NAME

search − SWISH++ searcher

SYNOPSIS

search [*options*] *query*

DESCRIPTION

search is the SWISH++ searcher.

Query Input

The formal grammar of a query is:

query:	query optional_relop meta
	meta
meta:	meta_name = primary
	primary
meta_name:	word
primary:	(query)
	not primary
	word
	word∗
optional_relop:	and
	or
	(empty)

In practice, however, the query is the set of words sought after, possibly restricted to meta data, and possibly combined with the Boolean operators ''and,'' ''or,'' and ''not.'' The asterisk (∗) can be used as a wildcard character at the end of words. Queries are evaluated in left-to-right order, i.e., ''and'' has the same precedence as ''or.'' See the EXAMPLES.

Results Output

The results are output in the format of:

rank path-name file-size file-title

where *rank* is an integer from from 1 to 100. The *path-name* is relative to where the set of files was originally indexed. If the file is an HTML file, then *file-title* is its title; otherwise, it is its filename. A results line can be parsed easily in Perl with:

```
($rank,$path,$size,$title) = split( / /, $_, 4 );
```

In addition to results lines, comment lines may also appear containing additional information about the query results. Comment lines are in the format of:

\# *comment-key*: *comment-value*

The *comment-keys* are:

ignored: *stop-words*	The list of stop-words (separated by spaces) ignored in the query.
not found: *word*	The word was not found in the index.
results: *result-count*	The total number of results.

OPTIONS

−**c***config_file*   The name of the configuration file to use. The default is swish++.conf in the current directory. A configuration file is not required: if none is specified and the default does not exist, none is used; however, if one is specified and it does not

	exist, then this is an error.
−**d**	Dump the query word indices to standard output and exit. Wildcards are not permitted.
−**D**	Dump the entire word index to standard output and exit.
−**i***index-file*	The name of the index file to use. The default is swish++.index in the present working directory.
−**M**	Dump the meta-name index to standard output and exit.
−**m***max-results*	The maximum number of results to return overriding the compiled-in default (which is usually 100).
−**r***skip-results*	The initial number of results to skip. The default is 0. Used in conjunction with −**m**, results can be returned in ''pages.''
−**R**	Print the result-count only.
−**s**	Perform stemming (suffix stripping) on words during the search. Words that end in the wildcard character are not stemmed.
−**S**	Dump the stop-word index to standard output and exit.
−**V**	Print the version number of **SWISH**++ to standard output and exit.
−**w***size*[,*match*]	Dump a ''window'' of at most *size* lines around each query word matching *match* characters. Wildcards are not permitted. The default for *match* is 0. Every window ends with a blank line.

CONFIGURATION FILE

The following variables can be set in a configuration file. Variables and command-line options can be mixed, the latter taking priority.

IndexFile	Same as the −**i** option.
ResultsMax	Same as the −**m** option.
StemWords	Same as the −**s** option.

EXAMPLES

The query:

        librar*

will return all documents that contain ''library,'' ''libraries,'' or ''librarian.'' The query:

        mouse and computer

will return only those documents regarding the kind of mice attached to a computer and not the rodents. The query:

        cat or kitten or feline

will return only those documents regarding cats. The query:

        mouse or mice and not computer

will return only those documents regarding mice (the rodents) and not the kind attached to a computer. The query:

        mouse and computer or keyboard

is the same as:

        (mouse and computer) or keyboard

in that they will both return only those documents regarding either mice attached to a computer or any kind of keyboard. However, neither of those is the same as:

        mouse and (computer or keyboard)

that will return only those documents regarding mice and either a computer or a keyboard.

Examples Using Meta Data
The query:

        author = carroll

will return only those HTML documents whose author attribute contains ''carroll.'' The query:

        author = stevenson treasure

will return only those HTML documents whose author attribute contains ''stevenson'' and also regarding treasure. The query:

        author = (lewis carroll)

will return only those HTML documents whose author is Lewis Carroll. The query:

        author = (lewis carroll) or wonderland

will return only those HTML documents whose author is Lewis Carroll or that contain the word ''wonderland'' anywhere in the document regardless of the author.

EXIT STATUS
Exits with one of the values given below:

        0   Success.
        1   Error in configuration file.
        2   Error in command-line options.
        50   Malformed query.
        51   Unable to read index file.

FILES
        swish++.conf   default configuration file name
        swish++.index   default index file name

SEE ALSO
index(1), **perlfunc**(1), **swish++.conf**(4)

M.F. Porter. ''An Algorithm For Suffix Stripping,'' *Program,* 14(3), July 1980, pp. 130-137.

AUTHOR
Paul J. Lucas *<pjl@best.com>*