

## NAME

searchd.conf – searchd configuration file

## SYNOPSIS

**/usr/local/aspseek/etc/searchd.conf**

## DESCRIPTION

**searchd.conf** is a configuration file for **searchd(1)**. The following parameters can be defined:

### General

**DBAddr DBType**:*[[User[:Pass]@]Host[:Port]]/DBName/*

Defines SQL server connection parameters.

**DBType** is SQL server type, it can be *mysql* or *oracle8* for now.

**User** is a SQL server's user to connect as.

**Pass** is a **User**'s password. If this field is omitted, no password is used.

**Host** is a host name or IP address of host to connect to. If you are running SQL server on the same machine, use *localhost*.

**Port** is a port number on which database is listening for SQL queries. Default is the same as default port of used SQL server.

**DBName** is a name of the database used.

### Port *nnn*

Sets the port number on which **searchd(1)** is listens to **s.cgi(1)** queries. Default is *12345*.

### DBLibDir */some/dir*

Adds */some/dir* to list of directories to search for database backend library (*libdbname-version.so*). Default library search path is */usr/local/aspseek/lib*. Several such options can be used, each adding one more directory to the list. Last added directory is used first; compiled in path is last.

### AllowFrom *some.host.com | xxx.xxx.xxx.xxx[/yy]*

This implements access control list, so **searchd(1)** will only accept connections from host(s) specified. Several such options can be used. You can specify hostname, IP address, or subnet (IP address with mask in CIDR notation).

### DataDir */some/dir*

Sets directory in which delta files and files with information about words, subsets, spaces will be stored. Default is */usr/local/aspseek/var*.

### DebugLevel *none | error | warning | info | debug*

Sets the level of debugging. If set to *none*, nothing will be logged. If set to *debug*, you will get a bunch of messages. Default value is *info*.

### MinFixedPatternLength *nnn*

Sets the minimal length of fixed part of word with pattern (like *someh\**) to be allowed in search query. Words shorter than this value will be rejected with appropriate error. Setting this to less than 3 will open ASPseek to DoS attacks. Default value is *6*.

### MaxThreads *nnn*

Sets the maximum number of threads that search daemon can run simultaneously to process queries. Setting high value can result in big memory consumption. Setting low value can result in big response time for queries in high load conditions (as "extra" queries are queued). Default value is *10*.

### MultipleDBConnections *yes | no*

Sets whether to use separate connection to the database for each thread. If multiple connections are used, this leads to better concurrency between threads, especially when one or more threads perform pattern search and the other is trying to perform simple search. Default is *yes*.

### Include *file*

Includes the contents of *file* at this point, so you can specify some parameters in that included *file*. File name is relative to ASPseek etc directory (*/usr/local/aspseek/etc*).

### Database format parameters

These parameters tells **searchd(1)** what database format is used by **index(1)**, so their values should be set to the same values as in **aspseek.conf** file.

**HiByteFirst** *yes | no*

Sets the byte ordering used in field **wordurl[1].word** (only in Unicode version). Default is *no*.

**IncrementalCitations** *yes | no*

Sets whether the data produced by **index(1)** is in "incremental citations" format. Default is *yes*.

**CompactStorage** *yes | no*

Sets the index storage mode. Default is *yes*.

**UtfStorage** *yes|no*

This parameter has sense only in Unicode version and only for MySQL back-end. In UTF8 storage mode fields **wordurl[1].word** are stored in UTF8 encoding. This mode can reduce sizes of data and index files for **wordurl** table. To convert existing Unicode database to this mode, run **index -b**. Default value is *no*.

### Ispell support parameters

When **ASPseek** is used with ispell support, **searchd(1)** can optionally find all forms for all specified words (example: 'create' -> 'create' OR 'created' OR 'creates'). This scheme retains exact search possibility. Note that only ispell suffixes are supported by now; prefixes are usually change the word meanings, for example if somebody searches for the word **tested** he hardly wants *untested* to be found.

Ispell affixes file contains rules for words and has the following format:

flag V:

```
E      >  -E,IVE      # As in create > creative
[ ^E ] >  IVE         # As in prevent > preventive
```

flag \*N:

```
E      >  -E,ION      # As in create > creation
Y      >  -Y,ICATION # As in multiply > multiplication
[ ^EY ] > EN         # As in fall > fallen
```

Ispell dictionary file contains words themselves and has format like this:

```
wop/S
word/DGJMS
wordage/S
wordbook
wordily
wordless/P
```

Note that if you add ispell support to already existing database, re-indexing is not required.

