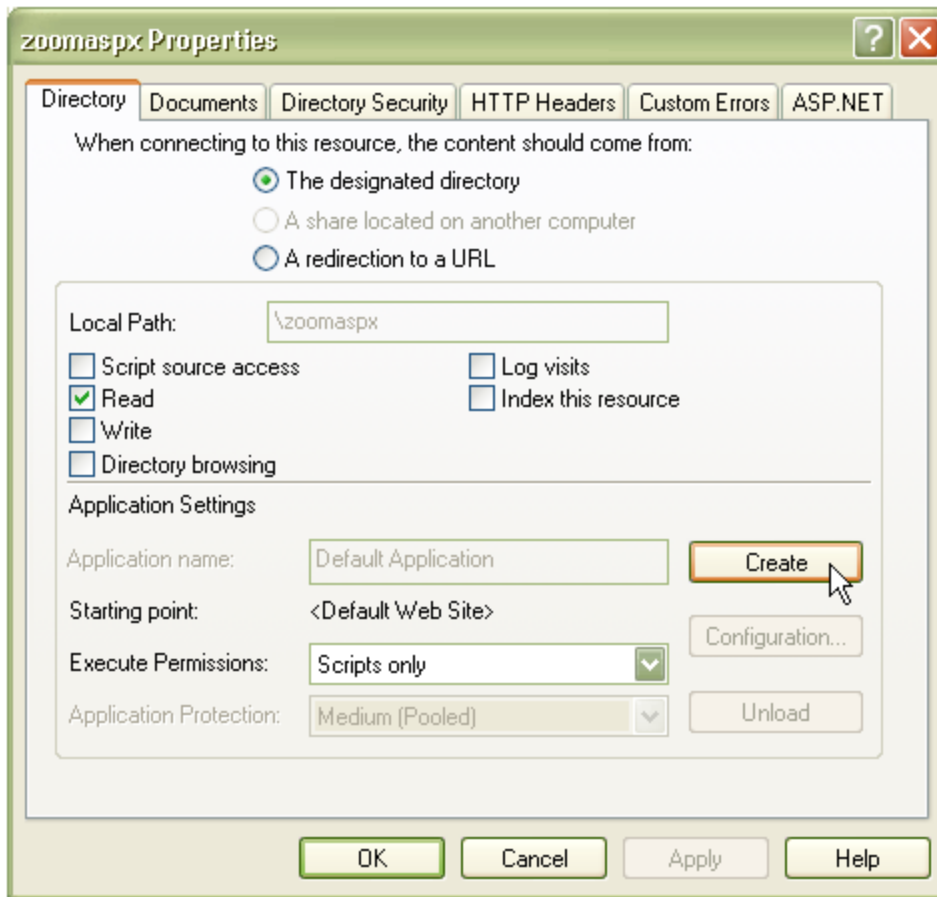
When you launch the MSI installer, you will be prompted for a path to install the control to. You should specify the folder which you wish to host the search page from.

Once you have done this, you will need to open up the IIS Manager and locate the folder in your "Web Sites" list. You will need to configure IIS to allow this folder to execute as an application. This can vary with different versions of IIS.

In IIS 7, you will need to right click and select "Convert to Application". Accept the default properties and click OK.



In IIS 5 and IIS 6, right click on the folder and select "Properties". Next, click the "Create" button and select the default options by clicking "OK".



Note: If you have multiple versions of ASP.NET installed, you may need to specify the version you wish to use for this application. You can do this from the "Properties" windows, under the "ASP.NET" tab where a drop-down box should be available for you to select the ASP.NET version to use. Please specify a version that is 2.0 or later. If you do not have the required version installed, you will need to download and install this from the Microsoft website.

At this point, you should be able to load the "search.aspx" page on your website (for example, if you have installed the control in a folder named "zoomaspx" at the root level of your website, it would be accessible from a URL such as "http://www.mywebsite.com/zoomaspx/search.aspx" in your browser). You should get a response that looks something like this:

```
Content-type: text/html Unable to open settings.zdat or file is invalid.
Check file permissions and that file exists
```

This is normal and indicates that the ASP.NET control is working, and it just needs the index files to continue.

If you have not already done so, you will now need to index your website as described in Chapter 2, "[Indexing your website](#)". This will create a set of index files and a search template which you can upload or copy over to the same folder where you have the ASP.NET control installed.

Once the index files (all ZDAT files and the "search_template.html" page) are hosted in the same folder, you should find that your search page will now be fully operational and you can customize the

appearance of your search page as described in, ["How do I customize the look of my search page?"](#).

Troubleshooting

There are currently known issues with displaying non-English characters in the search results using the ASP.NET native control.

Some .NET web page generation tools will put a HTML <form> that surrounds the entirety of the document. By default this form submits using POST, whereas this control uses GET variables. It is recommended to move this control outside of the form, but if that is not possible (i.e. the form is on the master page) then try placing the following lines of code within the ASPX code behind load function:

C#

```
this.Form.Method = "GET";
this.Form.EnableViewState = false;
this.ClientScript.RegisterClientScriptBlock(GetType(), "clr", @"document.
getElementById('__VIEWSTATE').value = '';", true);
```

Visual Basic

```
Form.Method = "GET"
Form.EnableViewState = False
Page.ClientScript.RegisterClientScriptBlock(Me.GetType(), "clr", "document.
getElementById('__VIEWSTATE').value = '';", True)
```

Part

4

**Client-side search engine
(Using JavaScript)**

4 Client-side search engine (Using JavaScript)

When to use

- When you intend to distribute the search engine on a CD-ROM or DVD, or any local Intranet where there is no running web server.
- When your web server does not support PHP or ASP, or provide CGI executable access.
- When your website is hosted on an intranet or local web server, where traffic and network latency is not an issue.

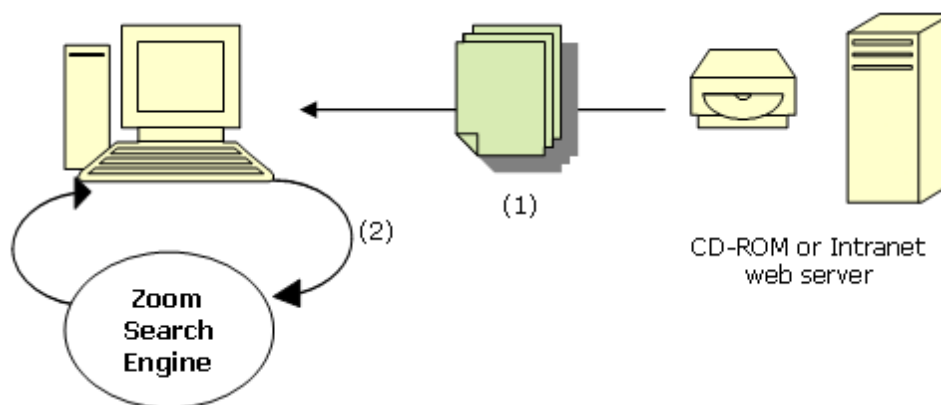
For CD-ROM or DVD searching

If you wish to put the search engine on a CD-ROM or DVD distribution, you will need to use the client-side (Javascript) search option. This is because it is usually not practical to run a web server off a CD or DVD and also since the client-side search script is more convenient in that it only requires the search files (see "[Files required](#)") and a web browser. However, note that the capabilities of Javascript are much more limited, and many features available to server-side scripting are not available (such as context descriptions and exact phrase matching).

If the capabilities and limitations of the Javascript option do not meet your requirements, you should consider running the CGI search script with a specialised web server on the CD. This would require additional software. For more information, see "[Using the CGI or PHP version without a web server](#)".

How it works

- Works regardless of web server platform. Can be used online, or to provide offline searching running off a local disk, CD-ROM, hard disk, etc.
- Search queries are processed on the user's computer.
- Performance depends on the user's computer, browser, and network connection.
- Requires the entire index data to be loaded onto the user's computer.



- (1) All index data must be transferred and loaded onto the computer of the user performing the search query. This is not an issue if we were running the search off a CD-ROM or a local Intranet server, but can be extremely slow and traffic heavy if it is hosted on an Internet web server. Note that, in this

context, we are not referring to the need for a user to click on a file and select “download”, but that the files will be automatically downloaded when a user visits the page (costing bandwidth and download time), much like how a big image on your web page can make it load slower.

(2) The search query is processed on the user’s computer.

See also:

[Files required for Javascript](#)
[Limitations of Javascript](#)

4.1 Files required

Note that all files required will be created in the Output Directory at the end of indexing.

Filename	Purpose	Notes
search.html	Search page / search template	This is the file to open in your browser to access the search page. You can also customize the appearance of the search page by modifying this file (see “How do I customize my search page?” for more information.). If this file does not already exists in the output directory, a default template is created.
search.js	Search script	This file contains the core functionality of the searching process.
settings.js	Settings file	This file defines the settings for the search engine.
zoom_index.js	Search index data file	Data file
zoom_pageinfo.js	Pages index data file	Data file

4.2 Limitations of Javascript

While Javascript has the advantage of having no server-side requirements, it does have some significant disadvantages due to the fact that it is a relatively basic, browser scripting language.

- **Capacity:** The amount of resources (both memory, and CPU) made available to the script interpreter is often very limited. As such, it can not deal with large volumes of data in memory and can only handle around 5 MB of indexed data (that is, the generated .js files can total around 5 MB in size), which may be approximately around 50 MB of site content. Your mileage may vary depending on the browser you are using.
- **Performance and functionality:** Due to the fact that the script runs inside a browser, more intricate processing methods are restricted, making it unfeasible to implement certain features (listed below).

Search features **not available** in the Javascript version are:

- Context descriptions

- Exact phrase matching
- Spelling Suggestions
- Search word logging

Note that all of the above features are available with the server-side search options available: PHP, ASP, and CGI (see "[Server-side search engine](#)"). If you wish to publish your search function without a web server, but require the features available in the server-side search options, take a look at "[Using the CGI or PHP version without a web server](#)".

4.3 Using the CGI or PHP version without a web server

While Javascript is a good solution for distributing small sites on CD it is limited in what it can do. Most notably, the Javascript version can not handle large volumes of data and struggles with sites of up to 100,000+ (this can depend on the browser used).

The CGI version can handle much greater volumes of data, and also offers additional functionalities such as exact phrase searching and "Google-like" context search results.

The solutions below apply to CD's, DVD's, USB drives and memory sticks. The process is the same in each case despite the media being different. Each of these options require the Zoom Search Engine.



Tip: You can find a solution comparison table (and more information) on our online support page for this topic here:

http://www.wrensoft.com/zoom/support/faq_CD_search.html

The available options are as follows:

Use our FlyingAnt web server on a CD with the CGI option

FlyingAnt is a product we have developed that allows you to run CGI's directly off a CD/DVD. It is the most powerful and flexible solution of the ones listed here, allowing you to create a CD with all the search capabilities available in Zoom and your CD will work for all browsers, as well as cross-platform support for your audience on Windows, Mac OSX, and most Linux distributions.

See the FlyingAnt website for more information: <http://www.wrensoft.com/flyingant/index.html>

Use the Zoom Javascript option to fully run the search in a browser

This is [default Javascript option](#) described in the rest of this chapter. It is the least capable due to the technical limitations of the JavaScript platform, but the most convenient. It requires no additional software or setup, and will run in any modern web browser.

Use third party IE plugin DLL to use the ASP scripting language on a CD.

There are third party solutions available to running ASP scripts on a CD. These include packages such as the HSP (HTML Scripting Pages). You may prefer this option if you have other ASP scripts that you wish to serve on the CD along with Zoom. For more information, visit our online support page here: <http://www.wrensoft.com/zoom/support/aspcd.html>

Use third party webserver software to use the PHP or CGI scripting language on a CD.

There are third party solutions available to running PHP scripts on a CD. These include packages such as Server2Go. You may prefer this option if you have other PHP scripts that you wish to serve on the CD along with Zoom. For more information, visit our online support page here: <http://www.wrensoft.com/zoom/support/cgicd.html>

Use our free Zoom Front End software on a CD with the CGI option.

For advanced software developers, we provide a sample C/C++ project with source code to demonstrate how to use the CGI within a C/C++ application and execute this off a CD-ROM or DVD-ROM without any web server.

The project also comes with a compiled sample Front End for Windows. You can take the binary Front End application and use this in your distribution, or modify the source code and compile your own Front End application.

