

Contents

[Overview](#)

[Select directory/files](#)

[Read Internet file](#)

[Parse Text/Csv Files](#)

[Sorting Urls](#)

[Search Options](#)

[Build Domains](#)

[Export functions](#)

[Options](#)

[Delete all with ...](#)

[Generate date patterns](#)

[Hit rates](#)

[Unescape/Cut off Parameters](#)

[Text Window](#)

[Misc](#)

[Command Line](#)

[Tips](#)

[Disclaimer of Warranty](#)

[Registration](#)

Unescape/Cut off Parameters

Some Urls contain ASCII characters in a so called **escaped** form. This means they are represented as **%XY** where **XY** stands for the hexadecimal notation of the character. For example a blank can be represented by **%20** or a question mark by **%3F**. Use **Unescape Urls** to reconvert such characters to their normal form.

Endings of Urls are often used to carry queries and/or parameters passed by an HTTP server to a cgi script looking like: `var1=id1&var2=id2&var3=id3 ...`
An example would be

`http://groups.google.com/groups?q=%22netscape+history%22+delete&hl=en`

If you want to get rid of such endings use **Cut off Parameters**. The preceding example then simply becomes `http://groups.google.com`.

Overview

UrlSearch extracts web addresses or other text strings from local or remote files. Results can be edited and exported in various ways. Web addresses can be sorted by domains and exported to your browser's bookmarks (MSIE, Netscape, or Opera). UrlSearch can also be used as a Download agent and can calculate hit rates, view history and cookie files, and edit 'Typed Urls' lists. A separate Text window offers some manipulating functions for HTML pages.

UrlSearch runs on Win9x/Me, WinNT 4.0 or later, and Win2K/XP/2K3. It consists of the files [urlsearch.exe](#), [urlsearch.hlp](#), and [readme.txt](#).

UrlSearch needs no Install/Uninstall. Simply place these files into a separate directory. To uninstall, remove the directory from the disk.

Select Files/Directory

When selecting a directory you can choose **Search all files of the form** or **Search all files except**. In either case you may specify up to twelve expressions of the form 'X*.Y' where X stands for the beginning term of a filename and Y for its extension. The first entry of the drop down list can be saved on exit (see Options). You can also choose **Search all non binary files**. In this case UrlSearch stops searching a file when it meets ASCII characters smaller than 0x20 (except line feeds and tabs).

Alternatively you can choose a number of files from the standard Open Files Dialog. To choose more than one file, keep the Shift key pressed (or Control key, respectively) when clicking.

You may also start a search from another application by **Drag and Drop** to the UrlSearch window.

When reading from Clipboard, preceding results are not deleted so that new results can be added. For example, you can use your Browser's page source function to copy a web page to the Clipboard and then use 'Read from Clipboard' to extract its web addresses by UrlSearch.

When you look for some 'lost' Url, it may be a good idea to search in the folder Windows\History or in your Browser's Cache folder.

Search Options

You can specify up to 10 strings which must be part of the url and/or title, or which must not be part of it, respectively. Terms may be separated by spaces or commas. Type \, if the comma should be interpreted as a string by itself. If you choose **Suspend**, the specifications will not be applied until you choose **Ok** again. Press the paper basket button to get rid of strings no longer needed.

Build Domains

This will sort the results displayed in the main list in an Explorer-like style window with domains displayed as folders. This will make it easier to look through the search results. This is however only available when a search for 'http and ftp' has been performed and it also needs the list to be sorted by the ascending Url column. If necessary click on the left most section of the Url column header to sort the list correctly. The export functions may be applied to each folder and will reflect its subfolder structure. Once the domains are created the toolbar's tree-view button can be used to toggle view between the main list and the domain's window. Accordingly, the toolbar buttons for exporting items will either refer to the main list or to the folder selected in the domains' window. Deleting url's from the domains' window will automatically delete them from the main list and vice versa. When the Menu item **Rebuild domains after search** is checked, the domains will automatically be rebuilt after each search. This can be saved on exit by checking **Search Options**.

