

Sphider-plus Manual

Content

Titel	Page
1. Introduction	2
2. Version and legal info	2
3. Installation	2
3.1 Install from scratch	3
3.2 Upgrade from original Sphider to Sphider-plus	4
3.3 Updating to new Sphider-plus releases	4
4. Settings and customizing	5
5. Indexing options	6
6. Using the indexer from command line	7
7. Keeping pages from being indexed	8
7.1 robots.txt	8
7.2 Must include / must not include string list	8
7.3 Ignoring links	8
7.4 Ignoring parts of a page	8
7.5 Ignored words	8
8. UTF-8 Support and 'Preferred Charset'	9
9. Search modes	11
9.1 Search with wildcards *	11
9.2 Strict search !	11
9.3 Tolerant search	11
9.4 Link search site:	12
10. Chronological order for result listing	12
11. PDF converter for Linux/UNIX systems	13
12. Clean resources during index / re-index.	14
13. Enable real-time output of logging data	14
14. FAQs	16
14.1 UTF-8 support does not work.	16
14.2 Can't search for long words.	16
14.3 Can't search for words with non-Latin characters.	16
14.4 How to bypass the Admin log in.	16
14.5 Links are not followed during Re-index, only main URL is indexed (option 1)	17
14.6 Links are not followed during Re-index, only main URL is indexed (option 2)	17
14.7 How to integrate Sphider's search field into existing pages.	17
14.8 Error message: "Warning: set_time_limit() . . ."	18
14.9 Error message: "Unable to flush table 'addurl' "	18
14.10 Error message: " Access denied; you need the RELOAD privilege. . . "	18
14.11 Fatal error: "Allowed memory size of xxx bytes exhausted (tried to allocate yyy bytes)"	18
15. Change log	19
15.1 Version 1.0	19
15.2 Version 1.0.a	22
15.3 Version 1.1	22
15.4 Version 1.2	23
15.5 Version 1.3	24
15.6 Version 1.3.a	24
15.7 Version 1.4	25
15.8 Version 1.5	27
15.9 Version 1.6	29

Last update: Saturday, September 06, 2008

1. Introduction

Sphider-plus is a search engine based on the original Sphider scripts created by Ando Saabas (www.sphider.eu).

In front of original Sphider additional mods, functions, template designs and debugging have been performed. For details about all changes, please notice the chapter [Change Log](#)

Sphider-plus requires PHP 5.

The names of Sphider-plus folders and scripts are often the same like those of original Sphider. But the scripts are not interchangeable between Sphider and Sphider-plus.

Some messages have been added in the language files. You are invited to translate your native language and then to share the files with the community. Also mods, improvements and of course bug fixes are very welcome for future releases of Sphider-plus.

2. Version and legal info

Name: Sphider-plus
Version: 1.6
Created: 06 September 2008

Based on original Sphider version 1.3.4 released 29.04.2008 by Ando Saabas <http://www.sphider.eu>

This program is licensed under the GNU GPL by Rolf Kellner [Tec] [sphider\(a\)tibk-kellner.de](mailto:sphider(a)tibk-kellner.de)

Original Sphider GNU GPL licence by Ando Saabas [ando\(a\)tcs.ioc.ee](mailto:ando(a)tcs.ioc.ee)

Updates and support for Sphider-plus are available at:

http://sourceforge.net/project/showfiles.php?group_id=214642
[sphider\(a\)tibk-kellner.de](mailto:sphider(a)tibk-kellner.de)

If you like Sphider-plus and want to promote further development, your donation at PAYPAL account

r.k@ibk-kellner.de

is highly appreciated. Also the footer of search.php and admin.php should only removed if you contribute (donate) to the Sphider-plus project. Note that this is a requirement under the GPL licensing agreement, which Sphider-plus acknowledges.
Thank you very much.

3. Installation

There are three ways to install Sphider-plus:

- Install from scratch
- Upgrade from original Sphider to Sphider-plus
- Update to new Sphider-plus releases

In order to get correct function of Sphider-plus, please follow the instructions as described below.

3.1 *Install from scratch*

1. Unpack the files, and copy them to the server, for example to:
C:\programs\xampp\htdocs\public\sphider-plus\
 - a) at command prompt type (to log into MySQL):
mysql -u <your username> -p
Enter your password when prompted.
 - b) in MySQL, type:
CREATE DATABASE sphider-plus_db;

Of course you can use some other name for database instead of sphider-plus_db.

 - c) Use exit to exit MySQL.
For more information on how to create a database and give/get the necessary permissions, check MySQL.com
3. In .../sphider-plus/settings/ directory, edit the file database.php and change:
\$database
\$mysql_user
\$mysql_password
\$mysql_host
to your personal requirements to correct values (if you don't know what \$mysql_host should be, it should probably stay as it is - 'localhost').
4. Open the file .../admin/install_all.php in your browser. This script will create the tables necessary for Sphider-plus to operate.
(http://localhost/public/sphider-plus/admin/install_all.php)
5. In admin directory, edit the file auth.php to change the
- administrator user name
and
- password
(Default values are 'admin' and 'admin').
6. Open admin.php in your browser and start indexing
(<http://localhost/public/sphider-plus/admin/admin.php>)
Installing from scratch your site table is empty. You will be asked to add any URL to be indexed.
You should enter something like:
<http://www.abcd.de>
7. search.php is the default search script
(<http://localhost/public/sphider-plus/search.php>)

3.2 Upgrade from original Sphider to Sphider-plus

1. Create a backup of your current Sphider database and all original Sphider files.
2. Unpack the downloaded files, and copy them into a new folder for your Sphider-plus installation, for example to:
C:\programs\xampp\htdocs\public\sphider-plus\
3. In .../sphider-plus/settings/ directory, edit the file database.php and change:
\$database
\$mysql_user
\$mysql_password
\$mysql_host
to those values you used up to now for the database together with original Sphider.
4. In admin directory, edit the file auth.php to change the
- administrator user name
and
- password
(default values are 'admin' and 'admin').
5. In front of original Sphider, Sphider-plus needs some additional tables in your database. They will be created by calling the file .../admin/install_sphider-plus.php in your browser:
(http://localhost/public/sphider-plus/admin/install_sphider-plus.php)
6. Open admin.php in your browser and define the settings for your specific application.
(http://localhost/public/sphider-plus/admin/admin.php)
7. search.php is the default search script
(http://localhost/public/sphider-plus/search.php)

Additional info:

When upgrading from original Sphider, after defining the Admin settings, it is absolutely necessary to run an 'Erase & Re-index' procedure.

Also you should keep in mind that for full Sphider-plus support your database needs to be created with utf-8 collation.