Due to the technical nature of the project files, we would recommend this option only for more experienced programmers. For more information, and download links, visit the web page here: <http://www.wrensoft.com/zoom/support/cgifrontend.html>

Part


5

**Publishing your search
engine on your website or
CD-ROM**

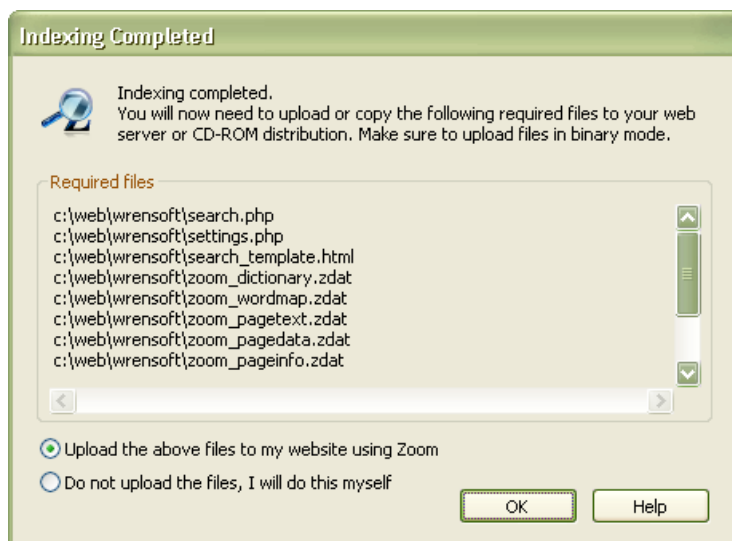
5 Publishing your search engine on your website or CD-ROM

5.1 What to do after indexing

Once you have indexed your website, you will need to upload or copy the files created by the Indexer to your web-server or CD-ROM/DVD distribution.

 **Note:** If you have selected a [server-side platform](#) (such as PHP, ASP, ASP.NET or CGI), your search files will not work until they are uploaded to a server with the appropriate scripting or application support. For example, you cannot test the PHP search page on your offline desktop computer (it will show up as script code). You need to upload or copy the files to a web server with PHP support to allow it to execute. A [client-side platform](#) (e.g. JavaScript) however, can be tested/executed on your local computer without needing to be uploaded to a web server.

The list of files required are given in the window that pops up after indexing (shown below).



These files are created in the Output directory you specified. Note that the list includes all the files you will need on your website for your search engine to run.

Here you can select from two options to continue with, by default, Zoom can upload the files to your website. If this option is selected, Zoom will open the FTP upload window when you click OK, and prompt you for FTP information.

If you would like to upload the files yourself (or you do not need to upload the files because you are indexing a site for a CD/DVD distribution or an Intranet, etc.), select "Do not upload the files" and click OK.

See also:

[Uploading to your website](#)
[CD-ROM distribution](#)

5.2 Uploading to your website

You will need to upload the files listed under "Required Files" to your web server. You can either use Zoom to upload the files (see "[Upload files to server \(FTP\)](#)"), or you can use your preferred FTP client. If you choose to use an external FTP client, make sure that the files are uploaded in binary mode and NOT text mode.

Also, make sure that all the files listed are uploaded to the same directory, and to remove or overwrite any files on your web-server from previous indexing sessions.

5.3 CD-ROM distribution

If you are putting your search engine on a CD-ROM or DVD (Javascript platform only), you should copy the [Required Files](#) listed to where your CD-ROM content is.

Note that, if you are using relative paths for your base URL, it will be important where you copy these files to make sure that the search result links will still work. For instance, if you specified a base URL of "." then your search files should be placed on the same directory level as where your CD-ROM content is. Refer to the Appendix chapter, "[Base URL for CD-ROM distribution](#)", for more information on relative and absolute paths.

5.4 Additional notes for uploading CGI

If you are using the CGI version, note that many web servers are set up to only execute CGI from a certain directory in your web account.

For Apache servers, this would usually be a directory named "cgi-bin". Check with your web host for more details, and upload your files to the appropriate directory.

For IIS servers, you may need to enable "Execute permissions" for "Scripts and Executables" on your selected hosting folder. You can do this from the IIS Control Panel, right clicking on the folder, and selecting "Properties".

Once your files are uploaded to the web server, you must also set execute permissions on the file "search.cgi", and public read permissions on the other files (all the .zdat files and the search template HTML file). This can be done via your FTP client or through a Unix shell account by using the "chmod 755 <filename>" command.



Tip: Zoom can be configured to set execute permissions for the "search.cgi" file automatically via FTP, immediately after uploading the files. This option can be found on the FTP tab of the Configuration window. See "[Set execute file permission after uploading](#)" for more information. Note that this option may not work on some servers where remote file permission changes are not allowed.

5.5 Where is my search page?

Once you have uploaded or copied the required files across to your server, you can open the following file in your browser to access the search page:

- search.php (for PHP)
- search.asp (for ASP)

- search.html (for Javascript)
- search.cgi (for CGI)

You should create links elsewhere on your website to the appropriate search page. Alternatively, you can create a search form on your website which passes the query to one of the above mentioned scripts. See "[How do I add a search form](#)" for more information.

Part

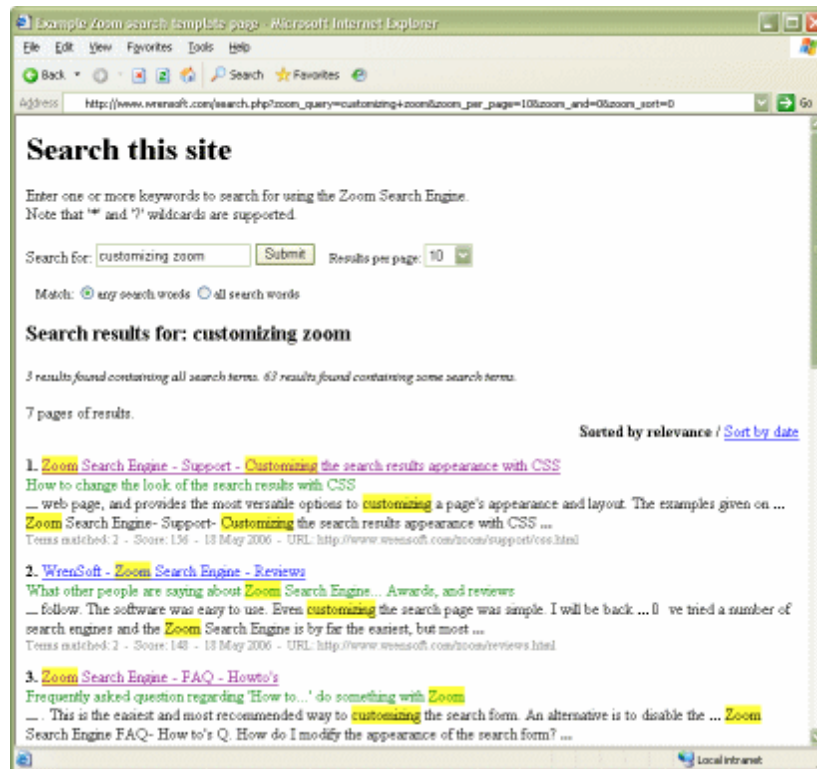
6

**How do I customize the look
of my search page?**

6 How do I customize the look of my search page?

You can completely alter the appearance of the search page, from changing the fonts and layout to including headers, footers or site navigation menus.

The following is a screenshot of how your search page will look by default after indexing and uploading (before any customisations have been made).

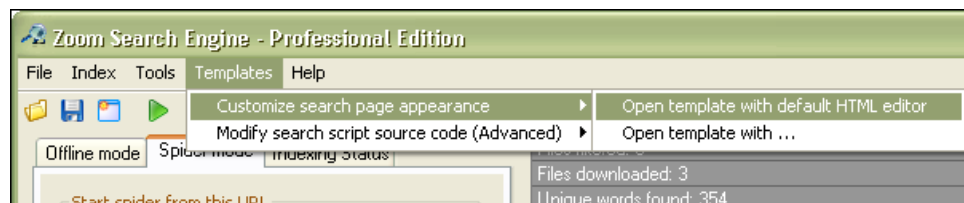


First, you will need to have indexed your site successfully. If so, a default search template file should have been created in the output directory (alongside the index files). This is a normal HTML file that can be edited like any other ordinary web page with your favourite web page editor.



Note: the search template file is named "search_template.html" for the PHP, ASP, and CGI versions. For the Javascript version, it is named "search.html".

You can also open this file from the Indexer, by clicking on the "Templates" menu, and selecting "Customize search page appearance".



This will open up the HTML search template in your default HTML editor. Alternatively you can select a different editor to modify the template with. Note that you should not use MS Word to edit the template

file due to known problems with Word's support for HTML editing* (see "[Changing your default editor for templates](#)").

You can modify this page as you see fit, however, you must retain the following "tag" in place of where you want the search function to be displayed:

```
<!--ZOOMSEARCH-->
```

This is different for the JavaScript version, which uses the following line instead:

```
<script language="JavaScript">ZoomSearch();</script>
```

If you need to go back to the default search template, you can delete or rename your search template file, and re-index your site. A new "default template" will be created in the output directory at the end of indexing, when it finds that the search template is missing.

To change the fonts, colours, and appearance of your search results (everything after the search form), edit the CSS styles defined at the top of the template HTML file. These styles are listed in "[CSS class listing](#)".

If you are not familiar with CSS (Cascading Style Sheets), you can look up more information on attributes and definitions you can use to alter everything from margins, font sizes, background images, and more. There are many tutorials and websites online (search for "CSS tutorial" on Google).

Just by changing this search template file, the appearance of your search page can be completely changed. The following screenshot is an example of a customized search page. See our website at <http://www.wrensoft.com/zoom/> for more examples of search pages in the "Example sites" section.

Advanced template options

For more advanced users, who may wish to modify the layout of search results and/or the search form outside of what is possible with CSS, they should refer to the "[Advanced template options](#)" section.

See also:

[Customizing the search form](#)
[Customizing the search results](#)
[CSS class listing](#)

6.1 Customizing the search form

The following are the class names used to define the style and appearance of the search form. (Note: that this only applies to the search form generated by the script. You can also define your own search form).

The image shows a search form with several labeled components:

- zoom_searchform**: The entire search form area.
- zoom_searchbox**: The search input field containing the text "customizing".
- zoom_button**: The "Submit" button.
- zoom_results_per_page**: The "Results per page:" dropdown menu set to "10".
- zoom_categories**: The "Category:" dropdown menu set to "All".
- zoom_match**: The "Match:" radio buttons, with "all search words" selected.

6.2 Customizing the search results

The following are the class names used to define the style and appearance of the search results.

The image shows search results for "customizing css in all categories" with the following callouts:

- searchheading**: Points to the search heading "Search results for: customizing css in all categories".
- summary**: Points to the summary text "23 results found containing all search terms."
- result_pagecount**: Points to the page count "3 pages of results."
- sorting**: Points to the sorting options "Sorted by relevance / [Sort by date](#)".
- result_title**: Points to the first result title "1. [Zoom Search Engine - Support - Customizing appearance of images and thumbnails in search results](#)".
- description**: Points to the result description "How to change the look of the search results with CSS".
- highlight**: Points to the highlighted text "... web page, and provides the most versatile options to customize a page's appearance and layout. The examples given ... Home Products Zoom Search Engine Support Using CSS How do I customize the appearance of my search page with ...".
- context**: Points to the context text "... Home Products Zoom Search Engine Support Using CSS How do I customize the appearance of my search page with ...".
- infoline**: Points to the infoline text "Terms matched: 2 - Score: 588 - 12 Dec 2006 - URL: http://www.wrensoft.com/zoom/suppo".
- result_block**: Points to the first result block.
- result_alblock**: Points to the second result block.
- category**: Points to the category text "2. [Zoom Search Engine - FAQ - appearance of images and thumbnails in search results](#) [FAQs] [Support]".

There are some common CSS examples available (such as changing the colour of the highlighting, or the

fonts used for the search results, etc.) on our FAQ page here:
<http://www.wrensoft.com/zoom/support/css.html#examples>

6.3 Customizing the recommended links

The following are the class names used to define the style and appearance of the search results.



You can specify the recommended links, titles and descriptions to appear here from the "[Recommended links](#)" tab of the Configuration window.

6.4 CSS class listing

Style class	Description
.highlight	The appearance of the highlighted search words in the results, when highlighting is enabled. For example, <code>.highlight { font-weight: bold; }</code> will make highlighted text bold (with no coloured background).
.searchheading	The appearance of the "Search results for..." heading
.summary	The appearance of the summary information for your search results ("129 results found on 2 pages")
.suggestion	The appearance of text regarding spelling suggestions ("Did you mean...") or search suggestions ("You may find more results ...")
.results	The appearance of the search result listing.
.category	The appearance of the category tag next to the result link, e.g. "[News article]"
.cat_summary	The appearance of the category summary, e.g. "Refine your search by category: News (3) Forum (13)"
.sorting	The appearance of the text links to switch between "Sort by relevance" and "Sort by date" when sorting by date is enabled.
.result_title	The appearance of the title line (result number, page title/link, and category name). Note that the link itself will be based on the styles defined for hypertext links (ie. "a:link", "a:visited", "a:hover" and "a:active").
.description	The appearance of the meta description or page description for a search result.