Export functions

Add to Bookmarks

You can add the search results to the bookmarks of MSIE, Netscape, or Opera, respectively. UrlSearch will save them in a new folder named for example **UrlSearch, Saturday November 17, 2003, 10:35:40** with the second term being the date and time of creation of the new folder. When a folder from the domains' window is exported the new bookmark folder will be named **UrlSearch, Name of exported folder** and its subfolders will keep their name. Note that since Browsers keep their bookmarks in memory while running, it may be necessary to reload the bookmarks before you can use the folder added by UrlSearch. Especially, when you use older versions of Opera you will not be able to add bookmarks while Opera is running since, for example, Opera v3.21 writes its bookmarks from memory back to disk without checking for changes made by other programs. Adding a bookmark to MSIE means a new file is created. Creating large amounts of files at a time (let's say more than 1500) can make the running session very slow. You can press the toolbar's stop button to cancel the process.

Save as

You can save the results as a text file. When choosing an existing file for saving you will be asked whether the new entries should be appended to the file. Results can also be saved as an HTML file. Using the 'Replace Text' function each Url can be appended a delimiting comma, for example. The main list can also be saved as a CSV file.

Print, Copy to Clipboard

Before printing the results you can choose a printer font from the corresponding menu item or toolbar.

You can choose in the Option's dialog which list columns should be exported. For example, to **export titles** check column 2. If you choose any of the export functions from the list's Context Menu (Shift+F10) via **Selected Items** then only the selected list entries will be exported.

Options

Search for

except for http and ftp addresses you can alternatively search for mailto, news, gopher, wais or telnet addresses, Java Script functions, email addresses, IP's, or for a single user defined string. If **prefix** is checked each output string will start with the search string itself. If **url like** is checked the output string is not allowed to contain spaces or other characters not allowed in Urls. If the minimal length of the output string is set to a value that exceeds the length of the search string it may happen that no items are obtained although the search string may be contained in the file.

Specifications made in this section are saved when 'Save on exit, Search Options' is checked. You can also save all the current search settings in a separate file which can be loaded again for later use. Use the Menu item 'Options, Save Search Settings'.

Some Options can be modified from the bottom of the Main Window. Additionally, there is a



second combo box filled with a constant set of search strings read from the text file usearch.txt located in the UrlSearch directory.

Note that UrlSearch replaces many HTML specific strings by their corresponding ascii characters, for example **>** with **>**. Therefore an output string may contain characters not actually contained in the file.

When searching for IP's, UrlSearch will NOT list expressions of the form **ip1.ip2.ip3.ip4** where $ip1+ip2+ip3+ip4$ is smaller or equal to 10 (or the value you specified). Also, **ip1** must be greater than 10. The check boxes **url like**, **prefix**, and **entire line** apply in this case to the context string.

Save on exit 'Search Options'

if this option is checked UrlSearch will save the settings made in the 'Search Options' dialog including whether the restrictions should be applied or not. It will also save the first entries of the dropdown lists in the 'Search directory' dialog so that they will always reappear as first entries. Use the paper basket symbol to delete entries from the drop down lists. Current search settings can also be saved in a separate file for later use (see above).

Export

choose the list columns that should be exported. For example, check column 2 if you want to

export titles. Prefixes like 'http://' and 'ftp://' can be omitted or replaced by different strings using the 'Replace Text' function from the Find Menu.

Locations

here you should specify the location of your bookmarks so that UrlSearch can append Url's to them. Note that it may be necessary to reload the browser's bookmarks before you can use the new ones inserted by UrlSearch. UrlSearch will suggest the **Initial search directory** the first time you choose any of the Open File Dialogs.

Delete all with ...

You may specify in each box up to 5 strings separated by commas, or simply by spaces. For example, the specification

`cgi, banner`

made in the 'containing' box will delete all entries containing `cgi` and/or `banner` from the list.

Type `cgi\`, if you want items containing `cgi`, (including the comma) to be deleted. The specification

`computer, altavista`

made in the 'not containing' box will delete all entries containing neither `computer` nor `altavista`. Hence `http://www.geocities.com/Avenues/Computers_and_Technology` would not be deleted. Note that this is not case sensitive. Note also that the 'containing' box will have priority over the 'not containing' box.

The same applies to the `Select all with...` dialog.