3.3 Updating to new Sphider-plus releases

If you already use Sphider-plus and want to update to the newest release, you may copy only the new scripts over the existing you used hitherto. In order to find out, which scripts are involved for the different releases, please notice the chapter "[Change log](#)". Some releases of Spider-plus also require additional database tables or just some new rows/columns for an existing table. The corresponding change log informs about necessary procedures.

But Sphider-plus takes care about its database. If you press a 'Save settings' button in Admin settings and Sphider-plus recognizes a database inconsistency, a message will be displayed. Also you will be asked to run the according installation script. Something like:

Please run the .../admin/install_bestclick.php file.

4. Settings and customizing

If you want to change settings, behavior and design of Sphider-plus, you can do so by means of the Admin interface. There is a wide range of settings foreseen for Sphider-plus. Separated into different submenus like:

- Sites
 - Add Site
 - Index only the new
 - Re-index all
 - Erase & Re-index
 - Approve sites
 - Banned domains
- Categories
 - Add, edit, delete
 - Create new subcategory under
- Index
 - Basic indexing options
 - Advanced options
- Clean
 - Clean keywords not associated with any link
 - Clean links not associated with any site
 - Clear Temp table
 - Clear Search log
 - Clear Spider log, separate and bulk delete
- Settings
 - General Settings
 - Index Log Settings
 - Spider Settings
 - Search Settings
 - Suggest Options
 - Page Indexing Weights
- Statistics
 - Top keywords
 - Larges pages
 - Most popular searches
 - Search log
 - Spidering logs
 - Server Info
- Database
 - Backup
 - Restore
 - Optimize

Three different template designs are prepared for Sphider-plus, which you may select in submenu 'Settings'. If the layout does not fit the design of your site (which is normal), you may create new template designs and modify the appropriate file `.../templates/My_template/thisstyle.css`

5. Indexing options

As part of the Admin Site settings you may select the following options:

Full: Indexing continues until there are no further (permitted) links to follow.

To depth:

Indexes to a given depth, where depth means how many "clicks" away the page can be from the starting page. Depth 0 means that only the starting page is indexed, depth 1 indexes the starting page and all the pages linked from it etc.

Re-index:

By selecting this mod, indexing is forced even if the page already has been indexed. Re-index only detects changes of the pages to be re-indexed. Modifications in Admin settings are not recognized.

Erase & Re-index:

By selecting this mod, indexing is forced even if the page already has been indexed.

Additionally this mod will

Clear Spider-plus database before the re-index process. It will leave the following untouched:

- Categories
- Query log
- Sites and all options: spider-depth, last indexed, can leave domain, title, description, Url Must include, Url must Not include.

If settings have been modified in Admin section this mod should be selected to update the database.

Spider can leave domain:

By default, Spider never leaves a given domain, so that links from domain.com pointing to domain2.com are not followed. By checking this option Spider can leave the domain, however in this case its highly advisable to define proper must include / must not include string lists to prevent the spider from going too far. This option must be activated, if a htaccess. file is used for redirect directives.

Must include / must not include: See below for an [explanation](#).

6. Using the indexer from command line

It is possible to spider web pages from the command line, using the syntax:
`php spider.php <options>`

where <options> are:

<code>-all</code>	Reindex everything in the database
<code>-u <url></code>	Set the url to index
<code>-f</code>	Set indexing depth to full (unlimited depth)
<code>-d <num></code>	Set indexing depth to <num>
<code>-l</code>	Allow spider to leave the initial domain
<code>-r</code>	Set spider to reindex a site
<code>-m <string></code>	Set the string(s) that an url must include (use \n as a delimiter between multiple strings)
<code>-n <string></code>	Set the string(s) that an url must not include (use \n as a delimiter between multiple strings)

For example, for spidering and indexing `http://www.domain.com/test.html` to depth 2, use:
`php spider.php -u http://www.domain.com/test.html -d 2`

If you want to reindex the same url, use:
`php spider.php -u http://www.domain.com/test.html -r`

7. Keeping pages from being indexed

7.1 *robots.txt*

The most common way to prevent pages from being indexed is using the robots.txt standard, by either putting a robots.txt file into the root directory of the server, or adding the necessary meta tags into the page headers.

7.2 *Must include / must not include string list*

A powerful option Sphider-plus supports is defining a 'Must include / Must not include' string list for a site (to be found in Sites / Options / Edit). Any Url containing a string in the 'Url must Not include' list is ignored. Any Url that does not contain any string in the 'Url Must include' list is likewise ignored.

All strings in the string list should be separated by a newline (Enter). For example, to prevent a forum in your site from being indexed, you might add `www.yoursite.com/forum` to the 'Url must Not include' list. This means that all Urls containing the string will be ignored and wont be indexed. Using Perl style regular expressions instead of literal strings is also supported. Every string starting with a '*' in front is considered as a regular expression, so that `*[a]+'` denotes a string with one or more a in it.

7.3 *Ignoring links*

Sphider-plus respect `rel="nofollow"` attribute in `<a href...>` tags, so for example the link `foo.html` in `` is ignored.

7.4 *Ignoring parts of a page*

Sphider-plus includes an option to exclude parts of pages from being indexed. This can for example be used to prevent search result flooding when certain keywords appear on certain part in most pages (like a header, footer or a menu).

Any part of a page between

`<!--sphider_noindex-->` and `<!--/sphider_noindex-->`

tags is not indexed, however links in it are followed.

7.5 *Ignored words*

The list of common words that are not indexed must be placed in `.../include/common.txt`.

The list of file types that are not checked for indexing are given in `.../admin/ext.txt`.

8. UTF-8 Support and 'Preferred Charset'

Starting with version 1.2, Sphider-plus provides Unicode assistance. The following 63 charset are supported and will be converted into UTF-8 Unicode:

WINDOWS

windows-1250 - Central Europe
windows-1251 - Cyrillic
windows-1252 - Latin I
windows-1253 - Greek
windows-1254 - Turkish
windows-1255 - Hebrew
windows-1256 - Arabic
windows-1257 - Baltic
windows-1258 - Viet Nam
cp874 - Thai - this file is also for DOS

DOS

cp437 - Latin US
cp737 - Greek
cp775 - BaltRim
cp850 - Latin1
cp852 - Latin2
cp855 - Cyrylic
cp857 - Turkish
cp860 - Portuguese
cp861 - Iceland
cp862 - Hebrew
cp863 - Canada
cp864 - Arabic
cp865 - Nordic
cp866 - Cyrylic Russian (this is the one,
used in IE "Cyrillic (DOS)")
cp869 - Greek2

MAC (Apple)

x-mac-cyrillic
x-mac-greek
x-mac-icelandic
x-mac-ce
x-mac-roman

ISO

iso-8859-1
iso-8859-2
iso-8859-3
iso-8859-4
iso-8859-5
iso-8859-6
iso-8859-7
iso-8859-8
iso-8859-9
iso-8859-10
iso-8859-11
iso-8859-12
iso-8859-13
iso-8859-14
iso-8859-15
iso-8859-16

MISCELLANEOUS

gsm0338 (ETSI GSM 03.38)
cp037
cp424
cp500
cp856
cp875
cp1006
cp1026
koi8-r (Cyrillic)
koi8-u (Cyrillic Ukrainian)
nextstep
us-ascii
us-ascii-quotes

DSP implementation for NeXT

stdenc
symbol
zdingbat

And specially for old Polish programs
mazovia

It is only a small checkbox in Admin settings. But in consequence of your selection, the impact will be dramatically.