.context	The appearance of the contextual description. You can specify font, size, etc. here. You can also specify margins and padding to make the context indent from the main result link.
.infoline	The appearance of the small information line for each search result ("Terms matched", "Score", and "URL").
.zoom_searchform	The appearance of the overall search form. You can specify the overall font you'd like to use for the text around the form, or give the form a border, margin, or background. For example: <pre>.zoom_searchform { font-size: 100%; border: dashed; border-width: 1px; background: #DDDDDD; }</pre> Will set a dashed border around the search form with a light grey background.
.zoom_results_per_page	The appearance of the "Results per page" drop-down box.
.zoom_match	The appearance of the "Match ... any search words / all search words" option.
.zoom_categories	The appearance of the categories drop-down or checkbox list (only when categories are enabled). See " Categories " for more information. The following will set your category drop-down box to green with smaller fonts: <pre>.zoom_categories { background: #00FF00; font-size: 70%; }</pre> When multiple categories are enabled, a checkbox list is used and you can change the appearance of the list via the "ul" and "li" sub-items of this CSS class. For example, <pre>.zoom_categories ul { display: inline; margin: 0px; } .zoom_categories li { display: inline; list-style-type: none; }</pre> will set the multiple category checkbox list to be horizontal without list bullets.
input.zoom_button	The appearance of the search button. You can change the colour of this button or set it to an image. For example: <pre>input.zoom_button { background-color: #008080; color: #FFFFFF; }</pre>
input.zoom_searchbox	The appearance of the search box where a user enters their search query. You can change the background colour, size, and font from here.
.result_image	The appearance of the image to be displayed alongside your search results (see " Icons and thumbnails "). Note that this is the class wrapping around the image itself. You can specify CSS for the img file itself via ".result_image img". For example: <pre>.result_image img { margin: 10px; width: 100px; border: 0px; }</pre> will set your images to a fixed width of 100 pixels (and automatically determine the height) with no border and a 10 pixel margin. For more examples, see " Customizing the appearance of your icons or thumbnails ".
.result_block	The appearance of one row of the search result. You can use this to specify lines between each search result entry, or an alternating background corner etc. For example, <pre>.result_block { border: 1px; margin: 15px; clear: left; }</pre> will specify a 1 pixel wide border and 15 pixel margin around the search result item (enclosing the titles, description, etc.). Note that you will need to specify the same for .result_altblock if you want all of your result items to have this appearance, or you can specify something different.
.result_altblock	The appearance of every alternating row of the search result (to .result_block above). You can specify something different here to .result_block and have alternating background colour changes for

	<p>example, like the following:</p> <pre>.result_block { background-color: #FFFFFF; } .result_altblock { background-color: #AAAAAA; }</pre> <p>will specify every result item to have a white background, and every second item to have a grey background.</p>
.result_pages	The appearance of the text links to the different pages of search results at the bottom of the page ("Result Pages: 1 2 3 Next >>")
.result_pagescount	The appearance of the text at the top of the results displaying the number of pages of results found ("10 pages of results")
.searchtime	The appearance of the "Search took x seconds" text displayed at the bottom of the page (when enabled, see " Configuring search page ")
.recommended	<p>The appearance of the recommended links part of the page that appears at the top of the search results (when "Recommended links" are enabled). The following will show all recommended links in a green coloured box with a dotted border on top and bottom:</p> <pre>.recommended { background: #DFFFBF; border-top: 1px dotted #808080; border-bottom: 1px dotted #808080; margin-top: 15px; margin-bottom: 15px; }</pre>
.recommended_heading	The appearance of the heading text for 'Recommended links'.
.recommend_block	The appearance of one row/item of a recommended link (similar to <i>.result_block</i> above, but this is for the recommended links).
.recommend_title	The appearance of the title text link for a recommended link (similar to <i>.result_title</i> but for the recommended links).
.recommend_description	The appearance of the description for a recommended link (similar to <i>.result_description</i> but for the recommended links).
.recommend_image	<p>The appearance of the image to be displayed alongside a recommended link. Note that this is the class wrapping around the image itself. To make your recommended link images appear on the left side of the link, use the following:</p> <pre>.recommend_image { float: left; display: block; }</pre> <p>You can specify CSS for the img file itself via ".result_image img". For example:</p> <pre>.recommend_image img { margin: 10px; width:100px; border:</pre> <p>will set your images to a fixed width of 100 pixels (and automatically determine the height) with no border and a 10 pixel margin. For more examples, see "Customizing the appearance of your icons or thumbnails".</p>
.recommend_infoline	The appearance of the information line for a recommended link (similar to <i>.result_infoline</i> but for the recommended links).

Custom Meta Field CSS classes

The following classes are only needed for sites using [Custom Meta Search Fields](#).

.result_custommeta	The appearance of one Custom Meta Field that are listed next to a search result. This is the block that encloses the inline "result_metaname_MYFIELD" and "result_metavalue_MYFIELD" tags. Changing this allows you to specify how to space out the different
--------------------	---

	meta fields, whether they should appear on the same line or not. For example, <pre>.result_custommeta { display: block; }</pre> will make each meta field appear on a new line.
.result_metaname_MYFIELD	This is the style for the individual meta field labels that appear next to your search results. For example, if you have a custom meta field for "PRICE" which is displayed as "Price: 8" in the search results. This would be the "Price:" part of the search result. So the following, for example: <pre>.result_metaname_PRICE { font-weight: bold; }</pre> will make the word "Price:" appear in bold, while the value will remain normal (e.g. " Price: 8 ")
.result_metavalue_MYFIELD	This is the style for the individual meta field values that appear next to your search results. For example, if you have a custom meta field for "PRICE" which is displayed as "Price: 8" in the search results. This would be the "8" part of the search result. So the following, for example: <pre>.result_metavalue_PRICE { font-style italic; }</pre> will make the value "8" appear in italics.
.zoom_metaform	This is the part of the generated search form containing the meta search fields/criteria that the user can search by.
.zoom_metaform_numeric	This is the input text box for meta fields which are specified as the "Numeric" type.
.zoom_metaform_dropdown	This is the dropdown box for meta fields which are specified as the "Dropdown text" type.
.zoom_metaform_multi	This is the multi-select box for meta fields which are specified as the "Multi-select" type.
.zoom_metaform_money	This is the input text box for meta fields which are specified as the "Money" type.

6.5 How do I modify the search form on the search page?

If you wish to modify the search form, you can either edit the script, or define your own search form, with the following HTML form code:

If you wish to use your own search form, you may want to turn off the form displayed by the script from the Search Page configuration tab, under the Form Format setting. See '[Configuring Zoom](#)' for more information.

```
<form method="GET" action="search.php">
  <input type="text" name="zoom_query" size="20">
  <input type="submit" value="Search">
  Results per page:
  <select name="zoom_per_page">
    <option selected>10</option>
    <option >20</option>
    <option >100</option>
  </select>
  <br><br>
  Match:
  <input type="radio" name="zoom_and" value="0" checked>
  any search words
  <input type="radio" name="zoom_and" value="1">
  all search words
</form>
```



Note: You must change "search.php" in the above HTML to the appropriate search page: that is, "search.asp" for ASP users, "search.cgi" for CGI users, and "search.html" for Javascript users.

If you have categories enabled, and you wish to include the category drop-down boxes in your custom search form, then you will need to manually specify these categories in HTML. Because these options vary depending on your categories configuration, we can not provide the exact HTML you will require. The best method is to simply allow the search script to generate the full search form once, click on "View Source" in your browser and copy and paste the categories HTML required for your particular configuration.

6.6 How can I add a search form to my menus, main page, etc.?

If you want to add a simple search form on your website's main page, or navigation bar, etc, you can do so by specifying the following HTML on that page.

```
<form method="GET" action="search.php">
  <input type="text" name="zoom_query" size="10">
  <input type="submit" value="Search">
</form>
```

Please change "search.php" in the above HTML to the appropriate search page: that is, "search.asp" for ASP users, "search.cgi" for CGI users, and "search.html" for JavaScript users.

Note also, the relative location of the search script from where this HTML is placed (eg. if the form is used in a page that is one directory down from the search script, you would need to link to "../search.php").

Alternatively, you can use an absolute path to the search script (so that no matter where you place the above HTML, it will still find the file), for example:

```
<form method="GET" action="http://www.mysite.com/search.php">
  ...
</form>
```

If you wish to include more search options (such as match any search words / all search words, etc.) refer to the example in "[How do I modify the search form](#)".

6.7 Advanced template options

Zoom now provides advanced template options so you can customize the appearance of your search results outside of what is possible with CSS, and without having to modify the search script. This allows you to customize the location of headings, summary information, search boxes, and more. Simply by placing additional tags in your template file.

The default `<!--ZOOMSEARCH-->` tag will provide the complete standard results layout. This is recommended for users who are happy with the default layout, or who are new to Zoom customizations.



Note: The `<!--ZOOMSEARCH-->` tag and all of the following template tags are case sensitive. This means you must specify them in uppercase as shown here.

For more advanced users, you can now specify individual elements as necessary and remove the `<!--`

ZOOMSEARCH--> tag from use. Items can be repeated multiple times or not shown at all.

For example, one commonly requested feature in the past was to be able to reproduce the "Results Pages: 1 2 3 Next >>" links at the top of the results. You can do this by simply adding a <!--

ZOOM_SHOW_PAGENUMBERS--> tag before your results.

Here is a listing of the advanced template tag options available:

Template tags in "search_template.html"	Description
<!--ZOOM_SHOW_HEADING-->	This is the search result heading, e.g. "Search results for: cat food"
<!--ZOOM_SHOW_SUMMARY-->	This is the summary text at the top of the results, e.g. "x results found"
<!--ZOOM_SHOW_CATSUMMARY-->	This is only available when Categories are enabled. It is a breakdown of the categories that your search results belong to. E.g. "Refine your search by category: News (3) Knowledge base (2) PDF files (10)"
<!--ZOOM_SHOW_SUGGESTION-->	This is any spelling suggestion, or tip to search again when there are few results.
<!--ZOOM_SHOW_PAGESCOUNT-->	This is the number of pages of results found. E.g. "x pages of results."
<!--ZOOM_SHOW_SORTING-->	This is the option to switch between "Sort by relevance" and "Sort by date" when sorting by date is enabled in the Indexer.
<!--ZOOM_SHOW_SEARCHTIME-->	This is the time taken to perform the search. E.g. "Search took 0.2 seconds"
<!--ZOOM_SHOW_RECOMMENDED-->	This is the recommended links.
<!--ZOOM_SHOW_PAGENUMBERS-->	These are the links to the different pages of results, e.g. "Results Pages: 1 2 3 Next >>"
<!--ZOOM_SHOW_RESULTS-->	This show the actual search results.
<!--ZOOM_SHOW_SEARCHFORM-->	This produces the default search form. You do not need to use the <!--ZOOM_SHOW_FORMSTART--> and <!--ZOOM_SHOW_FORMEND--> tag for this. The single tag itself will create the entire search form for you.
<!--ZOOM_SHOW_FORMSTART-->	This specifies the start of a search form.
<!--ZOOM_SHOW_FORMEND-->	This specifies the end of a search form.
<!--ZOOM_SHOW_SEARCHBOX--> <!--ZOOM_SHOW_SEARCHBUTTON--> <!--ZOOM_SHOW_RESULTSPERPAGE--> <!--ZOOM_SHOW_MATCHOPTIONS--> <!--ZOOM_SHOW_CATEGORIES-->	These should be pretty self explanatory. They are the different elements that make up the search form. They must be within the <!--ZOOM_SHOW_FORMSTART--> and <!--ZOOM_SHOW_FORMEND--> tags mentioned above. Use these tags if you want to create a custom search form that is different to the default generated by <!--ZOOM_SHOW_SEARCH_FORM-->

```
<!--
ZOOM_SHOW_CUSTOMMETAOPTIONS-->
```

Advanced template options in JavaScript

The JavaScript version has a very different syntax for the template options, due in nature to the way the JavaScript language works. However, it is possible to configure the layout of most elements in a similar fashion to the other versions as described above.

By default, your "search.html" page will contain the following line to indicate where you want your search form, and results to appear within your page.

```
<script language="JavaScript">ZoomSearch();</script>
```

However, if you wish to specify the individual elements, you will need to remove this line, and insert each element separately.

To do this, first you must insert the following line before any other "Zoom" function call:

```
<script language="JavaScript">ZoomInitSearch();</script>
```

Then after this, you can specify any of the following tags to show the individual elements.

For a search form, you will need to begin and end it like this:

```
<script language="JavaScript">ZoomShowFormStart();</script>
...
<script language="JavaScript">ZoomShowSearchBox();</script>
<script language="JavaScript">ZoomShowSearchButton();</script>
...
<script language="JavaScript">ZoomShowFormEnd();</script>
```

Only form elements need to be enclosed in the `ZoomShowFormStart` and `ZoomShowFormEnd` calls. Everything else only needs to be after the `ZoomInitSearch()` line.

The full list of available "tags" are below:

Template tags in "search.html" for the JavaScript version	Description
<code><script language="JavaScript">ZoomShowHeadings();</script></code>	This is the search result heading, e.g. "Search results for: cat food"
<code><script language="JavaScript">ZoomShowSummary();</script></code>	This is the summary text at the top of the results, e.g. "x results found"
<code><script language="JavaScript">ZoomShowCategories();</script></code>	This is only available when Categories are enabled. It is a breakdown of the categories that your search results belong to. E.g. "Refine your search by category: News (3) Knowledge base (2) PDF files (10)"
<code><script language="JavaScript">ZoomShowPages();</script></code>	This is the number of pages of results found. E.g. "x pages of results."
<code><script language="JavaScript">ZoomShowSort();</script></code>	This is the option to switch between "Sort by relevance" and "Sort by date" when sorting by date is enabled in the Indexer.