Generate date patterns

The contents of the left pane is read from the file update.txt which may be edited freely using the following format identifiers:

%%	Character %
%a	Abbreviated weekday name
%A	Full weekday name
%b	Abbreviated month name
%B	Full month name
%c	Date and time
%d	Two-digit day of month (01 - 31)
%H	Two-digit hour (00 - 23)
%I	Two-digit hour (01 - 12)
%j	Three-digit day of year (001 - 366)
%m	Two-digit month as a decimal number (1 - 12)
%M	2-digit minute (00 - 59)
%p	AM or PM
%S	Two-digit second (00 - 59)
%U	Two-digit week number where Sunday is the first day of the week (00 - 53)
%w	Weekday where 0 is Sunday (0 - 6)
%W	Two-digit week number where Monday is the first day of week the week (00 - 53)
%x	Date
%X	Time
%y	Two-digit year without century (00 to 99)
%Y	Year with century
%Z	Time zone name, or no characters if no time zone

Hit rates

The number of hits per file or over all files, respectively, can be calculated from the output list. When interested in Hit rates you should have checked **Do not omit duplicates** in the Option's dialog because otherwise UrlSearch automatically suppresses duplicates which follow closely each other in the same file, or even omits all duplicates if that option is checked. Also make sure that the **minimal length** of the output string is set to a value not exceeding the length of the search string because that also could result in omitting items. The resulting list can be saved as an HTML table from the context menu.

Misc

Join Download Files

Suppose you have downloaded a file `down.htm` with associated sub directory `down_files`. Then UrlSearch will move the file `down.htm` into the sub directory `down_files`, rename the file to `_index.htm`, modify the corresponding links within the file, and finally remove the suffix `”_files”` from the directory name. **Warning: do not apply this function to subfolders of `down_files`.** You can associate this dialog to a UrlSearch desktop icon if you give it the `“-join”` argument. Then files or directories can simply be dropped on that icon.

Edit Typed Urls

these are the Urls that appear in the Browser’s drop down list for entering a new Url in the address bar. UrlSearch lets you edit these Urls and save them as the Browser’s new drop down list (choose **Save to Browser** from the context menu).

Edit ‘About Urls’

When typing `about:title` into the MSIE address bar the corresponding Url is taken from the `title` entry of the registry key

`Hkey_Local_Machine\Software\Microsoft\Internet Explorer\AboutUrls`.

You can easily edit these entries using UrlSearch (you could of course also use Regedit.exe). Existing entries should not be modified since some of them are used internally by MSIE.

Current History, Index files

For MSIE v5.0 or higher ‘Current History’ displays information of the files `Index.dat` from the History directory and its sub directories. For Netscape, information of the file `Netscape.hst` is displayed. For MSIE this can also be used for the `Index.dat` file of the Cache folder (Temporary Internet files) or Cookie folder. These two features are not supported for Netscape 6.0 or higher.

The ‘Current History’ function for MSIE also includes a ‘Delete from File’ option which overwrites the selected Urls in the corresponding History file by some default bytes. This should only be applied to Urls that can not be deleted through the History pane’s built in delete function. MSIE should not be running during this operation. Note that Urls appearing in the MSIE history pane are taken from different places, including the Cache folder.

List Cookies

this displays information stored in the File `index.dat` of the Cookie directory. Each cookie saves its data in an associated text file the name of which is contained in the ‘Text file’ column. If this text file no longer exists its name is preceded by a double star. Double click to view the contents of the file.

RAS Autodial addresses

this displays data contained in the Registry key HKEY_CURRENT_USER\Software\Microsoft\RAS Autodial\Addresses.

Files in use (only on Win9x/Me)

To see which files are used by a certain application proceed as follows:

- run 'Files in use',
- start the application,
- run 'Files in use' again,
- sort items by column 'B'.

Now the files used by that application appear at the top of the list. (Don't know why the Registry files never appear in the list).

Disclaimer of Warranty

This software is provided 'as is'. There is no warranty of fitness for a particular use, and no other warranty of any kind either expressed or implied regarding this software. Accordingly, the entire risk as to the use, results and performance of the software is assumed by the user. In no event will the author be liable to the user or anyone else for any consequential, special, or indirect damages resulting from use, possession or malfunction of this product.

Registration

Among the benefits of the registered version are

*Search in sub directories,
Export functions for selected list items,
Unlimited saving/loading Sessions,
Reading from Clipboard,
Unlimited distance level for following links,
Correct summing up of Hit rates.*

UrlSearch can be registered through *RegNet* at <http://www.reg.net> (product ID number is 5579). The registration rate is US \$18 for a single-user license. You will then receive a full version via e-mail. For licenses other than single-user please contact the author.