You may also use ispell flags in this file if you know how to do it. This will allow not to write the same word with different endings to the rare words file, for example "webmaster" and "webmasters". You may choose the word which have the same changing rules from existing ispell dictionary and just to copy flags from it. For example, English dictionary has this line:

```
postmaster/MS
```

So, webmaster with MS flags will be probably OK:

```
webmaster/MS
```

You can get ispell affix and dictionary files for different languages from <http://fmg-www.cs.ucla.edu/fmg-members/geoff/ispell-dictionaries.html>

To make **ASPseek** support ispell the following parameters are used. *lang* argument is two letters language abbreviation. File names used are relative to ASPseek etc directory (`/usr/local/aspseek/etc`). Absolute paths can be also specified.

**Affix** *lang affix-file [charset]*

Load ispell affixes for language *lang* from file *affix-file*. If *charset* is given, file contents is assumed to be in that charset, otherwise the value from **LocalCharset** is used.

**Spell** *lang dict-file [charset]*

Load ispell dictionary for language *lang* from file *dict-file*. If *charset* is given, file contents is assumed to be in that charset, otherwise the value from **LocalCharset** is used.

**WordForms** *on | off | lang[,lang[,...]]*

Sets whether to search for different word forms by default. Argument can be *on*, *off*, or comma-separated list of languages. Value can be overridden by **fm** parameter of **s.cgi**(1).

**Ranking parameters****SiteWeight** *http://www.site.com nnn*

Specifies the priority for particular site. Default priority for all sites is 0. If priority of site is greater than 0, then it will always be displayed before all the other results. If priority of site is less than 0, then it will always be displayed after all the other results.

**AccountDistance** *on | off*

Specifies whether **searchd**(1) should take into account distance from the beginning of document section to search terms for ranking calculations. If this parameter is on, then documents with search terms closer to the beginning of section have higher priority over others, otherwise distance doesn't matter. Default is *on*. Value can be overridden by **ad** parameter of **s.cgi**(1).

**Results cache parameters**

**searchd** can implement results cache, so results for next page queries and for queries that are the same as were before will be taken from cache. The following parameters are used.

**Cache** *on*

If this line is present, results cache will be enabled. By default cache is disabled.

**CacheLocalSize** *number*

Size of cache, in entries (one entry for one query). Default value is *100*.

**CachedUrls** *number*

Number of resulting URLs to be stored in one cache entry. Default value is *200*.

**Charset configuration for non-Unicode version**

Charset configuration for non-Unicode version is usually stored in file `/usr/local/aspseek/etc/charsets.conf`. Charset files for non-Unicode version can be found in `/usr/local/aspseek/etc/charsets` directory. Langmap files can be found in `/usr/local/aspseek/etc/langmap` directory.

**CharsetTable** *charset lang file [lfile]*

Loads the table for *charset* of language *lang* from *file*. Optionally load langmap file *lfile*, which is used for charset guesser.

**CharsetAlias** *charset alias1 [alias2...]*

Defines *alias1*, *alias2*, ... as aliases (alternative names) for *charset*. This is needed because in many cases there is no "one true name" for the charset - different web servers and page authors use different names.

**LocalCharset** *charset*

Sets the local charset for ASPseek, so all data in the database is assumed to be in that charset.

**Charset configuration for Unicode version**

Charset configuration for Unicode version is usually stored in file `/usr/local/aspseek/etc/ucharset.conf`. Charset files for Unicode version can be found in `/usr/local/aspseek/etc/tables` directory.

**CharsetTableU1** *charset lang file [lfile]*

Loads the Unicode mapping for *charset* of language *lang* from *file*. Optionally load langmap file *lfile*, which is used for charset guesser.

**CharsetTableU2** *charset lang file [lfile]*

Loads the Unicode mapping for multibyte *charset* of language *lang* from *file*. Optionally load langmap file *lfile*, which is used for charset guesser.

**Dictionary2** *lang file [charset]*

Loads dictionary for *lang* from *file*. If *charset* is not specified, it is assumed that the file is in Unicode. Dictionary is used for tokenizing of text in Chinese, Japanese and Korean languages.

**Stopwords**

Stopwords configuration is usually stored in file `/usr/local/aspseek/etc/stopwords.conf`. Stopword files for different languages can be found in `/usr/local/aspseek/etc/stopwords` directory.

**StopwordFile** *lang file [charset]*

Loads stopwords for language *lang* from *file*. If *charset* is not specified, file contents is assumed to be in **LocalCharset**, otherwise it is in *charset*.

**FILES**

`/usr/local/aspseek/etc/searchd.conf`  
`/usr/local/aspseek/etc/charsets.conf`  
`/usr/local/aspseek/etc/ucharset.conf`  
`/usr/local/aspseek/etc/stopwords.conf`

**BUGS**

Many parameters are the same in **searchd.conf** and in **aspseek.conf(5)**.

**SEE ALSO**

**searchd(1)**, **aspseek.conf(5)**.

**AUTHORS**

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

Man page by Kir Kolyshkin <kir@aslinux.ru>