<code><script language="JavaScript">ZoomShowSearchTime()</code>	This is the time taken to perform the search. E.g. "Search took 0.2 seconds"
<code><script language="JavaScript">ZoomShowRecommendedLinks()</code>	This is the recommended links.
<code><script language="JavaScript">ZoomShowPageNavigation()</code>	These are the links to the different pages of results, e.g. "Results Pages: 1 <u>2</u> 3 <u>Next</u> >>"
<code><script language="JavaScript">ZoomShowResults()</code>	This show the actual search results.
<code><script language="JavaScript">ZoomShowSearchForm()</code>	This produces the default search form. You do not need to use the <code>ZoomShowFormStart()</code> and <code>ZoomShowFormEnd()</code> tag for this. The single tag itself will create the entire search form for you.
<code><script language="JavaScript">ZoomShowFormStart()</code>	This specifies the start of a search form.
<code><script language="JavaScript">ZoomShowFormEnd()</code>	This specifies the end of a search form.
<code>ZoomShowSearchBox();</code> <code>ZoomShowSearchButton();</code> <code>ZoomShowResultsPerPage();</code> <code>ZoomShowMatchOptions();</code> <code>ZoomShowCategories();</code> <code>ZoomShowMetaFields();</code>	These should be pretty self explanatory. They are the different elements that make up the search form. They must be within the <code>ZoomShowFormStart()</code> and <code>ZoomShowFormEnd()</code> tags mentioned above. Use these tags if you want to create a custom search form that is different to the default generated by <code>ZoomShowSearchForm()</code> .

6.8 More how-to's

For more how-to's regarding customizing your search page and integrating it with your site, visit our Support and FAQ webpage online at:
<http://www.wrensoft.com/zoom/support/>

6.9 Changing your default editor

When you edit a HTML or script file from Zoom (from either the Templates menu or the Configuration Window), it will open up in the default editor for that file type in Windows. However, some users' computers may not be setup to use their preferred applications for certain file types (not to mention a lot of applications "hijack" certain file associations upon installation). If you want Zoom to edit a file type with a different application, you will need to change your Windows file associations.

To do this, you need to open up the "Control Panel" in Windows, and select "Folder Options". Click on the "File types" tab and select the extension you wish to modify ("HTML" for the search page, and "PHP", "ASP", or "JS" for the search scripts). Then click on the "Advanced" button, and select the "Edit" action from the list. Click on the "Edit..." button and here you can specify the application to open this file with. For more information, consult Windows Help.

Once you have done this, selecting the "Customize search page appearance" and "Modify search scripts" options in Zoom will always open the file in your preferred editor.

Note: Due to known issues with the way it handles HTML files, MS Word is NOT recommended for editing HTML files and in particular, the search page template for Zoom. One such issue is when it automatically tries to strip, remove, or re-arrange HTML comments it considers unnecessary – which happens to be required by Zoom. There are also general-use problems with the additional meta data it adds to a web page, which causes it to not render correctly or at all in most versions of Internet Explorer.

Part



Advanced Options

7 Advanced Options

7.1 International / foreign language support

Zoom supports a wide variety of languages such as French, German, Italian, Hebrew, Cyrillic (Russian), Greek, Arabic, and others. But the search page must be configured accordingly, depending on the encoding you have used on your website.

The Zoom Indexer application supports Unicode, and thus, multi-byte character strings. This allows you to index and configure your search engine for many languages using the UTF-8 encoding. However, due to the nature of the scripting platform and the server restrictions of your web host, searching for some Asian languages such as Japanese, Chinese, and Korean are limited. Searches will still be functional but accuracy is hampered due to the lack of a proper language definition to identify words.

For the following language settings to work, make sure your search template HTML file is specified for the correct charset. For example,

```
<meta HTTP-EQUIV="content-type" CONTENT="text/html; charset=windows-1252">
```

Zoom will warn you if it finds that the search template in the output directory does not have the same character set selected for the indexing configuration (specified under "Languages" in the Configuration window).



Note: By specifying the use of Unicode and UTF-8, the indexer will also store all entries in the search engine data files as UTF-8 encoded characters. This means that the translation of accented HTML entities (such as "ü," etc.) will be converted to UTF-8 encoded characters. When UTF-8 is not selected, the translation happens using standard ANSI instead.

See also:

[European languages \(French, German, Danish, Swedish, etc.\)](#)

[Russian \(Cyrillic\)](#)

[Japanese](#)

7.1.1 European languages (French, German, Danish, Swedish, etc.)

European websites commonly use the "Western European" character encoding, CP1252 (also known as "windows-1252"). A few web sites will also use "iso-8859-1" which is almost the same as "windows-1252", and some are in Unicode UTF-8.

If your website uses the "windows-1252" encoding:

- Select "Specify other encoding" in the Configuration window, and the name of the encoding you are using (ie. "windows-1252"), before indexing your website.

If your website uses UTF-8 encoding:

- Select "Use Unicode (UTF-8 encoding)" in the Configuration window before indexing your website.

7.1.2 Russian (Cyrillic)

Russian websites commonly use one of three encodings: KOI8, CP1251 (aka "windows-1251"), and Unicode UTF-8.

If your website uses KOI8 or "windows-1251" encoding:

- Select "Specify other encoding" in the Configuration window, and the name of the encoding you are using (ie. "windows-1251"), before indexing your website.

If your website uses UTF-8 encoding:

- Select "Use Unicode (UTF-8 encoding)" in the Configuration window before indexing your website.

7.1.3 Japanese

Japanese websites commonly use one of two encodings: Shift-JIS, and Unicode UTF-8.

Zoom does not currently support indexing Shift-JIS websites.

If your website uses UTF-8, ensure that:

- Select "Use Unicode (UTF-8 encoding)" in the Configuration window before indexing your website.
- We also recommend enabling "Substring match for all searches" option in the Configuration window, to allow for matching on words that are not delimited by a space character (as is often the case in Japanese).

7.2 Translating the search page

You can modify the text that appears on the search page by editing the Zoom Language Files (.ZLANG files) from the Configuration window (under the "Languages" tab). Here you can edit or create new ZLANG files, or select one of the default language files to use.



Note: Before you start editing ZLANG files, you should first create your own language file (eg. "MyFrench.zlang"). This is because the default files that come with Zoom (English.zlang, French.zlang, etc.) will be overwritten when you re-install the software, or upgrade to a new version. If you do make all your changes in the default file, and you re-installed Zoom and find that your changes have been overwritten, you can find a backup of them in the "lang-backup" folder under your Zoom program directory.

The ZLANG file format is simply a line-by-line entry, beginning with a "STR_" key word (such as "STR_FORM_SEARCHFOR"), followed by an equals sign, and the actual text to be used. You must always leave the "STR_" part of the line intact, and only modify the text after the equals sign. The following is an example of a ZLANG file.

```
STR_FORM_SEARCHFOR=Search for:
STR_FORM_SUBMIT_BUTTON=Submit
STR_FORM_RESULTS_PER_PAGE=Results per page:
STR_FORM_CATEGORY=Category:
STR_FORM_CATEGORY_ALL=All
STR_FORM_MATCH=Match:
STR_FORM_ANY_SEARCH_WORDS=any search words
STR_FORM_ALL_SEARCH_WORDS=all search words
```

Note that an entry can not span multiple lines (so you should disable "Word Wrap" in Notepad). Also, the STR key words are predefined and can not be modified. Below is an extract of a ZLANG file.

Use the "English.zlang" default file as your reference for text which can be translated, and the overall format of the file.



Note: All ZLANG files must be in Unicode encoding. You can save a file in Unicode with Notepad, by clicking on the File menu, and selecting "Save As...". Locate the "Encoding" dropdown menu at the bottom of the Save As file window. Select "Unicode" here, and save your file.

7.3 Importing and Exporting additional start URLs

In spider mode, you can import additional spider URLs from a text file by clicking on the Import button and specifying its filename. The following is an example of a text file that can be imported by this feature.

```
http://news.mysite.com/  
http://www.myalternativesite.com/  
http://www.mysite.com/spider_sitemap.html  
http://www.friends_site.com/
```

These URLs will be imported into the list and given the default option of "index and follow".

It can also import from a list of URLs with comma-separated values to specify the indexing option of that URL. The format for this advanced import list is as follows:

```
<URL>, <URL option>, [<base URL> [, LIMIT=<limit>] [, WEIGHT=<weight>] ]
```

The following is an example list in the above specified format:

```
http://news.mysite.com/, INDEX_AND_FOLLOW  
http://www.myalternativesite.com/, INDEX_AND_FOLLOW  
http://www.othersite.com/, INDEX_AND_FOLLOW, http://pages.othersite.com/  
http://www.mysite.com/spider_sitemap.html, FOLLOW_ONLY  
http://www.myschool.edu/bob/index.html, INDEX_AND_FOLLOW, http://www.myschool.edu/, LIMIT=50, WEI  
http://www.friends_site.com/, INDEX_ONLY
```

The URL options available correspond to the "Advanced Spider URL options" described in section 2.1.4 as follows:

- INDEX_AND_FOLLOW: Index page and follow links (default)
- INDEX_AND_FOLLOW_ALL: Index page and follow all links
- INDEX_ONLY: Index single page only
- FOLLOW_ONLY: Follow links only
- FOLLOW_ALL: Follow all links on this page only

See "[Start spider URL](#)" for more details on each of these options.

The LIMIT= flag must be followed by a value specifying a restriction on the maximum number of files to index for a particular start point. This is the same setting as the one which you can specify on the "[Advanced Spider URL options](#)" window.

The WEIGHT= flag must be followed by a value which specifies the boosting or indexing weighting for the pages under this start point. This should be a value between +5 and -5 similar to the PAGEBOOST meta value. It is the equivalent of specifying PAGEBOOST meta tags for each of the individual pages that are

indexed under the start point in question. See "[Weightings](#)" for more information.

You can also export your current list of URLs to a CSV text file by way of the Export button, which will export in the latter layout as given in the above example.

7.4 Editing the search script

If you are familiar with scripting, you can edit the search script source code by clicking on the "Templates" menu in the Zoom application and selecting "Edit search script source".

In most cases, you would not need to do this, because most customizations can be made via Zoom's Configuration window and search template.

However, some advanced users may want to do this in order to customize certain responses, or simply to add some Server Side Includes (SSIs) for additional navigation bars and other page elements that require server side processing (as this will not work in the search template HTML file).

7.5 Skipping sections of a page

Sometimes there are situations where you would want to stop a section of a page from being indexed. This may be common headers, footers, or navigation menus which appear on every page of your site. This can be accomplished by enclosing the unwanted section of the HTML document within the following tags: `<!--ZOOMSTOP-->` and `<!--ZOOMRESTART-->`. Note that this tag must be used as it is, in upper case, with no space characters within the tags. For example,

```
<p>This text will be indexed</p>
<!--ZOOMSTOP-->
<p>This section is skipped</p>
<p>and no indexing will occur</p>
<!--ZOOMRESTART-->
<p>Indexing starts again here</p>
```



Note: A `<!--ZOOMSTOP-->` tag must be followed by a `<!--ZOOMRESTART-->` tag somewhere later in the page for it to take effect. You also cannot have nested ZOOMSTOP tags. Zoom will simply skip starting from the ZOOMSTOP tag found, up to the first occurrence of ZOOMRESTART.

This is often used to exclude navigation bars and menus. Note however, that the hypertext links within a ZOOMSTOP and ZOOMRESTART section would still be followed in Spider indexing mode.

If you would like to stop the spider from following links, you can use the `<!--ZOOMSTOPFOLLOW-->` and `<!--ZOOMRESTARTFOLLOW-->` tags. Links within these tags would not be followed, but when used alone, the text would still be indexed.

If you wish to skip a section of a page from being indexed AND from having its links followed, you should use both tags to enclose a section. For example:

```
<p>This text will be indexed</p>
<!--ZOOMSTOP-->
<!--ZOOMSTOPFOLLOW-->
<p>This section is skipped and <a href="test.html">this link will not be followed</a></p>
<!--ZOOMRESTARTFOLLOW-->
<!--ZOOMRESTART-->
<p>Indexing starts again here</p>
```

Note that when you are using the two tags combined, you must specify the ZOOMSTOPFOLLOW tag

within the ZOOMSTOP block, as shown in the above example. A ZOOMSTOP tag within a ZOOMSTOPFOLLOW block will be ignored.

7.6 Manually add words into the index

There is often the need to add words into the index that are not present in the text of the page itself. By doing this it is possible to have search matches on commonly misspelt words and synonyms of words that do appear on the page. Watch the search log closely in order to identify the common words that people are searching for which generate 0 results. These words may be candidates for manual insertion into the index.