First of all: if you select this option, the complete text and all keywords to be indexed, will have to be translated into Unicode. Consequence is an increase of time for indexing.

As also suggested by Yiannes [pikos], I integrated three steps to realize this procedure:

1. Detect charset of site, page or file that's content has to be translated.
This information is normally presented as part of the HTML header.
If not available, or for files without header like .doc, .rtf, .pdf, .xls and .ppt files, the 'Preferred charset' (as defined in Admin settings) will be used to translate the file into Unicode.
In other words: you can't convert DOCs, PDFs etc. that are coded in 'foreign' charset.
Only those with your personal charset will be converted correctly.
2. By means of the PHP function 'iconv()' content and keywords will be converted into UTF-8.
This step is successful, if the required charset (for the content to be translated) is part of your local PHP installation. In order to find out which charset are available in your installation, notice the files in server folder:
.... ../apache/bin/iconv/
Depending of the installation you will find about 200 charset files that iconv() is able to translate into UTF-8
3. If the PHP function fails, finally the class 'ConvertCharset' is invoked. This class, originally designed by Mikolaj Jedrzejak, enables translation for a lot of charset.
But it takes more time than the compiled PHP function 'iconv()'.

As result of this translation, you are also enabled to search for words that contain non-Latin characters.

Some additional remarks:

It is not enough to select the checkbox 'Convert all into UTF-8 charset'. You are obliged to do a fresh index or 'Erase & Re-index'. Also if you de-select this option, you must update your database again. Otherwise content of database and search query will not match, as different character coding is used.

Even if you didn't notice immediately: with release date of Sphider-plus version 1.2 all files in folder ../languages/ have been converted to Unicode. (By the way, I'm still hoping for your translations.)

In order to enable translation of all entities into UTF-8, the Unicode requires also upper case characters. So, if you enable the UTF-8 option, your query 'html' will not deliver results for sites and files that contain the string 'HTML'. Both keywords are stored separately. This behavior is in contrast to former versions of Sphider and Sphider-plus and should be remembered.

But I also improved the 'Did you mean' option. Let's assume there are no sites with 'html' but some with 'HTML'. If your query is 'html', Sphider-plus will not capitulate without results. You will be asked 'Did you mean HTML?'

Starting with version 1.6 Sphider-plus offers the additional option:

'Enable distinct results for upper- and lowercase queries'

If enabled in Admin settings, everything remains as described above. But if this checkbox is unchecked, result listing will deliver all results independent of the query input. HTML, html or even hTmL will produce the same (all) results.

The checkbox for this option is placed with full intention in section 'Spider settings', as activating and also deactivating requires always an 'Erase & Re-index' procedure.

9. Search modes

Beside the original Sphider search queries like:

- Search for a single word
- AND and OR search
- Search for a phrase

Sphider-plus offers 4 additional modes to enter queries:

- Search with wildcard
- Strict search
- Tolerant search
- Link search

Wildcard, strict and tolerant search modes are available only for single query word input.

9.1 Search with wildcards *

This mod enhances the Sphider-plus capabilities to search also for parts of a word. The mod is invoked by entering a * as wildcard for the unknown part of the search query.

Wildcards could be used like:

- *searchme
- *searchall*
- *search*more*

Depending on Sphiders keyword table, a lot more results may appear. In order not to confuse the user, the printout of relevance (weight/hits) is suppressed in result listing.

9.2 Strict search !

This variant is invoked by entering a ! as first character of the search query. If you search for '!plus' only results for the word 'plus' will be presented in the result pages. No results for words that contain 'spider-plus' or 'spiderplustec' will be shown. This is the reverse function of 'Search for part of a word by means of * wildcards'. Strict search only indicates results in the text part of the indexed pages. If utf-8 support is enabled and 'Distinct results for upper- and lowercase queries' is also activated, 'Strict search' will respond case sensitive.

9.3 Tolerant search

This mod enables a tolerant search for Sphider. Selectable in Search-form like AND, OR and Phrase Search a new item "Tolerant Search" is added.

If this item is selected, query input "perdida" will also deliver results for all sites that contain the word "pérdida". Inverse function is also implemented: "pérdida" input will deliver all results for "perdida".

If enabled, this mod equalizes search input for e=é=è=ê and all the other vowels like: ä=a=à=â , ü=u , o=ö etc. The upper-case letters like Ä=A are also taken into account. Tolerant search overwrites the 'Distinct results for upper- and lowercase queries' setting and will mark all results.

Natively developed to deliver most possible results for queries with entities and accents and also to simplify user input, this mod also delivers results that are "like" the query input. So something as the "Did you mean" facility is already integral part of this search method.

9.4 Link search site:

Invoked by starting the query input with ' site: ', the user is enabled to search for all pages of a domain. It is not necessary to enter the full domain address. For example if you enter 'site:sphider.eu' you will get a list of all pages that belong to the domain <http://www.sphider.eu>

If the search query is part of more than one domain address in Sphiders site table, a list of these domains will be presented as intermediate result. If you then click on the desired domain of this list, all links (pages) of this domain will be presented as final result listing.

10. Chronological order for result listing

Sphider-plus offers 5 methods how to sort the results:

- By relevance (weight/hits)
- Main URLs (domains) on top
- By URL names and then weight
- Only top 2 per URL (like Google)
- Most Popular Links on top

The according selection is placed in Admin settings section 'Page index weight'. This was done, because the method **'Main URLs (domains) on top'** changes the weight of words in all involved pages. The normal relationship (weight) as defined for:

- Word in web page Title tag
- Word in the Domain name
- Word in the Path name
- Word in web page Keywords tag
- Word in full text

remains valid and is calculated as before. But additionally all words in main URLs are weighted with an additional factor. In other words:

The weight of each word found in a page that is accepted as main URL, will become an increased weight. A multiplier performs this. This multiplier may be reduced or increased by modifying 'Multiplier for words in main URLs'. So you may influence the weight of main URLs in front of all other pages.

