

NAME

s.cgi – aspseek search CGI front-end

SYNOPSIS

http://your.host.com/cgi-bin/s.cgi[?parameter=value[¶meter=value...]]

DESCRIPTION

s.cgi is a search front-end of aspseek. It is a CGI program, so usually it is run by web server. **s.cgi** parses its options, reads its configuration and template from **s.htm(5)** file residing in `/usr/local/aspseek/etc` directory, makes a query to **searchd(1)**, formats the results according to template, and outputs HTML page containing results found, or error, or empty search form in case no query was specified.

Functionality of **s.cgi** related to template and meta symbols is described in **s.htm(5)**.

OPTIONS

q=query Query string. Input line in search form, used for query entering, should be named "**q**".

ul=sql_pattern [*sql_pattern* ...]

ul=site [*site* ...]

URL mask. Several masks can be used, delimited by exactly 1 space. This parameter can also be used multiple times, values are concatenated. *sql_pattern* specifies subdirectory of site, and record in **subsets** table containing value of *sql_pattern* in the field **mask** must exist. *site* specifies the whole site, it should be in the form `http://www.sitename.com/`. Note that *sql_patterns* and *sites* can not be mixed together. Examples:

ul=http://www.aspstreet.com/

ul=http://www.aspstreet.com/directory/%

r=query Previous query string. Used for searching in results, when **in** parameters is equal to *on*.

s=rate | date

Result sorting type. *rate* means sorting by rate (relevance), and *date* means sorting by last modification date. Sorting order is descending, so results with greater rate or latter modification date will be first ones.

in=on Used for "search within results" feature. If value of this parameter is *on*, then actual query is composed from parameters **q** and **r** combined by AND. Any other value is ignored. Checkbox "Search within results" in query form, that is displayed in search results, should be named **in**. Also, the following string should be put in query form:

```
<INPUT TYPE="hidden" NAME="r" VALUE="$q">
```

t=query Used for "Take me there" feature. If this parameter is present, then **s.cgi** finds first results matching query and generates output for redirection to that URL.

ch=url "Cached" URL. This parameter is used for displaying "cached" copy of web page. Can be combined with **q**, **cs** and **tmpl** parameters. If **q** and **cs** parameters are present, then appropriate query terms will appear highlighted in "cached" page.

np=number

Result page number. Default value is *0*.

ps=number

Number of results per page. Overrides value set by "preferences" cookie.

gr=off

Grouping results by site. If value of this parameter is *off*, then results are not grouped by site. Any other value is ignored.

st=number

Site ID. Value of this parameter is used to restrict search by specified site. Generated in search results, as result of `$$H` meta symbol, which is used in **moreurls** template.

tmpl=file

Template file name. *file* is relative to `/usr/local/aspseek/etc` directory. Default is constructed from the part of **s.cgi** name before point with *.htm* added, so **s.cgi** will look for **s.htm**, and if you rename it to **another.cgi**, it will look for **another.htm**.

spN=on

Space ID. *N* is one or more digits value. If value of this parameter is *on*, then search is restricted by web space with ID equal to *N*. Any other value is ignored. This parameter can be used multiple times, all spaces are being ORed.

To assign particular site to web space, insert record with appropriate values or site ID and

space ID into table **spaces**.

cs=charset Source charset. Tells **s.cgi** which *charset* is used in input query. This is required parameter if non-ascii characters are used in query. Results of query will also be presented in that *charset*.

fm=on | off | lang[,lang,...]

Word forms. Can be comma-separated list of languages or just *on* or *off*. In case it is not set to *off*, **s.cgi** will search for all forms of specified words, and results with exact word forms will be displayed first.

Example: if word 'create' is specified, then documents containing either 'create' or 'creates' or 'created' will be found.

ad=on Account distance. If value is *on*, then **s.cgi** ranks documents with search terms closer to the beginning higher. Any other value is ignored. Overrides **AccountDistance** parameter in **searchd.conf(5)**.

ln=url Link to page. Tells to **s.cgi** to find pages which have link to the specified *url*. If protocol in *url* is omitted, then *http://* is implied. If *url* is not found and last symbol is not '/', than **s.cgi** tries to find URL with '/' at the end.

Limiting search scope to certain parts of HTML pages

Parameters **bd**, **ds**, **kw** and **tl** can be used together.

bd=on Search in body. If value of this parameter is **on**, then advanced search will be performed only within HTML `<BODY>...</BODY>`. Any other value is ignored.

ds=on Search in description. If value of this parameter is **on**, then advanced search will be performed only within HTML `<META NAME="DESCRIPTION"...>`. Any other value is ignored.

kw=on Search in keywords. If value of this parameter is **on**, then advanced search will be performed only within HTML `<META NAME="KEYWORDS"...>`. Any other value is ignored.

tl=on Search in title. If value of this parameter is **on**, then advanced search will be performed only within HTML `<TITLE>...</TITLE>`. Any other value is ignored.