To add words into the index for a certain page, you need to create a "ZOOMWORDS" meta-tag in the HTML header of that page. For example, on a page about music, some additional words could be added. It is assumed that these words are not already in the text of the page.

```
<meta name="ZOOMWORDS" content="melody, mellody, tune, song">
```

The standard `<meta name="keywords">` tag is supported and can also be used. However, note that some external Internet search engines (such as Google) ignore or even penalise a page's rank if they find what they consider to be excessively repeated or misspellings, so the ZOOMWORDS meta-tag is provided to allow you to 'force' page ranks internally, within your own website, without affecting external search engines.



Note: You must check the "Meta keywords" box on the "Indexing Options" tab of the Configuration window in order to enable support for both ZOOMWORDS and Meta keywords.

7.7 Alternative page titles and descriptions

In some cases, you may have designed titles and meta descriptions which target Internet-wide search engines. For example, it may be part of your SEO (Search Engine Optimization) strategy, in having more generic titles that attract visitors to distinguish your site from another.

However, these titles and descriptions may not be so appropriate for the search results within your own website. For example, if every page of your site has a title which begins with "ACME Co. - Leading Provider of Electronic Goods - ..." then your internal search results (served by Zoom) will all look remarkably similar at first glance and hard to tell apart.

To avoid this, you can specify alternative titles and descriptions for Zoom to use. This is done by having a ZOOMTITLE and ZOOMDESCRIPTION meta tag.

Here is an example of what you could have in the header of a web page:

```
<title>ACME Co. - Leading Provider of Electronic Products - Latest News and
Announcements</title>
<meta name="description" content="News on ACME manufacturer and distributor of
everything from USB sandwiches to Firewire donuts!">
<meta name="ZOOMTITLE" content="Latest News and Announcements">
<meta name="ZOOMDESCRIPTION" content="Find out what's new here at ACME">
```

Note that meta tags need to be within the `<head>...</head>` section of the page, in compliance with HTML Standards.



Note: You must have "Title of page" enabled for indexing (on the "Indexing Options" panel of the Configuration tab) to index ZOOMTITLE meta tags, and "Meta description" for ZOOMDESCRIPTION tags.

7.8 Specifying a last-modified date for your web pages

You can configure Zoom to display the date and time for each search result and allow the user to sort their results chronologically (see section 2.3.2 and 2.3.6). This normally uses the last-modified date and time of the file according to the file system. However, this may not be very useful in cases such as dynamically generated web pages (where it would use the date and time of the server-side script as opposed to a date that is related to the content actually displayed).

In these cases, you can specify your own last-modified date and time in the form of a "http-equiv" Meta tag such as the following:

```
<meta http-equiv="Last-Modified" content="Sat, 07 Apr 2001 00:58:08 GMT">
```

Note that this is a standard Meta tag and would also be used by other spiders and applications. Note also that the exact syntax for specifying the date and time must be followed.

The expected format for the date value is:

```
ddd, DD MMM YYYY HH:MM:SS [Timezone]
```

- ddd is the first three letters corresponding to the day of the week (eg. "Mon", "Tue", ...)
- DD is the day of the month in numerals
- MMM is the first three letters for the name of the month (eg. "Jan", "Feb", ...)
- YYYY is the year in numeral format (eg. 2005)
- HH:MM:SS is the time in hours, minutes, and seconds respectively. This must be specified in 24 hour time.
- [Timezone] is optional (eg. "GMT", "GMT-5"). Note that Zoom will not perform any timezone conversions based on this.

7.9 Enable jump to match and highlight within document

You can setup highlighting to occur on the actual page of your website when you click on a search result. However, this requires you to add some Javascript to each page of your site where you want this feature to take effect. It will highlight the word that the user was looking for in the search engine, as well as scroll the window down to the first appearance of the word.

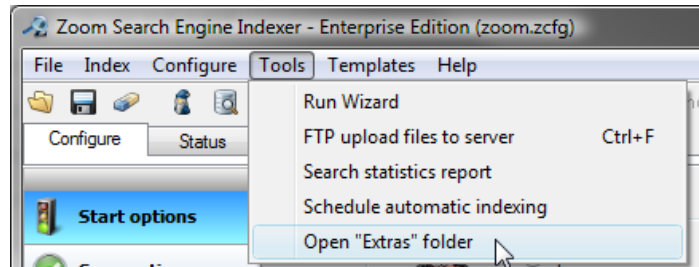


Note: This feature only works for HTML web pages, and does not work for Word, PDF, XLS and other document types. It also will not work for Javascript disabled browsers. For PDF documents, you can enable a similar behaviour within Acrobat Reader via the "Highlight and locate" option in the PDF plugin settings. See "[Configuring a plugin](#)" for more information.

First, you must enable this option in the Configuration window (under the "[Results Layout](#)" tab).

Once you have enabled this option and re-indexed your website, you must locate the "highlight.js" file

included with Zoom. This file can be found in the "extras" folder, which you can access from Zoom Indexer, by clicking on "Tools" in the menu and selecting the option to "Open 'Extras' folder" (also see "[Where is the Program Data directory?](#)").



Copy the "highlight.js" file from the "extras" folder to your website files, as you will need to link to the "highlight.js" file from each page of your site that requires the feature.

For example, you could paste the following HTML in your site's header or footer:

```
<style type="text/css">.highlight { background: #FFFF40; }</style>
<script type="text/javascript" src="highlight.js"></script>
```

Note that you will need to specify the correct path to "highlight.js" depending on where the file is located relative to the page. You will then need to modify the BODY tag on your page to include an "onLoad" attribute, such as:

```
<body onload="highlight();">
```

If for some reason you can not modify the body tag of your page, an alternative would be to put the following line after the </body> tag of your page:

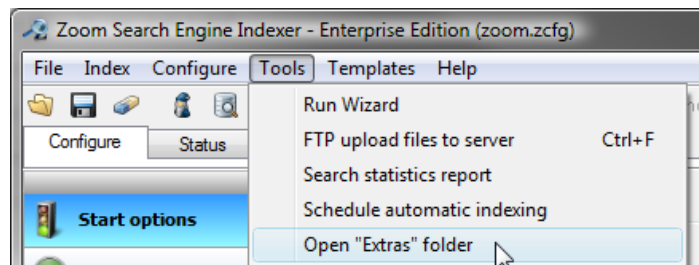
```
<script type="text/javascript">highlight();</script>
```

Once you have successfully applied the above changes, you should find that clicking on a search result will go straight to

7.10 Search Statistics PHP Script

If you have PHP support on your web server, you can use a provided script ("report.php") to provide online reporting for your search log. This allows you to provide live, up-to-date statistics on your website without needing to download the log file manually and generate/upload your report.

You can find this script in the "extras" folder. From the Zoom Indexer, click on the "Tools" menu and select "Open 'Extras' folder" to locate this file (also see "[Where is the Program Data directory?](#)").



Installation

Copy or upload this script to the same folder as the log file on the server.

If your logfile does not have the default name of "searchwords.log" open the php script in a text editor

and change this line:

```
$LOGFILENAME = "searchwords.log";
```

To correspond to the filename of your log file.

Customization

PHP Variables

Inside the php script near the top of the file are customizable options. The php script can be edited with any basic text editor. To change an option you should edit the value to the right of the equals sign. If the value is encased in quotation marks they should be left in place. Also, the semi-colon at the end of each line must remain in place as well for the script to work.

```
$LOGFILENAME = "searchwords.log";
```

The name of the log file that the statistics will be generated from.

```
$TEMPLOGNAME = "_tempsearchwords.log";
```

Name of temporary log file. This file will be created and written to during a trim operation. If the file already exists it will be overwritten.

```
$BAR_LENGTH = 300;
```

The length in pixels of the longest bar in each bar chart .

```
//$PASSWORD_REPORT = "password";
```

This is the password required to generate a report. Note that the line begins with //, this means that the line is commented and is not active. To enable password protection for generating a report remove the // from the beginning of the line and change "password" to whatever you would like the password to be.

```
//$PASSWORD_TRIM = "password";
```

This is the password required to trim a log file. Note that the line begins with //, this means that the line is commented and is not active. To enable password protection for trimming the log file remove the // from the beginning of the line and change "password" to whatever you would like the password to be.

```
$ENABLE_TRIM = 1;
```

This line enables or disables the ability of the script to trim a log file. By default this is set to 1 (on). If you would like to completely disable the trim functionality from the script please set this to 0.

```
$_DEBUG = 0;
```

Debug data on or off, you probably don't want to change this setting.

CSS Styles

Near the top of the PHP script are some CSS styles that can be used to change the look of the report. Feel free to edit these however you like.

Report Configuration

Start Date / Finish Date

The reports will be generated based on a date range specified in these fields. The date range is inclusive of the dates specified. The recommended format for entering dates is in the (year-month-day) format however the script is able to recognize many formats so you can try entering in the date however you like. The dates will be displayed in year-month-day format on the final report page so you will know if the script understood what you meant. Some example dates are:

- "now" – Today's date
- "3rd March 2007"
- "-1 Month" – Today's date minus 1 month

Top 10 Search phrases

This provides a bar chart of the top 10 search phrases made on your website. This gives you a good indication of what people are looking for on your website.

Top 10 Search Phrases that returned no matches

This provides a bar chart of the top 10 search phrases on your website which found "no results". This is useful for determining what your visitors are searching for on your website but not managing to find. You can use this information to provide content that better cater your website to your visitors (or determine what meta keywords to add to allow your users to find what they are looking for).

Searches per day

This provides a bar chart representing the number of searches that are made on your website per day.

Searches per week

This provides a bar chart representing the number of searches made on your website per week.

Searches per month

This provides a bar chart representing the number of searches made on your website per month.

List the top x searched words (sorted by popularity)

This option appends a HTML table of search words (sorted by popularity). You can specify the number of searched words you would like to list here.

Raw Data Output

A table containing the raw data from the log file for the date range specified. Because this data can be quite large a maximum of 5000 log entries will be displayed. If you need to see more consider opening the log file directly in a text editor.

Password

If you have enabled password protection for generating logs you will need to enter your password on the configuration page.

Trim Functionality

The trim section of the configuration page allows you to delete entries out of the log file, making it smaller and faster to generate reports from. Consider using this if report generation is taking too long or timing out. Please note however that this operation cannot be undone and any deletions are permanent.

Password

If you have enabled password protection for trimming the log file you will need to enter your password here.

Trim Log Entries Before

Any log entries before this date will be permanently deleted. Like the start/finish dates in the report configuration this field can take a number of formats, once you click the Trim Log button you will be taken to a confirmation page that will display the entered date in year-month-day format to confirm the script understood what you meant and to make sure you really want to delete the records.

7.11 Integrating Zoom in your own applications

Software developers looking to include a search engine feature as part of their own applications can integrate Zoom Indexer into their software and provide dynamic indexing and searching on demand.



Note: A single license of Zoom is required for each copy of the Zoom Indexer application installed on a computer. This does not permit you to distribute the registered version of the Zoom Indexer software. A special distribution license (and a custom built version of Zoom) must be purchased separately if you wish to distribute Zoom Indexer with your software. Please [contact us](#) for more details.

Console mode (Enterprise Edition only)

The Enterprise Edition of Zoom allows the Indexer to run in "console mode". When Zoom is running in this mode, it will not display a GUI nor require any user interaction. With the use of pre-configured .ZCFG files, the Zoom Indexer can then be spawned as a process to index files as needed.

All indexing progress, and error messages will be written to "stdout" and you can capture the output for internal analysis, or display to the user as you wish. The Indexer will also return an exit code indicating the outcome of the indexing.

Console mode must be specified alongside an AutoRun option, and a configuration filename. The following example will run the Indexer in console mode, indexing offline, using the configuration settings defined in "zoom.zcfg":

```
ZoomIndexer.exe -c -o zoom.zcfg
```



Note: If "stdout" is not redirected to a pipe or file, Zoom will automatically create a console window to display the output. Note also that Zoom will always display a splash screen, even in console mode. We can provide a custom build without the splash screen for developers who have purchased the special distribution license. Please [contact us](#) for more details.

Redirecting stdout

This can be done programatically (by the software spawning this process) or via the command line with the use of the greater than sign (">"). eg. redirecting stdout to file:

```
ZoomIndexer.exe -c -o zoom.zcfg > output.txt
```

Exit codes

The Indexer will also return an exit code indicating the outcome of the indexing. The possible exit codes are:

- 0 Indexer executed successfully.
- 1 Indexing completed with the maximum unique words limit having been reached.

- 2 Indexing completed with the maximum pages limit having been reached.
- 3 Invalid configuration
- 4 Indexing failed
- 5 Writing the index data output failed
- 6 Report generating failed
- 7 FTP upload failed
- 8 Installation error

7.12 The indexing process

For your information, we have documented the basic process in which text is extracted from the HTML source file. This may be useful if you are looking into the operation of the Indexer to determine why certain content is not being indexed, or if you are attempting to make advanced changes to the script.