Per default this multiplier is set to 5. In order to keep the results of the main URL on top, it might become necessary to increase this factor slightly. An according input field called 'Multiplier for words in main URLs (domains)' is also added to the same section of Admin settings. This multiplier is valid only for the method 'Main URLs (domains) on top'.

In order to be accepted as 'Main URL', the address must fulfill the following conditions:

1. The domain may include at maximum 2 dots. So valid are:
<http://www.abc.de> and <http://forum.abc.de>
But sub domains like <http://www.forum.abc.de> will be ignored.
2. As file path are only accepted: index.php, index.html, index.htm or a blank path.
If additional file names should become valid, open the file `.../admin/spiderfuncs.php` and search for: `$act_path = str_replace ('index.htm', '', $act_path);`
Beyond this row include a row like above, but holding the additional file name.

The current selection of 'order for result listing' is visible for the users as additional headline on all result pages. If not desired, this option can be de-selected in Admin setting: 'Show mode of chronological order for result listing as headline'.

In order not to confuse the user, for the 3 methods

- By URL names and then weight
- Only top 2 per URL (like Google)
- Most Popular Links on top

the output of relevance (weight/hits) is suppressed in result listing.

Additional reminder: As weight of each keyword is already defined during index and re-index, any modification of weight or 'Order for result listing' requires an 'Erase & Re-index'.

For the '**Most Popular Links on top**' method, Sphider-plus uses the before learned link acceptance. If a user leaves the result listing by clicking on any of the offered links, Sphider-plus will memorize this decision. The user is temporary redirected to the script `.../include/click_counter.php`, which stores the users link decision, last query, time and date before leading the user to the real destination.

This link specific 'best click' counter is used as teach-in to define the chronological order of result listing. In order to prevent promoted clicks on a specific link, there is a delay timer before the next user click will be accepted. To be set in Admin /Settings/ Index Log Settings, the setting defines the idle-time in seconds.

If there are more results as rated links, the rest of the result listing will be presented by relevance (weight), using the weighting of the last index / re-index.

As the 'Most Popular Links On Top' item overwrites all other order of result listing, it might be selected without re-index procedure.

For the method of result ordering '**By relevance (weight/hits)**' the weight is calculated as described above in this chapter. Situation changes, if in Admin settings the item 'Instead of weighting %, show count of query hits in full text' is activated. Now only the hits in full text are used to calculate the order of result listing. Keyword hits in URL name, path, title tag etc. are not taken into consideration.

11. PDF converter for Linux/UNIX systems

Starting with version 1.5 Sphider-plus includes one PDF converter for Windows systems and another converter for LINUX/UNIX systems. The Windows converter is ready for use, but the Linux version needs your attention. So, before using the Linux converter you need to do the following steps:

1. Identify the physical path of your web site. if available, Admin / Statistics / Server / PDF-converter should present this info. Otherwise your hoster should provide this information anywhere.
2. Take this full path as above and use simple slashes (not double backslashes) and include your individual path to the file `.../admin/converter/pdftotext`
Second line of that file begins with a slash and ends with the minus sign.
3. Set permissions of both `pdftotext` and `pdftotext.script` to 755 or 777 (whatever needed to run correctly).
4. Set permissions of the converter dir to 777! Otherwise indexing fails because of `pdftotext` is unable to write a temp file needed!
5. Adapt the variable `$pdftotext_path` in file `.../settings/conf.php` to your personal needs:
`$pdftotext_path = '/PATH/TO/YOUR/WEB/DOWN/TO/converter/pdftotext';`
Do this modification with closed Admin browser window.

The DOC, RTF, PTT and XLS converter are not available for LINUX/UNIX systems.

12. Clean resources during index / re-index.

In order to prevent performance problems and memory overflow for large amount of URLs, Sphider-plus may clean unused resources during index / re-index. Selectable in Admin settings, this item periodical will:

- Free memory that is allocated to unused MySQL resources.
- Unset PHP variables, which are no longer required.

As this clearing work is done several times during index / re-index of every URL, additional capacity is required. Consequently overall indexing time will increase. So this item should be selected only for huge amount of URLs. Depending on

- Memory size allocated to PHP
- Total number of URLs
- Number of internal and external links
- Size of text to be indexed for each page
- CPU clock rate
- System RAM

there will be an individual limit when to enable this feature. Following the discussion on the Sphider forum this feature should be activated only if > 100 sites are to be indexed, or when Sphider-plus dies a silent death during index procedure, not indexing any more sites.

Please take notice of the FAQ chapter:

Error message: "Unable to flush table 'addurl' "

and

Error message: " Access denied; you need the RELOAD privilege. . . "

13. Enable real-time output of logging data

Up to version 1.5 of Sphider-plus, during index / re-index there was no printout available because:

- Several servers, especially on Win32, buffer the output from the script until it terminates before transmitting the results to the browser.
- Server modules for Apache do buffering of their own that will cause flush() to not result in data being sent immediately to the client.
- Browser may buffer its input before displaying it. Netscape, for example, buffers text until it receives an end-of-line or the beginning of a tag, and it won't render tables until the </table> tag of the outermost table is seen.
- Some versions of Microsoft Internet Explorer only start to display the page after they have received 256 bytes of output.

As progress was not presented during index / re-index procedure, waiting for results became a pain in the neck.

Selectable in Admin setting together with the update interval (1 - 10 seconds), AJAX technology was the approach to realize this feature.

Pressing one of the 'Start index / re-index' buttons, three additional scripts are involved.

(onclick="window.open('real_log.php')"))

.../admin/real_log.php By opening a new browser window / tab, this script takes over to display latest logging data. Requesting fresh data from the JavaScript file 'real_ping.js', all new logging data will always be placed into <div id='realLogContainer' /> So, better not to press the 'Reload' button of your browser. The current <div /> might be already empty.

- .../admin/real_ping.js Script that transfers requests from HTML client to PHP server script and vice versa. Handling refresh for real-time logging during index and re-index procedure by means of asynchronous requests (AJAX) to the server.
- .../admin/real_get.php This script delivers 'refresh rate' and latest 'logging data', requested from the JavaScript file 'real_ping.js'. Also performs the reset of the 'real_log' table in Sphiders database.

Latest logging data is delivered by the .../admin/messages.php script that, besides writing into the normal log file, feeds the table 'real_log' in Sphiders database. This is the buffer for latest logging data.

Prerequisites are the enabled 'Log spidering dates' and 'Log file format = HTML'. When activating the real-time output, both pre-conditions are automatically selected.