Limiting search scope to certain period of dates

dt=back | er | range

Type of time limit. See below.

dp=date If 'dt' is 'back', that means you want to limit result to recent pages, so you specify that "recentness" in *date* value, that is specified in the form *xxxA[yyyB...]*, there *xxx*, *yyy* are numbers (which can also be negative), and *A*, *B* can be one of the following (the letters below are the same as in **strptime(3)** and **strptime(3)** functions):

Character	Meaning
<i>s</i>	second
<i>M</i>	minute
<i>h</i>	hour
<i>d</i>	day
<i>m</i>	month
<i>y</i>	year

Examples of values for **dp** parameter:

String	Meaning	Value, s
<i>4h30m</i>	2 hours and 30 minutes	16200
<i>1Y6m-15d</i>	1 year and six month minus 15 days	45792000
<i>1h-60M+1s</i>	1 hour minus 60 minutes plus 1 second	1

Note that **ASPseek** do not use minutes and seconds of document's last modification date, so specifying something more precise than hour is useless (but still allowed).

dx=1|-1 If **dt** is *er* (which is short for newer/older), that means the search will be limited to pages newer or older than date given. Parameter **dx** is newer/older flag, value *1* means "newer" or "after", and value *-1* means "older" or "before". The actual date is separated into **dd**, **dm**, **dy** fields as follows.

dd=number

Day of month (1...31)

dm=number

Month (0 - January, 1 - February, ..., 11 - December)

dy=number

Year (four digits, for example 2001)

db=dd/mm/yyyy

de=dd/mm/yyyy

If **dt** is *range*, that means search within given range of dates. Parameters **db** and **de** are used in this case and stands for beginning and ending date, respectively. Each date is in the form *dd/mm/yyyy*, there *dd* is day of month number (1...31), *mm* is month number (1...12), and *yyyy* is four-digits year.

fr=value

to=value These parameters are passed to **s.cgi** in subsequent search pages; they contains date and time in internal format used by **s.cgi**.

Advanced search

iq=words Include *words*. Used in advanced search. If this parameter is not empty, then parameter **q** is not used.

xq=words Exclude *words*. Used in advanced search. All words found in value of this parameter will be added to query with - sign before them.

im=p Include mode. If value of this parameter is equal to *p*, then value of parameter **iq** is double-quoted, which means pages containing phrase must be found.

xm=p Exclude mode. If value of this parameter is equal to *p*, then value of parameter **xq** is double-quoted, which means pages containing phrase must be excluded.

is=site_pattern [*site_pattern ...*]

Include sites. This parameter is used to restrict search by sites those names contains *site_pattern*. If several patterns delimited by space are used, then they are ORed. Examples: **is=www.google**, **is=.com**.

xs=site_pattern [*site_pattern ...*]

Exclude sites. This parameter is used to to exclude sites those names contains *site_pattern* from results. If several patterns delimited by space are used, then they are ORed.

o=number Use "*number*+1"th template sections if possible. See "Defining different output formats" subsection in **s.htm**(5).

ENVIRONMENT

QUERY_STRING

This variable is usually set by web server and contains all the parameters for **s.cgi** described in OPTIONS above.

ASPSEEK_TEMPLATE

UDMSEARCH_TEMPLATE

Template file to use. Overridden by **tmpl** parameter. If both variables are used, **ASPSEEK_TEMPLATE** is preferable.

HTTP_COOKIE

Cookies can be sent by client browser together with request. Web server forms **HTTP_COOKIE** environment variable from it. **s.cgi** parses it and uses value of **PS** section as the number of results per page. This is overridden by **ps** parameter.

HTTP_HOST

Value is set by web server and is used together with self name to form HREF references to other result pages.

SCRIPT_NAME

Value is set by web server and is used to determine the self name of the script and the name of template file to load (see description of **tmpl** parameter).

REDIRECT_STATUS**REDIRECT_URL****PATH_TRANSLATED**

These variables can possibly be set by web server (for example, by Apache using its "AddHandler" and "Action" directives). If **REDIRECT_STATUS** is set, **s.cgi** will get the self name from **REDIRECT_URL**, and template file name from **PATH_TRANSLATED**.

FILES

/usr/local/aspseek/etc/s.htm

BUGS

It appears that **tmpl** parameter is broken since version 1.2.8. Use either **ASPSEEK_TEMPLATE** environment variable, or rename **s.cgi**.

NOTES

When parameters are passed in URL, they should be encoded according to RFC 1738. If values are coming from form input fields, they are encoded by browser (so, for example, space character becomes either '+' or '%20'). **s.cgi** does all appropriate decoding of parameter values and encoding of links to the other pages it generates whenever needed. In this page values are shown in non-encoded form.

SEE ALSO

s.htm(5), **searchd(1)**, <http://www.aspseek.org/>.

AUTHORS

Copyright (C) 2000, 2001, 2002 by SWsoft.

Man page by Kir Kolyshekin <kir@asplinux.ru> and Alexander F. Avdonkin <al@asplinux.ru>.