1. Remove all text between PHP markers (`<? ... ?>`)
2. Remove all text between ASP markers (`<% ... %>`)
3. Remove all text between Zoom comments (`<!--ZOOMSTOP--> ... <!--ZOOMRESTART-->`)
4. Remove all text between HTML comments (`<!-- ... -->`)
5. Remove all Java scripting sections (`<script> ... </script>`)
6. Remove all style sections (`<style> ... </style>`)
7. Extract the Title information and description meta-tag information from the file header.
8. Remove all HTML tags (`<...>`)
9. Convert HTML character entities and numerical entities back to plain text.
10. Index all remaining text.
11. Index any additional key words found in the "ZOOMWORDS" and "Keywords" meta-tag.

7.13 Word delimitation when indexing

When building the index, the indexer must decide how to split a sentence up into words. Any characters in the following range are considered to be part of a word.

- Lower case characters, 'a' to 'z'
- Upper case characters, 'A' to 'Z'
- Numbers, '0' to '9'
- Foreign characters, 'À' to 'ÿ'
- A join character (defined by the user – eg. dot (.), dash/hyphen (-), underscore, etc.) immediately followed by another valid character (one of the above), eg. "2.5" and "F.B.I.". See "[Indexing options](#)" for more information.

Any characters not in this range will force the current word to end and a new word to start. For example,

based on the default configuration, this sentence,

"Record number 653-45+ABCD is invalid"

will be broken up into 6 words,

"Record", "number", "653-45", "ABCD", "is", "invalid"

7.14 Word delimitation when searching

When a visitor on your search page enters a search query, the script must decide how to split a search phrase up into search words. This is now determined based on the same settings you make for indexing (see "[Word delimitation when indexing](#)" and "[Indexing options](#)" for more information), so if words containing dots are indexed as word joining characters, then searches containing dots will also be treated as a single word.

7.15 Other tips

- A small index is a fast index. Filter directories, pages and words that contain redundant or useless information (from a search engine point of view) by use of the Skip options. This can reduce the size of the data files and speed up the search process.
- Save your configuration for each website, and create a shortcut that you can just double click on to re-index the whole site. See the command line parameter section for more information on how to do this. You can also schedule Zoom to index automatically on a regular basis.
- If you need to boost a page up the search result ranking for a particular search term but don't want to change the text, use the ZOOMWORDS meta-tag and include the search term a few times.
- Turn on the search term logging feature. If users are matching too many pages use the Boolean AND option for multiple search terms. If visitors are often getting zero results try adding some additional keywords in the ZOOMWORDS meta-tag.

Part

8

**Troubleshooting and
support**

8 Troubleshooting and support

8.1 Frequently Asked Questions (FAQs)

We provide an extensive and up-to-date online support section with many Frequently Asked Questions (FAQs) and solutions.

Please visit our support site at:

<http://www.wrensoft.com/zoom/support/>

8.2 How to enter in my license key

After purchasing the software a license key is sent out via E-mail. This license key needs to be entered into the Zoom software (under the "Help" menu, select "Enter license key").

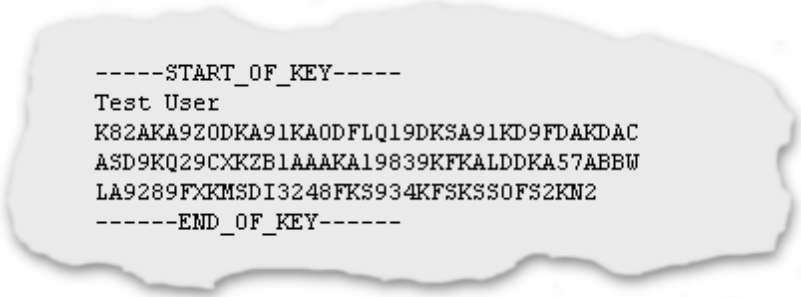
When entering a license key, copy and paste the license key from the E-mail. Doing a copy and paste will avoid the possibility of a typing mistake.

License keys purchased for previous major versions of the software will not work in the latest (e.g. license keys for V5 will not work with V6), and vice versa. So check the software version matches the key you purchased. If you have an old license key, you can upgrade to the latest version or you can continue to use the older version. More information on upgrades can be found on our website here:

<http://www.wrensoft.com/zoom/upgrade/upgrade.html>

Find your license key

After you have placed an order you will receive an e-mail that contains details about your order, your user name and your license key. It should look something like this:



```
-----START_OF_KEY-----  
Test User  
K82AKA9ZODKA91KAODFLQ19DKSA91KD9FDAKDAC  
ASD9KQ29CXKZB1AAAKA19839KFKALDDKA57ABW  
LA9289FXKMSDI3248FKS934KFSKSSOFS2KN2  
-----END_OF_KEY-----
```

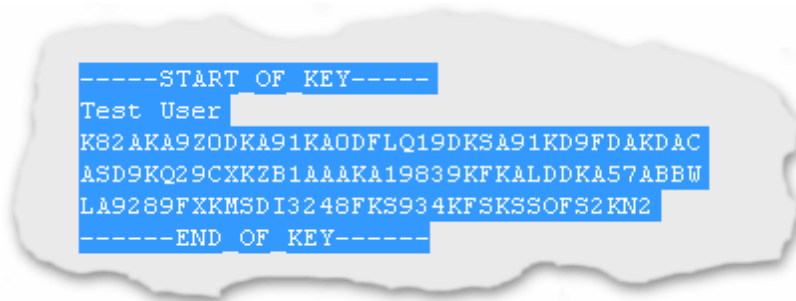
Note that the keys may vary in length and be shorter or longer than the examples above.

Step 1 - Make sure you have the right software

Make sure that the product that you have downloaded and installed, matches the version of the product you have purchased. Note that the key should be entered in the Free Edition of the software to transform it to the registered edition you purchased. Download and install the latest version of the software, if required.

Step 2 - Copy your user name and key from the E-mail

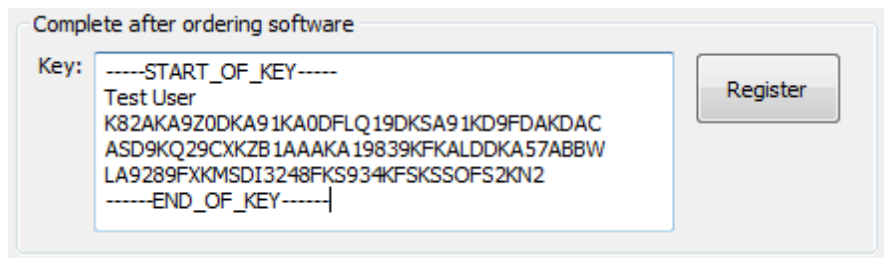
Select the entire key, including the -----START_OF_KEY----- and -----END_OF_KEY----- flags:



Copy your key to the clipboard. This can be done by using the Edit / Copy menu item in most E-Mail programs. Alternatively you can use the CTRL-C key combination on the keyboard.

Step 3 - Paste your user name and key into the software

Start up the software and select "Help"->"Enter license key" in the menus. Paste the key in the window provided by right clicking and selecting "Paste" or by using the CTRL-V key combination on the keyboard.



Click on "Register". If the user name and key was accepted, the program will restart and identify itself as the registered edition of the software in the title bar of the window.

Remember to keep your key safe

The e-mail containing the license key should be kept in a safe place in case the software ever needs to be reinstalled. Your User Name and Key will also be required to be re-entered when software upgrades are released.

Still have a problem?

If you still have a problem, check the following.

- No extra characters were included, be especially careful about not copying extra space characters or new line characters.
- Your user name is exactly as it appears in the E-Mail, using a different user name will not work.
- If you typed in your user name or key, rather than copying and pasting, check that you have not made a typing mistake and check that upper and lower case characters are correct. Upper and lower case are important.

Contact us

If the above doesn't fix your problem, [contact us](#) and describe the problem you have encountered and include your order number and key.

8.3 Discussion forums

You can visit our online discussion forum for help and advice on using Zoom. It is an active community made up of many Zoom users sharing and seeking hints and tips, and is also frequented by the Wrensoft developers and support team.

Please drop by and say hello, or feel free to post a question:

<http://www.wrensoft.com/forum/>

8.4 Known issues

The following are some known issues, technical limitations and their possible solutions or workarounds. Note that we can provide custom development and support. So if you need a new feature, a work-around to a known issue, or some serious technical support, e-mail us at zoom@wrensoft.com for a quotation.

- When given links (or a start URL in spider mode) without the filename (eg: <http://mywebsite.com/>), it may be indexed twice if there is another link somewhere else on the website which points to the same place, but with the actual filename specified (such as <http://mywebsite.com/index.html>, <http://mywebsite.com/home.htm>, etc.). To prevent this, use the "[Duplicate page detection](#)" option.
- When URLs with identical content are found using the "[Duplicate page detection](#)" option, only the first URL encountered will be indexed.
- Spidering dynamic websites that utilise session IDs as part of their URL (as GET parameters) will vary in effectiveness depending on how the website uses the session information. There is no easy solution to identifying session IDs, so you will have to configure the indexer to scan accordingly, if you do not wish Zoom to index the parts of your website which require session information.
- Some documents in PDF format may make use of complicated layouts such as columns, etc. This can confuse the indexer's context descriptions (finding words from different columns but on the same line) and exact phrase matching. Try the alternate PDF scan mode (see "[Configuring a plugin](#)" for more information) if you are having this problem.
- Very long words that have more than 35 characters in the word are split into two words in the index.
- Highlighting may not work on some words with special accented characters.
- An exact phrase can not be matched if it begins with a skipped word. The user will get a message saying that the phrase is in the skipped word list.
- When "substring matching" is enabled, a skipped word may return results. For example, if the skipped word "of" is searched with this function, it may return results containing "office" or "often". Note also that the highlighting feature may highlight occurrences of the word "of" in this case, despite it not being used as part of the searching criteria (i.e. the word was "skipped" in the indexing and searching, as intended).

8.5 Technical limitations

We document technical limitations of the software because we understand the importance for advanced users or developers to have the most comprehensive information possible about their websites and development environment. Note that we can provide a quote for custom modifications if you require different specifications.

Maximum word length	35 characters. (Fixed limit)
Maximum title length	512 characters. (Fixed limit)
Maximum URL length	2083 characters (Fixed limit)
Maximum number of words to skip	400 (Fixed limit)
Maximum number of entries in the Page Skip list	1000 (Fixed limit)
Maximum number of extensions to scan for	50 (Fixed limit)
Maximum length of meta keywords tag (This is a shared limit with the ZOOMWORDS meta tag, and the Dublin Core meta tags)	20000 characters (Fixed limit)
Maximum number of dropdown values for any single Custom Meta field	255

8.6 Notes for users upgrading from Zoom 4.x

Due to the introduction of new CSS classes (and the removal of some old CSS classes), we would recommend that you allow Zoom to create a new copy of "search_template.html" in the output directory (that is, remove or rename your existing search template). This way you can view the new default search template, and adopt the CSS styles used to specify the appearance of the recommended links, and other new attributes.

See the "[CSS class listing](#)" for more detailed information on the new CSS classes used.

8.7 Notes for users upgrading from Zoom 3.x

Users upgrading their sites and indexing configurations from Zoom 3.x should note some significant differences with the use of template files.

- The Indexer will no longer overwrite the "search_template.html" (or "search.html") file in the output directory IF the file already exists.
- You should NOT modify "search_template_src.html" and "search_src.html" in the Zoom installation folder. You should now only need to modify the search template file in the OUTPUT directory (where the index files are written to).

All existing users should allow Zoom to overwrite the "search_template_src.html" and "search_src.html" files found in the Zoom installation folder (usually "C:\Program Files\Zoom Search\scripts\..."). Backup your customized version of these files, if necessary, before hand.

The new method of how you should go about changing the search template is as follows:

1. When you index a site for the first time, Zoom will create a copy of the default template in the output directory. The default templates are the "search_template_src.html" and "search_src.html" files (which you should no longer modify).

2. Subsequent indexing sessions will NOT overwrite the search template in the output directory. This means that you can make your modifications to the search template file directly, and save your changes to the output directory.
3. You no longer need to manage multiple "Custom search template paths" in the Advanced configuration tab. You simply have different search templates in different output folders.
4. When a search template file is missing in the output directory, Zoom will create a copy of the default template there the next time it indexes. This means that, if you need to return to the default template at any point, you can delete or rename your template in the output directory, and re-index.

It is recommended to start with a fresh new default search template from Version 5.0 and adapt that to the look of your 3.x template. This means you should index once to an empty directory, so that it creates the default template file for you to use as a reference. There have been some changes with new stylesheet classes defined, and more importantly for the Javascript version, it now needs to link to different script (.js) files than before.

8.8 Contacting us

Web: <http://www.wrensoft.com/zoom/>

E-mail: zoom@wrensoft.com

Visit our website for more contact information.

Part

9

Appendix

9 Appendix

9.1 What is PHP, ASP, CGI, or JavaScript?

This chapter serves to provide beginners with an explanation of some of the technologies and terminologies used in Zoom.

9.1.1 What is PHP?

PHP (PHP: Hypertext Preprocessor) is a software package installed on web servers to provide scripting capabilities. It is commonly found on web servers running Apache (as it is now built-in by default), but it can be installed on IIS servers also. Since March 2004, there are a reported 15,528,732 website domains which use PHP (source: Netcraft Survey).