14. FAQs

14.1 UTF-8 support does not work.

In order to enable correct utf-8 support, charset of the search page and also of the result page must be set to utf-8. So, if you integrate (embed) Sphider-plus into your site design, the html header of your files should include something like:

```
<meta http-equiv="content-type" content="text/html; charset=utf-8">
```

14.2 Can't search for long words.

Current length of words to be indexed is set to 30 character. This might be not enough for several applications. In order to increase this value to 255, the database table \$mysql_table_prefix 'keywords' must be manually modified. Set keyword varchar(255) not null

Also in ../admin/spiderfuncs.php the row

```
if (strlen($word)<= 30) {
```

must be modified to

```
if (strlen($word)<= 255) {
```

Please remember to do an 'Erase & Re-index' after you made this modification.

Starting with version 1.4, the database installation and spider scripts were changed to the new values.

Example:

The Gregorian word

᠓ᠪᠣᠨᠵᠢᠨᠭᠣᠷᠭᠢᠨ (13 letters)

is represented in Sphider-plus database as:

áf>áfœáf~áf~áf•áfœáf"áfšáf áf•áf áfœáf~ (39 letter string)

14.3 Can't search for words with non-Latin characters.

in Admin settings the two checkboxes:

- Convert all into UTF-8 charset
- Enable distinct results for upper- and lower-case queries

must be selected. Even if the site to be indexed is UTF-8 coded. In order to get the query input UTF-8 coded and also to present the results correctly, Sphider-plus needs these two checkboxes to be activated. If the site to be indexed is not UTF-8 coded, the limitation for words to be queried is defined by the 63 charsets currently available to be transferred to UTF-8. For details, please take notice of the chapter:

[UTF-8 Support and 'Preferred Charset'](#)

14.4 How to bypass the Admin log in.

For Intranet applications and during debugging it might be more comfortable to bypass the Admin authorization. There are two possibilities:

1. This version still shows the **Log In** page as warning that you now enter into the Admin section, but you just have to click on the Login button.


```
In ../admin/auth.php set
    $admin = "";
    $admin_pw = "";
```

2. This version removes the **Log In** page totally:

```
Rename the file ../admin/auth.php into auth_backup.php
Rename the file ../admin/auth_bypass.php into auth.php
```

14.5 Links are not followed during Re-index, only main URL is indexed (option 1).

It is not a bug, it is a feature. If 'Follow sitemap.xml' is activated in Admin settings, links will only be followed if:

- 'last modified' date in sitemap.xml is newer than Sphiders 'last indexed' date.
- New link that is not yet known in Sphiders link table.

The main URL will always be indexed, because status and content of the sitemap file is required for further decision what necessarily has to be indexed. Because only relevant pages will be indexed, this approach significantly reduces the time required for index and re-index.

14.6 Links are not followed during Re-index, only main URL is indexed (option 2).

If you use a htaccess. file on your server in order to redirect requests, or to 'produce' seo friendly link names, you must enable the checkbox 'Spider can leave domain' in Admin/Sites/Options/Edit/. Otherwise Spider will not follow the redirect directive of your .htaccess. file.

14.7 How to integrate Sphider's search field into existing pages.

Add the following code at the according position into the HTML code of your page and personalize the path_to_sphider-plus address relativ to the HTML code:

```
<form action="/path_to_sphider-plus/search.php" method="get">
<table border="2" width="150" cellpadding="0" cellspacing="2">
<tr>
<td align="center"><input type="text" name="query" size="30" value="" /></td>
<td align="center"><input type="submit" value="Search" />
<input type="hidden" name="search" value="1" /></td>
</tr>
</table>
</form>
```

This simple example does not support all facilities of Sphider-plus. It is foreseen only as first step into your personal adaption. For example if you add

```
<input type="hidden" name="mark" value="markyellow" />
```

the found keywords will be marked yellow.

More details and examples how to integrate Sphider-plus into existing pages may be found on the Sphider forum. For example at:

<http://www.sphider.eu/forum/read.php?2,4505>

14.8 Error message: "Warning: set_time_limit() . . . "

Sphider does not work if the server is in 'safe' mode. That server setting must be disabled in the PHP initialisation file (e.g.: ../apache/bin/php.ini).

```
safe_mode = Off
```

The current state is shown in Admin / Statistics / Server Info / php.ini file key: safe_mode

Before modifying this value, stop your server and afterwards restart the server again.

14.9 Error message: "Unable to flush table 'addurl' "

Sphider has not enough privileges to close the tables of your database. Sphider needs the privilege 'Reload' to perform the flush instruction (MySQL-Manual chapter 13.5.5.2). Please check your database installation, grant enough privileges to Sphider and shut down other scripts that could use the Sphider database.

If you don't succeed with these fundamentals because you use a shared hosting server, open the file
../admin/db_backup.php
and delete the row
mysql_query("FLUSH TABLE \$row[0]") or die("Unable to flush table \$row[0].");

Also open the file
../admin/spiderfuncs.php
and delete the row
mysql_query("FLUSH QUERY CACHE");

Please keep in mind that by deleting these rows you will loose parts of the 'Optimize database' and 'Clean resources during index/re-index' functions.

14.10 Error message: " Access denied; you need the RELOAD privilege. . . "

The same problem as error message: "Unable to flush table 'addurl' " This time your server sends the error message. Sphider has not enough privileges to flush the tables of your database. Sphider needs the privilege 'Reload' to perform the mysql flush instruction. For more details see chapter above.

14.11 Fatal error: "Allowed memory size of xxx bytes exhausted (tried to allocate yyy bytes)"

This is a limitation of your server that does not allow PHP to allocate enough memory. In order to prevent this error message, increase the memory size in the PHP initialisation file (e.g.: ../apache/bin/php.ini)

```
memory_limit = 64M
```

The currently allocated memory size is shown in Admin / Statistics / Server Info / php.ini file
key: memory_limit

Before modifying this value, stop your server and afterwards restart the server again.

15. Change log

15.1 Version 1.0

Release date: February 15, 2008

Based on the original Sphider v.1.3.4.a by Ando Saabas, the following items are modified:

Define min. relevance level (weight %) for results to be presented at result pages.

To be defined in Admin settings.

Enable user suggestion for new Url to become part of Sphider-plus database (addurl by user).

- To be activated in Admin settings, the user is enabled to suggest sites.
- If enabled, a link at the footer of the result page leads to the suggestion form.
- The user will have to fulfil 'Url', 'Title', 'Description' and 'Dispatchers e-mail account'.
- Checked for valid input, DNS availability and MX-RR validation of dispatchers account.
Suggested Url will be stored in the Sphider-plus database until Admin decision.
- Suggested sites are presented in Admin submenu 'Approve sites' so that the admin may decide to
 - accept
 - reject
 - bann
- Result of decision will be mailed to the dispatcher (if selected in Admin settings).
- Included is also the submenu 'Banned domains' to refuse all sites not welcome for this search-engine.