Read Internet file

You can search through an Internet file requested from an HTTP or FTP server. When 'Follow Links' is checked UrlSearch will also search through further Urls found on that page. The unregistered version only allows a distance level of 1 so that no further child pages are read. Strings specified in the 'Links must/must not contain' boxes must be separated by blanks (not by commas). Urls with ending gif, jpg, ... usually are binary files and will therefore not contain any links to follow. There is therefore a default string '.gif .jpg .tif .mp3 .zip .exe .pdf' in the 'Links must not contain' box. Settings can be saved to text files with ending .usj. When associating such files to UrlSearch they will be loaded on double click. When saving an HTML file to disk UrlSearch will insert the original Url at the top of the file.

Sorting Urls

For sorting the list by a specific column, click on the left most section of the column header. Apart from standard sorting, urls can be sorted by their length and their ending. For this go to the 'Edit' Menu and choose 'Sort Urls'. Note however that some functions require the list to be sorted by the ascending Url column. These functions are 'Remove Dupes', 'Build Domains' and 'Hit rates'.

Tips

To launch a Url not contained in the first column keep the Shift or Control key pressed

If you associate file extensions .urs, .usj, and .uss to UrlSearch such files will be loaded on double click

To use 'Read Internet Files' with more than one Remote File leave the remote file edit box empty and press Ok. Then copy the target urls into the Clipboard and paste them into the 'Read Internet Files' window.

Use the "Load source" item from the context menu to open the currently selected file in the Text window

Ctrl+c will always copy first column items regardless of what is checked at Export Options

Ctrl+n with n an integer will copy the n'th column item

Parse Text/Csv Files

Parse Text Files

This is a tool to retrieve multiple items or lines from Text or HTML files. When nothing entered simply a list of all words of the text file is generated. If only **Entire Line** for column A is checked all lines of the text file will be listed. This may be used to merge a number of text files. If you enter a search string for column B and check **Entire Line** for column B as well as for column A the lines containing the search string together with one preceding line will be listed. If you want two preceding lines listed use column C instead of B. Text strings in HTML files are often marked like

```
<FONT FACE="Arial, Helvetica" size=2>Aug 19, 2003</FONT>
```

If you want to retrieve all such items from a file (arbitrary line length) you would enter the search term **Helvetica" size=2>** so that the column would be filled with **Aug 19, 2003**. Use then the Replace Text function from the Find Menu to cut off **** so that only **Aug 19, 2003** remains in the column. If “per Line” is checked each line generates a new record.

Parse CSV Files

This can be used to simply extract a number of columns from a CSV file.

Command Line

UrlSearch can be used with the following **command line arguments**:

UrlSearch **-stdin** reads from stdin (for example, the command "dir | urlsearch -stdin"
 scans the output of the "dir" command)

UrlSearch **-join** open Join Download files dialog

UrlSearch **-bdown** open Batch Download dialog

UrlSearch [File/Dir, or **-jobFile**] [**-save2File**, or **-saveh2File**]

Examples:

UrlSearch E:\Test\test.txt **-save2**E:\Test\save.txt
read the file test.txt, save the results to save.txt (this file must not exist), and exit

UrlSearch E:\Test\test.txt **-saveh2**E:\Test\save.htm
read the file test.txt, save the results in html format to save.htm (this file must not exist),
and exit

UrlSearch E:\Test
read all files in E:\Test (no sub directories)

UrlSearch **-job**E:\Test\test.usj
carry out the job saved in test.usj

Text Window

The Text Window offers some manipulating functions for HTML pages like Strip Tags, Remove comments, and decode Hex or Ascii numbers. For example when decoding Hex Html text an expression like `http://%31%392.%385.2%331.%312/z/r%65%6do%76e.h%74m%6c` becomes <http://192.85.231.12/z/remove.html>, and when decoding Ascii numbers the expression <http://192.85.231.12/z/remove.html> also becomes <http://192.85.231.12/z/remove.html>.

When decoding Hex Script text the expression `window.alert("\u0068\u0065\u0078\u0020\u0063\u006f\u0064\u0065')` becomes `window.alert('hex code')`.

You may choose to start UrlSearch in Text Window mode by checking the corresponding option at Options->View. To undo this press the “New Window” Button and go to Options again (or edit the TextWin entry in urlsearch.ini). UrlSearch will also start in Text Window mode if it gets the parameter `txt_pane`.