You can check with your web host to find out if PHP is available on your hosting account. Alternatively, you can upload a text file containing the following (as test.php):

```
<?php
phpinfo();
?>
```

When you open this file up from your browser (<http://www.mysite.com/test.php>), it should provide detailed information on the PHP installed on the server. If not, this means you do not have PHP support configured on your server and you should consult your web host or use one of the other platform options available.

9.1.2 What is ASP (Classic ASP)?

ASP (Active Server Pages), also known as Classic ASP, is the Microsoft equivalent to PHP, and comes packaged with most default IIS (Internet Information Services) web servers. Chances are if you have a Windows-based web server, you will have ASP available to you. It can also be available on some non-IIS servers, but they are less common.

Note that ASP is not the same as ASP.NET (Microsoft's new server-side platform), and they are not compatible.

Check with your web host to find out if ASP is available on your hosting account.

9.1.3 What is ASP.NET?

ASP.NET is a web application framework from Microsoft. It is NOT the same as ASP (which is also referred to as "Classic ASP") and have very different requirements. It is generally only available on IIS web servers.

This option is only recommended for advanced web developers who are working on existing .NET websites and need to integrate the search engine into their .NET web application. It requires a Zoom ASP.NET Native Control to be installed directly onto the web server. Please see "[Installing the ASP.NET native control](#)" for more information.

9.1.4 What is CGI?

CGI (Common Gateway Interface) is a method of running programs on a server over the web. This is different to PHP and ASP in that it does not have to load and interpret a "script", and is not limited by the

technical capabilities of a scripting platform. In fact, the scripting engines for PHP and ASP are CGI applications themselves. As such, CGI provides a way to run web applications requiring maximum performance and efficiency, and you will find it used on most enterprise-scale sites such as popular sites like eBay, Google, and Yahoo.

Due to the less restrictive nature of CGI applications, some web hosts (especially those offering cheaper packages) do not provide CGI support for security reasons. In addition to this, setting up and installing CGI applications can be more complex, especially if you have never installed one before.

For information on installing and uploading the Zoom CGI, please refer to "[Additional notes for uploading CGI](#)".

9.1.5 What is Javascript?

Javascript is a scripting language that allows a web page to tell the browser what to do. It is a client-side language, meaning that it is interpreted and processed on the computer viewing the web site – not on the server. This means that it is usually very limited in what it can do, but can be convenient since it does not rely on any special requirements on the server-side, and can also run off a CD with no web server present.

Please note that Javascript is not the same as Java, and they bear only marginal similarities in syntax. Java is a powerful, general-purpose programming language and can run outside of a web browser.

Zoom provides a Javascript version of the search script to cater for users who wish to run their search engines on a CD-ROM, or who have a very restricted hosting environment with no server-side scripting capability. However, due to the technical constraints of Javascript on most browsers, we are unable to provide the full number of features (such as context descriptions) that are available with the other platform options. It is also unsuitable for large volumes of data, and sites of over 1000 pages or with a large number of unique words, will not run on some browsers.

There is an alternative to running Zoom on CD/DVDs, and that is to include a specialised mini-web server on your CD, so that you can run the CGI or PHP version off the CD. More information is available in "[Using the CGI or PHP version on a CD/DVD distribution](#)".

9.2 Base URL for CD-ROM distribution (absolute and relative paths)

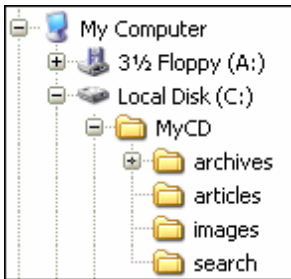
Understanding absolute and relative paths are important to determine the most consistent and effective way of linking to files, and it is useful knowledge to all web developers.

The following is a brief explanation for those who are not yet familiar with these forms of file paths, which are especially necessary for defining the base URL for CD-ROM and offline distribution.

Absolute paths

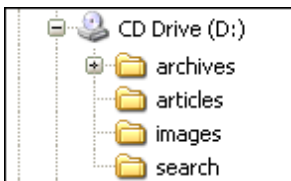
Absolute paths refer to specific locations, independent of the current location of the file. These usually begin with a "/" or even the domain of the site (eg. "http://mysite.com/"). When a path begins with a "/", this is a reference to the root directory.

For instance, if you have the following files:



Now assuming you have your search files in the search folder, and you need to link to a file in the articles folder. An absolute link would be `"/MyCD/articles/page1.html"`. Note that this link refers to the location of the file from the root directory, and thus it will refer to the same file regardless of where you place this link.

However, when you copy the contents of the "MyCD" folder to your CD-ROM, it may now look something like this:



As you can see, a reference to `"/MyCD/articles/page1.html"` would no longer be valid. The link would now have to be `"/articles/page1.html"`.

Because of this issue, we generally recommend against using absolute paths if you intend to publish your files on a CD-ROM. It also causes similar issues to those described above on different Operating Systems, such as the Mac, which refers to the CD drive as a part of the path (eg: the path would be `"/cdrom/articles/page1.html"`).

Relative paths

Relative paths refer to file locations based on the current directory. That is, they often refer to locations in the form of "up one directory level" and often begin with one or two dots ("`..`"). An example would be `../index.html` which would refer to the file "index.html" one directory level up.

A key to relative paths:

- A single dot ("`.`") refers to the current directory, eg. `./test.html` would refer to the file "test.html" in the current directory.
- Double dots ("`..`") refer to the previous directory, eg. `../test.html` would refer to the file "test.html" in the previous directory.
- To refer to a location which is more than one directory level up, use a combination of double dots, eg. `../../test.html` would refer to a file two directory levels up.

In the case of the illustrated example above (for absolute paths), you would link to the same file with a relative path of `../articles/page1.html`. This would then still be correct when your files are moved off to a CD-ROM, because the articles folder is always one level up from the search folder.

What should I use as the base URL for a CD or DVD?

If you need your search files to run on a CD-ROM (this also assumes you are using the Javascript version), in most cases, a base URL of `./` would work. This assumes that your search files will be located in the root directory of your CD.

If however, you host your search files on a subdirectory, eg. a "search" folder off the root directory of the CD, then you would need a base URL of `../` to refer to the linked files. Refer to the above section on relative paths for more information on determining the correct relative path.

9.3 Where is the Program Data directory?

The Program Data directory is used to store files which are used by Zoom. This includes the plugins, source code to the search scripts, the ZLANG language files, and some "extra" scripts for advanced features.

Previous versions of Zoom stored these files within the "Program Files" directory, but with the introduction of Vista, Microsoft has introduced policies whereby files such as these must be stored in a variety of different locations depending on the operating system you are using.

Windows Vista

On Vista, this would typically be the following:

```
C:\ProgramData\Wrensoft\Zoom Search Engine Indexer\
```

You may need to enable viewing of hidden directories to see the "ProgramData" folder - to do this, press ALT-T in Explorer, select "Folder Options"->"View"->"Files and Folders"->"Hidden files and folders"->"Show hidden files and folders"

Windows XP and Windows 2000

On XP and 2000, it is usually something like this:

```
C:\Documents and Settings\All Users\Application Data\Wrensoft\Zoom Search Engine Indexer\
```

You may need to enable viewing of hidden directories to see the "Application Data" folder - to do this, press ALT-T in Explorer, select "Folder Options"->"View"->"Files and Folders"->"Hidden files and folders"->"Show hidden files and folders"

9.4 End User License Agreement (EULA)

PassMark® Software Pty Ltd trading as Wrensoft™ ('Wrensoft')
End User Licence Agreement ('EULA')

IMPORTANT! PLEASE READ THE FOLLOWING TERMS AND CONDITIONS.

YOU, THE INSTALLER OF THIS SOFTWARE, AGREE THAT ALL OF THE TERMS AND CONDITIONS DESCRIBED BELOW APPLY TO YOU AND ANYONE ELSE WHO USES THIS SOFTWARE, IF EITHER; YOU CLICK THE "ACCEPT" BUTTON, OR YOU COPY, INSTALL, OR USE THIS COPY OF WRENDOFT SOFTWARE, OR YOU PERMIT OR ENABLE OTHERS TO COPY, INSTALL OR USE THIS WRENDOFT SOFTWARE.

IF YOU DO NOT AGREE TO BE BOUND BY THE TERMS AND CONDITIONS DESCRIBED BELOW,

PLEASE TERMINATE INSTALLATION IMMEDIATELY AND DO NOT USE THIS SOFTWARE**SOFTWARE COVERED BY THIS LICENCE**

This licence agreement (“Agreement”) applies only to the version of the software package [Zoom Search Engine](#) with which this Agreement is included. Different licence terms may apply to other software packages from Wrensoft and licence terms for later versions of [Zoom Search Engine](#) may also be changed.

TITLE

Wrensoft or its licensors own the [Zoom Search Engine](#) software package, including all materials included with the package. Wrensoft owns the names and marks of ‘Wrensoft’™, ‘[Zoom Search Engine](#)’ under copyright, trademark and intellectual property laws and all other applicable laws. Wrensoft is both a trademark and brand used by PassMark Pty Ltd to market and promote the [Zoom Search Engine](#).

TERMINATION

This licence will terminate automatically if you fail to comply with any of the terms and conditions, limitations and obligations described herein. On termination you must destroy all copies of the Wrensoft package and all other materials downloaded as part of the package.

DISCLAIMER OF WARRANTY

Wrensoft disclaims any and all warranties express or implied, including any implied warranties as to merchantability or fitness for a particular purpose. You acknowledge and agree that you had full opportunity to test [Zoom Search Engine](#) before any live, public or production use, that you assume full responsibility for selecting and using [Zoom Search Engine](#) and any files that may be created through the use of [Zoom Search Engine](#) and that if you use [Zoom Search Engine](#) improperly or against instructions you can cause damage to your files, software, data or business. The entire risk as to quality and performance of [Zoom Search Engine](#) is borne by you. **This disclaimer of warranty constitutes an essential part of the agreement.** Some jurisdictions do allow exclusions of an implied warranty, so this disclaimer may not apply to you and you may have other legal rights that vary by jurisdiction.

LIMITATION OF LIABILITY

In no event shall Wrensoft, its officers, employees, affiliates, contractors, subsidiaries or parent organisations be liable for any incidental, consequential, or punitive damages whatsoever relating to the use of [Zoom Search Engine](#), files created by [Zoom Search Engine](#) or your relationship with Wrensoft. Some jurisdictions do not allow exclusion or limitation of liability for incidental or consequential damages, therefore the above limitation may not apply to you.

HIGH RISK ACTIVITIES

The [Zoom Search Engine](#) is not fault-tolerant and is not designed or intended for use or resale as on-line control equipment in hazardous environments requiring fail-safe performance, such as in the operation of nuclear facilities, aircraft navigation or communication systems, air traffic control, direct life support machines, or weapons systems, in which failure of [Zoom Search Engine](#) could lead directly to death, personal injury, or severe physical or environmental damage (“High Risk Activities”). Wrensoft and its suppliers specifically disclaim any express or implied warranty of fitness for High Risk Activities.

LINKS TO THIRD-PARTY SITES

Wrensoft is not responsible for the contents of any third-party sites or services, any links contained in third-party sites or services, or any changes or updates to third-party sites or services. In the case where Wrensoft is providing those links and access to third-party sites and services to you only as a convenience, and the inclusion of any link of access does not imply an endorsement by Wrensoft of the third-party site of service.

ADDITIONAL SOFTWARE

This EULA applies to updates, supplements, add-on components or internet based services components of the software that Wrensoft may provide to you or make available after the date you obtain your initial copy of the software, unless they are accompanied by separate terms.

UPGRADES

To use software identified as an upgrade, you must first be licensed for the software identified by Wrensoft as eligible for the upgrade. After installing the upgrade, you may no longer use the original software that formed the basis of your upgrade eligibility, except as part of the upgraded software (unless otherwise authorised in writing by Wrensoft).

EXPORT RESTRICTIONS

You acknowledge that the software is subject to Australian export jurisdiction. You agree to comply with all applicable international and national laws that apply to the software including destination restrictions issued by Australia and other governments.

SOFTWARE TRANSFER

You may transfer your copy of the software to a different device. After the transfer, you must completely remove the software from the former device (except for Free Edition).

Transfer to Third Party of Free Edition

The Free Edition of the software may be freely distributed provided that no fee is charged and the product is distributed complete, intact and unedited by anyone other than Wrensoft.

Transfer to Third Party of other Editions (Standard, Professional & Enterprise)

This licence is granted exclusively to you, the original licensee, and therefore no right to resell, transfer, or re-assign the license is granted. An exception may exist for manufacturers, distributors and dealers/resellers of computer systems or computer software who have specifically negotiated for such an exception with Wrensoft to resell a particular license key as part of an installed system or as an authorised reseller of the software on its own.

See www.wrensoft.com for specifications of different editions.