Create a sitemap during index/re-index.

- Compatible with <http://www.sitemaps.org/schemas/sitemap/0.9>
this module automatically creates a sitemap.xml file.
- In Admin settings the folder name for the sitemaps can be defined.
- The xml files will be individually named like 'sitemap_www.abc.de.xml'
- When running a 'Re-index', 'Re-index all' or 'Erase & Re-index'
existing sitemaps will be overwritten with the actual data set.

For index/re-index follow sitemap.xml (to be activated in Admin settings).

If available Sphider-plus will use the sitemap to follow all links of that domain.

This increases significant the speed for index and re-index.

The mod will also force Sphider-plus to re-index only links that are:

- New and not yet known in Sphiders link table
and
- Links whose 'last modified' date is newer than Sphider's 'last indexed' date.

Search for part of a word by means of * wildcards.

This mod enhances the Sphider-plus capabilities to search also for parts of a word.

Invoke this mod by entering a * as first character of your search query.

You may use * wildcards like:

*searchme
searchall
*search*more*

Search !strictly for the search query.

Invoke this variant by entering a ! as first character of your search query.

If you search for '!plus' only results for the word 'plus' will be presented in the result pages.

No results for words that contain 'spider-plus' or 'spiderplustec' will be shown.

This is the reverse function of 'Search for part of a word by means of * wildcards'

Search for all pages of a site.

This utility searches for all that pages, which belong to a domain.

Initialize your search query with 'site:' followed by the domain you want to check.

Also parts of domain names like 'site:www.abc.de' or 'site:abc.de' are valid search queries.

The mod searches for all links in Sphider's link-table but not in the stored keywords.
The search output has the same look and feel as usual in Sphider-plus search results.

Enabled search for dates like 2008-09-06, 06/09/2008 or 06.09.2008

Enabled suggestion also for search queries that containing upper case characters.

Automatically adapt Sphider's dialog to user language.

This mod detects the language of visitors client and selects the according language from Sphider's language folder. If not available, Sphider will use the language as defined in Admin settings.

Auto-detection may be enabled by checkbox in Admin settings

Show 'Most popular searches' table at the bottom of result pages.

Selectable in Admin settings, the most popular queries are presented on the bottom of each result page.
Count of rows for 'Most popular searches' is also to be defined in Admin settings.

Warning message if search string is only found in Url or <title> tag.

If the search string will be found only in title or Url, but not in the HTML body or meta tags, there is no short description for that Url with no possibility to highlight the search string.

A warning message will be displayed instead: "Search string was found only in page title or Url."

This mod is Admin selectable.

Index only new sites.

Additional item in Admin Sites submenu for bulk indexing of all the new sites that were added since last index/re-index.

Erase & Re-index.

Additional item in Admin Sites submenu that will clear the database and perform a re-index.

Clear database done before the re-index will leave the following untouched:

- Categories
- Query log
- Sites and all options: spider-depth, last indexed, can leave domain, title, description, url must include, url must not include.

Limit max. link count to be indexed for each Url.

In Admin settings the count of links to be followed per Url is selectable.

Will be followed by:

- Index
- Index only the new

Perform a link-check instead of re-index.

Selectable in Admin settings, a fast running link-check can be performed.

Unreachable links are automatically deleted from Sphiders database.

Define max. length of title presented in result pages.

An additional input field in Admin "Search Settings" is presented for Admin determination.

Dynamic adaptation of <title> and <h1> tags.

In order to create an individual title for the result pages,

a new input field in Admin settings 'Search Settings' is presented.

Additionally the result page <title> in HTML-header is provided with

- User defined title
- Category (if selected)
- Search query
- Page number of results

New Admin Sites Option menu design with additional utilities.

- Based on the XHTML valid Admin by Peter__LT

3 new template designs selectable in Admin settings

- Based on preparatory work by Peter__LT

The template folder contain only those files that are responsible for the design

Additional Admin Sites submenu: List all pages that belong to the selected site.

To be found in Sites / Options / Pages a list is shown with:

- Page Url
- Last indexed date
- Page size

Validate all user input for security acceptance.

All entries are checked

- Delete quotes
- Place backslash in front of special characters
- Shell commands, XSS attacks and SQL injections are blocked

Additional .htaccess security file.

Prepared for:

- Prevent listing of folder content (files)
- Redirect client queries to search.php
- Prevent delivery of internal files

Sort Admin's Site table in alphabetic order.

Selectable by checkbox, the table is presented in alphabetic order or by index date.

Export all current Url's from Admin section.

A file 'url.txt' will be created with all existing Url's in folder .../admin/urls/

Import url.txt file from folder .../admin/urls/

The content of file 'url.txt' will be copied into Sphider-plus database.

Existing Url's will be lost and overwritten.

Following rules are valid for the url.txt file:

- Url's must be in format: url,spider-depth,category
like:
http://www.abc.de
http://www.abc.de,2
http://www.abc.de,-1,Info
http://www.abc.de,3,Funny things
- Rows must be separated with 'LF'
- Url, spider-depth and category must be separates by commas.
- If you don't specify spider-depth it is automatically set to '-1'.
- Also category is optional. If not specified the new site will be stored without category.
- Not specifying spider-depth but category requires: url,,category-name

Delete Spider log.

Spider log files now can be deleted separately or as bulk delete.

Added in the submenus:

- Admin / Clean / Clean Spider log
- Admin / Statistics /Spider logs

Search in categories: Four bugs fixed.

The following items are modified for proper function of Sphider's Category search:

- The 'Search' button now also sends the variable \$catid to the search script.
- Selecting 'Next' or the other page selections (on bottom of the result page), now transfers also the variable \$catid and \$category to the search script.
- The check boxes 'Search only in category . . ' and 'All sites' are no longer pre-selected.
So, once selected 'Search only in category . . ', you may now select search result page 2, 3, 'Next' and 'Previous' together with the category search.
- If 'Search only in category . . ' is selected, an additional headline is presented.
So the user is informed about the actual situation.

Database Backup and Restore. Bug fix by re-writing the complete Database Management.

Before backup, the 'Optimize Database' function is automatically performed.

Separated folders for each backup task.

Backups now are stored in individual files for each table.
Backup utility selectable for: 'Structure only' or 'structure plus data'
Unlimited file size for restore function is ensured.
Backup files compatible to phpMyAdmin.

Optimize Database. Bug fix by re-writing the complete Database Management.

Links that do not contain page name are now correctly followed (Bug fix by BenRosey)
Original Sphider does not except links like link text
Thanks to the bug fix of BenRosey Sphider-plus follows correctly.

Links that do not contain slash at the end of the Url are now correctly followed (Bug fix).
Original Sphider does not except links like: http://www.abc.de
Sphider-plus adds the required slash automatically like: http://www.abc.de/

Correct template selection for different css files in different template folders (Bug fix).

15.2 Version 1.0.a

Build up with Sphider v.1.3.4.b

Bug fixed in function validate_email

Involved files that have been modified / added for this release:
.../include/commonfuncs.php

15.3 Version 1.1

Build up with Sphider v.1.3.4.b

Included converters for indexing PDF, DOC, RTF, XLS and PPT files.
To be activated individually in Admin settings
Warning message during index process when deactivated file was found

Captcha protection for Submission Form '*Suggest a new Site*'.
Use of Captcha to be activated in Admin settings

Automatically adapt Sphider's dialog to user language.
Improved version by ^demon

Bug fixed in language depending user dialog.

Bug fixed in function check_robot_txt.

Involved files that have been modified / added for this release:
addurl.php
search.php
.../converter/ all files
.../converter/charsets/ all files
.../admin/configset.php

.../admin/ext.txt
.../admin/messages.php
.../admin/spiderfuncs.php
.../include/captcha.TTF
.../include/make_captcha.php
.../languages/ all files
.../settings/conf.php

15.4 Version 1.2

Build up with Sphider v.1.3.4.b

UTF-8 support for (nearly) all charsets.

- Selectable in Admin settings the translation into UTF-8 charset can be enabled.
- Index and search functionality for Unicode.
- Please notice the important information and details to be found in chapter: UTF-8 Support and 'Preferred Charset'

Individual preferred charset.

- Charset for result page can be defined in Admin settings.
- This option will be overwritten by the UTF-8 option.

Use of '*Default results per page*' (10, 20, 30, 50) also for Sites table in Admin section.

Use of '*Default results per page*' (10, 20, 30, 50) also for Link search (site:).

Included PHP version check before admin.php could be used.

Translated Danish language file.

- Thanks to Brian Jorgensen

Media files excluded from index/re-index procedure.

- Enlarged file list in .../admin/ext.txt

Improvements and bug fixes in:

- 'Admin settings' dialog
- 'Did you mean' option
- !strict search
- Converter for non-HTML files
- Site search (site:)
- Addurl suggest form

Involved files that have been modified / added for this release:

.../addurl.php
.../search.php
.../admin/admin.php
.../admin/admin_header.php
.../admin/auth.php
.../admin/configset.php
.../admin/db_main.php
.../admin/spider.php
.../admin/spiderfuncs.php
.../admin/ext.txt
.../converter/ConvertCharset.class.php
.../converter/charsets/ all files
.../include/searchfuncs.php

.../include/search_links.php
.../include/js_suggest/suggest.php
.../language/ all files
.../settings/conf.php

15.5 Version 1.3

Release date: March 31, 2008
Build up with Sphider v.1.3.4.b

Tolerant search

- Selectable in search-box like AND/OR/Phrase and as new item: *'Tolerant search'*
- Presents results that are 'like' the query as an integrated *'Did you mean'*
- Presents search results for queries with e=é=è=ê, ä=a, Ü=U etc.
- Results are independent whether the user enters e or é or ê in the search query

Clear Category table

- Additional item in Admin section *'Database & Log Cleaning Options'*
- Deletes all categories not associated with any valid site

Fixed charset to UTF-8 for User Suggestion Form (addurl).

Involved files that have been modified / added for this release:

.../addurl.php
.../search.php
.../admin/admin.php
.../admin/spider.php
.../include/searchfuncs.php
.../languages/ all files

15.6 Version 1.3.a

Build up with Sphider v.1.3.4.b

Individual *'Erase & Re-index'* function for single sites.

- Additional item in Admin sites submenu *'Manage Site Indexing Options'*
- *'Erase & Re-index'* functionality for selected site

Translated Spanish and Dutch language files.

- Thanks to Willy

Involved files that have been modified / added for this release:

.../admin/admin.php
.../languages/es-language.php
.../languages/nl-language.php

15.7 Version 1.4

Release date: May 28, 2008
Build up with Sphider v.1.3.4

In Admin settings the method of chronological order for result listing can be defined.

Results ordered by:

- Relevance (weight)
- Main URLs (domains) on top
- First URL names and then weight
- Only top 2 per URL

The mode of chronological order for result listing is shown as additional headline on top of the result pages.
To be activated in Admin settings.

Select method of highlighting for found keywords in result listing.

If 'Advanced search' is activated, the user may select:

- bold text
- marked yellow
- marked green
- marked blue

The default highlighting can be defined in Admin settings.

If in Admin settings the option '*Index words in Domain Name and URL path*' is activated, found keywords now are highlighted also in result listing (row URL).

If in users browser JavaScript is disabled, a warning message is displayed on top of the search form that full functionality of Sphider-plus will not be available (required for the suggest framework).

Improved printout for '*Show sites in category*'. If in Admins '*Site options*' the content for 'title' was not included, now title and short-description will be fetched from the HTML header (of the indexed sites). If also this information is not available, a warning message will be displayed.

Enable index and re-index for pages with duplicate content.

Additional item in Admin settings:

- If selected, pages with content that was already indexed by another page will also be indexed/re-indexed. A warning message together with the URL that also holds the duplicate content will be presented in spider log output.
- If not selected, the link (page) will be ignored. Never the less the message and URL info will be presented.

Improved function '*If available follow sitemap.xml*' in order to prevent '*Page is duplicate*' messages.

Improved printout if PDF files cause indexing problems.

If '*Follow sitemap.xml*' is activated and a valid sitemap was found, the log output

Links found: 0 - New links: 0

is no longer shown. Because all links are delivered from the sitemap file and new links are not searched during index / re-index.

An eventually non-existing log folder will be created automatically during index / re-index process. So, the message '*Logging option is set, but cannot open a file for logging.*' will be prevented.

If in Admin browser JavaScript is disabled, a warning message is displayed on top of Admin page that full functionality of Sphider-plus administration will not be available (required for warning messages).

Updated Romanian language file by CyBerNet.

Corrected Spanish language file by Willy.

Bug fixed in index / re-index function that caused problems to index words which consist only of upper case characters.

Bug fixed in index / re-index function that caused problems to index words containing the ' à ' character.

Some small improvements for result printout.

Length of words to be indexed is increased to 255 characters per word. For the required modification in Sphider-plus database, please notice the additional FAQ information ([Can't search for long words](#)) in the readme.pdf document. If not required, this item must not be installed. Functionality of Sphider-plus does not depend on this modification.