SITE LICENSES

If this software is being installed as part of a Site License purchase, then following conditions apply:

The software may be installed on an unlimited number of computer systems provided that:

- 1) The computers on which the software is installed belong to the one legal entity. Subsidiaries, parent companies, brother/sister companies, affiliates and/or agents are not considered to be the same legal entity and are therefore not entitled to have the software installed on their computer systems unless specific permission is granted by Wrensoft.
- 2) The computer systems must all be situated in the one country. It is permissible that the computers be located in different cities or states within the one country.
- 3) All such computers are the property of, or are being leased or borrowed by the licensee and are on the premises of the licensee.
- 4) In the event that the computers are leased or borrowed, the software must be removed prior to the computer being returned to its legal owner.

NO RENTAL/COMMERCIAL HOSTING

You may not rent, lease or lend the software.

LIMITATIONS ON REVERSE ENGINEERING, DECOMPIATION AND DISASSEMBLY

You may not reverse engineer, decompile, or disassemble the software, except and only to the extent that such activity is expressly permitted by applicable law notwithstanding this limitation.

APPLICABLE LAW

This Agreement and any dispute relating to the 'Software' or to this Agreement shall be governed by the laws of the state of New South Wales and the Commonwealth of Australia, without regard to any other country or state choice of law rules. You agree and consent that jurisdiction and proper venue for all claims, actions and proceedings of any kind relating to Wrensoft or the matters in this Agreement shall be exclusively in courts located in NSW, Australia. If any part or provision of this Agreement is held to be unenforceable for any purpose, including but not limited to public policy grounds, then you agree that they remainder of the Agreement shall be fully enforceable as if the unenforced part or provision never existed. There are no third party beneficiaries or any promises, obligations or representations made by Wrensoft herein.

ENTIRE AGREEMENT

This Agreement (including any addendum or amendment to this EULA which is included with the software) constitutes the entire Agreement between the parties with respect to the subject matter herein and supersedes all previous and contemporaneous agreements, proposals and communications, written or oral between you and Wrensoft. Waiver by Wrensoft of any violation of any provision of this Agreement shall not be deemed to waive any further or future violation of the same or any other provision.

EULA0601128ZSE_ENG01

Index

- . -

- .DESC files
 - use offline .desc files when indexing remote files 32
- .gif images 77
- .jpg images 77
- .png images 77
- .result_image 76
- .tiff images 77

- 2 -

- 2 GB files 88

- 4 -

- 4 GB memory 88

- 6 -

- 64-bit 88

- A -

- accent characters 45
- add list of new or updated pages 79
- add start points to existing index 79
- adding a search form to your website 117
- adding folders 24
- adding websites 21
- additional start directories 24
- additional URLs 21
- advanced options 52
- advanced spider URL options 21
- advanced start directory options 24
- advanced start folder options 24
- advantages over other solutions 8
- ampersand 39
- apostrophes 39
- appearance of search page 110
- authentication
 - cookie or session based authentication 42
 - HTTP authentication 42
- automatic index 86
- automatic login 42
- automatically upload files at the end of indexing 59
- auto-run options 85
- avg unique words per page 25
- avg words per page 25

- B -

- bar charts 67
- base URL 24
- beep at the end of indexing 52
- boolean AND (match all search words) 35
- boolean OR (match any search words) 35
- boosting words found in
 - filename 47
 - headings 47
 - meta description 47
 - meta keywords 47
 - title 47
- broken link detection 84
- broken link messages 27

- C -

- cache 32
- camera make 77
- camera model 77
- cascading style sheets 113
- catch files not belonging to a category 60
- categories
 - allow searching in multiple categories 60
 - category description 60
 - category name 60
 - category pattern 60
 - importing categories 60
 - specifying category per file (ZOOMCATEGORY) 60
- cents 64
- change base URLs 39
- changing the base URL 21
- character sets 45
- check thumbnail exists 32
- chronological sort 35
- client-side search engine

- client-side search engine
 - files required 102
 - how it works 101
 - when to use 101
- colon 39
- comma 39
- command-line parameters 85
- compressed files 92
- configuration files 84
- configuration window 28
- configuring zoom 28
- console mode 132
- content density 47
- content filtering 48
- cookies 42
- crawling JavaScript links 32
- CRC signatures 29
- CSS class names 113
- Currency 64
- custom meta field 64
- custom script source path 52
- customizing search form 116
- customizing search page 110
- customizing the recommended links 113
- customizing the search form 111
- customizing the search results 112
- cyrillic 124

- D -

- danish 123
- database 64
- derivative words 45
- development server 39
- diacritic characters 45
- did you mean 35
- disable charset enforcing on search script 52
- display same icon for all files 71
- distribution license 132
- do not show wizard on startup 52, 57
- do not upload search template 59
- dollar sign 39
- dots 39
- downloading messages 27
- drop-down box 60
- Drop-down text 64
- duplicate page detection 29
- duplicate pages 138

- E -

- editing search script 126
- elapsed index time 25
- embedding script 52
- encoding 45
- End User License Agreement 145
- Enormous websites 88
- entering a license key 136
- error messages 27
- e-shop 64
- EULA 145
- european languages 123
- exact match 64
- exact phrase 35
- exact phrases containing skipped words 138
- exit codes 132
- export start points 21
- exporting additional start URLs 125
- extensions 29
- extract and index all files inside ZIP archive 92

- F -

- feature list 9
- file i/o messages 27
- file:// URLs 32
- files downloaded 25
- files filtered 25
- files indexed 25
- files skipped 25
- filtered messages 27
- filtering files by content 48
- flash setting 77
- follow links only 21
- FOLLOW_ONLY 125
- force page as top result 54
- form action 42
- french 123
- FTP
 - automatically upload files at the end of indexing 59
 - do not upload search template 59
 - folder or path on server 59
 - server 59
 - upload with .tmp filenames 59

FTP
 username and password 59

- G -

german 123
 gigabytes 88
 google sitemap 49
 greater than 64
 greater than or equal to 64
 group files by pattern in URL or filename 60

- H -

hash sign 39
 header already defined 52
 hidden parameters 42
 hide all messages 27
 highlight.js 128
 highlighting issues 138
 highlighting options
 jump to match and highlight within documents 37
 words matched in search results 37
 Highlighting within XML results 82
 how zoom works 13
 howtos 120
 hyphens 39

- I -

icons and thumbnails
 associating an image with a particular page 71
 associating icon images with a file type 71
 associating thumbnail images with files 73
 customizing the appearance of icons or thumbnails 76
 enable images 69
 image technical information 77
 identical pages 138
 image indexing 77
 image meta information 77
 ImageInfo plugin 77
 import start points 21
 importing additional start URLs 125
 improving image search accuracy 77
 incremental indexing

add list of new or updated pages 79
 add start points to existing index 79
 command-line parameters for incremental indexing 79
 requirements 79
 updating an existing index 79
 view or delete pages from existing index 79
 index filename only of ZIP archive 92
 index log 51
 index messages 51
 index page and follow all links 21
 index page and follow links 21
 index single page only 21
 index.html issues 138
 INDEX_AND_FOLLOW 125
 INDEX_AND_FOLLOW_ALL 125
 INDEX_ONLY 125
 indexing completed 106
 indexing messages 27
 indexing modes
 offline mode 20
 spider mode 19
 indexing options 39
 indexing status 25
 information messages 27
 initialization messages 27
 inputs and outputs 13
 international foreign languages 123
 international languages 45
 introduction 8
 inventory 64

- J -

japanese 124
 JavaScript links 32
 jump to match and highlight within document 128

- K -

known issues 138

- L -

languages 45
 Large File Support 88
 last-modified date 128

layout 110
 less than 64
 less than or equal to 64
 LFS 88
 License Agreement 145
 license key 136
 ligature characters 45
 limit files for a specific start point 21
 limit files per start point 41
 limit words per file 41
 limitations of javascript 102
 limits 41
 line charts 67
 list the top x searched words 67
 loading indexer settings 84
 login 42
 logout 42
 long words 138

- M -

match all search words 35
 maximum description length 41
 maximum extensions 139
 maximum file size scanned 41
 maximum files to scan 41
 maximum meta keywords 139
 maximum skip pages 139
 maximum skip words 139
 maximum title length 139
 maximum unique words 41
 maximum URL length 139
 maximum word length 139
 maximum zoomwords 139
 meta field 64
 minimize down time 59
 minimum index word length 30
 misspellings 54
 modifying the search form 116
 Money 64
 More button 21, 24
 multiple base URLs 24
 multiple categories 60
 multiple domains 21
 multiple folders 24
 multiple threads 32
 multiple websites 21
 Multi-select 64

- N -

new features 8
 Numeric 64

- O -

offline mode 20
 open all DOC files in new window 29
 open all PDF files in new window 29
 open all plugin files in new window 29
 output directory 24
 overview 8

- P -

page boosting 47
 page skip list 30
 parsing JavaScript links 32
 password 42
 PASV mode 59
 PDF layout problems 138
 pdf plugin options 52
 pie charts 67
 platform 25
 plugin messages 27
 plugins
 configuring 90
 file formats supported 89
 installing 90
 overview 88
 upgrading 91
 using 90
 using custom description .DESC files 91
 plurals 45
 preference to short URLs 47
 previews 73

- R -

recommended links 54
 customizing appearance of recommended links 113
 reload all files 32
 required files 106
 requirements

- requirements
 - for indexing your site 12
 - for performing searches on your site 12
- reset to default messages 27
- results linking 35
- rewrite links 39
- rewrite URLs 39
- robots.txt 32
- russian 124
-
- S -**
-
- same page 138
- saving indexer settings 84
- scan extensions 29
- scheduling
 - add task 86
 - command-line parameters 85
 - edit time 86
 - remote task 86
 - troubleshooting scheduled task 86
- screenshot
 - example search results 15
- script platform 25
- search and replace URLs 39
- Search criteria name 64
- search form appearance 35
- search logging 52
- search page options 35
- search page text 45
- search results layout 37
- search statistics report
 - list the top x searched words 67
 - logging search words 52
 - options 67
 - overview 67
 - searches per day 67
 - searches per month 67
 - searches per week 67
 - top 10 "no result" phrases 67
 - top 10 search phrases 67
- searches per day 67
- searches per month 67
- searches per week 67
- server-side search engine
 - files required 95
 - how it works 95
 - using cgi or php version without a web server 103
 - when to use 95, 103
- session IDs 138
- sessions 42
- Shift-JIS 124
- shop 64
- short URLs 47
- show all index messages 51
- show all messages 27
- shutter speed 77
- single-case languages 45
- single-threaded downloading 32
- sitemap.xml 49
- sitemap_index.xml 49
- sitemaps
 - text file URL list 49
 - XML sitemap 49
- skip directories 30
- skip files 30
- skip options 30
- skip words less than x characters 30
- skipped messages 27
- skipping sections of a page 126
- skipping small image files 77
- sort results by date 35
- spelling suggestions 35
- spider mode 19
- spider options 32
- spider throttling 32
- spidering messages 27
- sponsored links 54
- start directory 24
- start index time 25
- start indexing 25
- start points 21, 24
- start points scanned 25
- start spider URL 21
- stdout 132
- stemmer 45
- stemming 45
- step-by-step wizard 57
- stop indexing 25
- submitting sitemaps to search engines 49
- substring matches for all searches 45
- summary messages 27
- swedish 123
- synonyms 54

system requirements 12

- T -

technical limitations 139
 test server 39
 Text 64
 thread info messages 27
 thumbnail checking 32
 thumbnails 73
 top 10 "no result" phrases 67
 top 10 search phrases 67
 total bytes scanned/downloaded 25
 total words found 25
 translate search page text 45
 translating the search page 124

- U -

UNC addresses 24
 underscores 39
 Unicode 45
 unique words found 25
 unsupported binary formats 138
 unzip 92
 updating an existing index 79
 upload with .tmp filenames and rename when completed 59
 uploading index files 59
 uploading messages 27
 uploading to your website 107
 URL rewrite 39
 urllist.txt 49
 URLs in spider queue 25
 URLs visited by spider 25
 use cookies from Windows and IE 42
 USELIMIT 125
 User-Agent text 52
 username 42
 UTF-8 45

- V -

verbose mode 51
 view or delete pages from existing index 79

- W -

warning messages 27
 what should i do first 14
 what's new 8
 wizard 57
 word join characters 39
 word rules 39
 word skip list 30
 word variations 54
 word weighting 47

- X -

XML sitemap 49
 XML/RSS output
 channel information 82
 configuring 52
 Highlighting within XML results 82
 OpenSearch compatibility 82
 OpenSearch description file 82
 overview 82

- Y -

yahoo sitemap 49

- Z -

ZCFG files 84
 ZIP files 92
 ZLANG file
 selecting your ZLANG file to use 45
 ZLANG files 124
 zoom indexer 19
 ZOOMCATEGORY 60
 ZOOMIMAGE 71
 ZOOMPAGEBOOST 47
 ZOOMRESTART 126
 ZOOMRESTARTFOLLOW 126
 ZOOMSTOP 126
 ZOOMSTOPFOLLOW 126
 ZOOMWORDS 127



Copyright © 2000-2010 Wrensoft