Involved files that have been modified / added for this release:

- search.php
- .../admin/admin_header.php
- .../admin/auth.php
- .../admin/configset.php
- .../admin/install_all.php
- .../admin/messages.php
- .../admin/spider.php
- .../admin/spiderfuncs.php
- .../include/searchfuncs.php
- .../include/search_links.php
- .../languages/ all files
- .../settings/conf.php
- .../templates/all_folders/thisstyle.css

15.8 Version 1.5

Release date: July 14, 2008
Build up with Sphider v.1.3.4

Improved Suggest Framework. Now suggestions are presented also for queries with accented letters.

Enable real-time output of logging data. Selectable in Admin setting together with the update interval (1 - 10 seconds).

In order to prevent performance problems and memory overflow for large amount of URLs, Sphider-plus may clean resources during index / re-index. Selectable in Admin settings, this item periodical will:

- Free memory that is allocated to unused MySQL recourses.
- Unset PHP variables, which are no longer required.

Define max. length of URL presented in result pages.

An additional input field in Admin "Search Settings" is presented for Admin determination.

For 'Maximum length of page title displayed in search results' the title now will be broken at the end of the word exceeding the defined length. Not inside a word at the character count limit defined in Admin setting.

PDF converter for Linux Operating System included.

Needs to be individualized according to readme.pdf documentation, chapter ['PDF converter for Linux server'](#). Thanks to rasc.

Additional item in Admin section: Server Info

To be found in submenu 'Statistics', important information are presented for:

- Server
- Environment
- MySQL
- PDF converter
- php.ini file
- PHP integration

Enlarged Admin interface if database is empty.

Improved printout for database connection problems. Now MySQL error message is included.

Improved printout if text converter could not extract words from PDF, DOC, XLS etc. files.

Improved printout for Database Backup Management.

Modified installation script. Thanks to Flemp.

Font file renamed to captcha.tff (former: captcha.TTF). Thanks to ethix.

All style sheets now are centralized in ../templates/all_folders/thisstyle.css

Consequently the file ../include/js_suggest/SuggestFramework.css is no longer required.

Function 'create sitemap()' improved for XML conformity and moved from script ../admin/spider.php to script ../admin/spiderfuncs.php.

Bug fixed in 'Phrase Search' if UTF-8 support is not selected.

Bug fixed in highlighting of found keywords on result page.

Some small bug fixed for mysql queries.

Involved files that have been modified / added for this release:

- search.php
- .../admin/admin.php
- .../admin/admin_footer.php
- .../admin/configset.php
- .../admin/db_backup.php
- .../admin/db_main.php
- .../admin/install_all.php
- .../admin/install_reallog.php
- .../admin/install_sphider-plus.php
- .../admin/messages.php
- .../admin/real_get.php
- .../admin/real_log.php
- .../admin/real_ping.js
- .../admin/spider.php
- .../admin/spiderfuncs.php
- .../converter/pdfotext
- .../converter/pdfotext.script
- .../include/captcha.tff
- .../include/commonfuncs.php
- .../include/searchfuncs.php
- .../include/js_suggest/suggest.php
- .../settings/conf.php
- .../settings/database.php
- .../templates/all_folders/navdown.jpg
- .../templates/all_folders/thisstyle.css

Attention: Starting with version 1.5, Sphider-plus supports real-time output of logging info during index / re-index procedure. This item requires an additional table for the database. If you update from a former version of Sphider-plus, please run the .../admin/install_reallog.php script. If you upgrade from original Sphider or install from scratch, you don't need to run this script. Its features are also included in the other installation scripts.

15.9 Version 1.6

Release date:	September 06, 2008
Build up with Sphider:	v.1.3.4
Sphider-plus vs. original Sphider:	93 items wrought

Additional item in Admin settings to select:

- Instead of weighting %, show count of query hits in full text.
Selecting this item will also influence the order of result listing. Now only the number of keyword hits in full text will define the position of a page in result listing.

Additional item in Admin settings to select the chronological order of result listing:

- 'Most Popular Links ' on top.
Activating this item, Sphider-plus will present the result listing in order of before learned link attractivity. Defined as those links with the best user acceptance (clicks).

Additional items in Statistics overview:

- Queries total
- Link clicks total

Additional item in Admin / Statistics:

- Most Popular Links.
Presenting the quantity of clicks individual for each link with date and time of last click.
Also the latest query before clicking that link is presented.

Additional item in Admin / Clean:

- Clear 'Most Popular Links' log.

Additional item for index/re-index procedures:

- Temporary ignore 'robots.txt'.

If utf-8 support is activated, result listing now is independent for queries with upper- or lowercase letters. Or alternatively, if selected in Admin settings, distinct results for case sensitive queries could be performed.

Improved utf-8 support for non-Latin characters.

Improved suggest framework for utf-8 support. Now offering suggestions

- for phrases
- for accented letters
- for non-Latin characters

Known issue: Well working for Firefox and Opera browser, for non-Latin characters IE is not cooperative. Need to rewrite the Suggest Framework completely for a browser independent presentation of the suggestions.

Improved search functionality for queries with accent letters without selecting the utf-8 support.

Phrase search improved, so that common words and too short (min_word_length) words could be used as part of the query phrase and are no longer marked as ignored.

Improved functionality for 'Most popular searches'. Now also

- Advanced search settings
- Categories
- Mode of highlighting
- Results per page

will be taken into account when clicking a 'Most popular searches' suggestion.

Bug fixed that seduced Sphider to follow links that are placed in HTML comments.

Bug fixed that created a wrong weighting calculation for keywords placed

- behind a word that did not match 'min_word_length'
- behind a 'common' word
- first found in full text

Bug fixed in 'Strict search' that caused invalid highlighting in result listing.

Involved files that have been modified / added for this release:

```
search.php
.../admin/admin.php
.../admin/configset.php
.../admin/install_all.php
.../admin/install_bestclick.php
.../admin/install_sphider-plus.php
.../admin/spider.php
.../admin/spiderfuncs.php
.../include/click_counter.php
.../include/commonfuncs.php
.../include/searchfuncs.php
.../include/js_suggest/suggest.php
.../include/js_suggest/SuggestFramework.js
.../languages/ all files
.../settings/conf.php
```

Attention: Starting with version 1.6, Sphider-plus supports logging of 'Most popular links'. This item requires additional rows in 'links' table of the database. If you update from a former version of Sphider-plus, please run the .../admin/install_bestclick.php script. If you upgrade from original Sphider or install from scratch, you don't need to run this script. Its features are also included in the other installation scripts.