

Name

indxbib – make inverted index for bibliographic databases

Synopsis

indxbib [-w] [-c *common-words-file*] [-d *dir*] [-f *list-file*] [-h *min-hash-table-size*] [-i *excluded-fields*]
 [-k *max-keys-per-record*] [-l *min-key-length*] [-n *threshold*] [-o *file*] [-t *max-key-length*]
 [*file* ...]

indxbib --help

indxbib -v

indxbib --version

Description

indxbib makes an inverted index for the bibliographic databases in each *file* for use with *refer*(1), *lookbib*(1), and *lkbib*(1). Each created index is named *file.i*; writing is done to a temporary file which is then renamed to this. If no *file* operands are given on the command line because the **-f** option has been used, and no **-o** option is given, the index will be named *Ind.i*.

Bibliographic databases are divided into records by blank lines. Within a record, each field starts with a % character at the beginning of a line. Fields have a one letter name that follows the % character.

The values set by the **-c**, **-l**, **-n**, and **-t** options are stored in the index: when the index is searched, keys will be discarded and truncated in a manner appropriate to these options; the original keys will be used for verifying that any record found using the index actually contains the keys. This means that a user of an index need not know whether these options were used in the creation of the index, provided that not all the keys to be searched for would have been discarded during indexing and that the user supplies at least the part of each key that would have remained after being truncated during indexing. The value set by the **-i** option is also stored in the index and will be used in verifying records found using the index.

Options

--help displays a usage message, while **-v** and **--version** show version information; all exit afterward.

-c *common-words-file*

Read the list of common words from *common-words-file* instead of */usr/local/share/groff/1.23.0/eign*.

-d *dir* Use *dir* as the name of the directory to store in the index, instead of that returned by *getcwd*(2). Typically, *dir* will be a symbolic link whose target is the current working directory.

-f *list-file*

Read the files to be indexed from *list-file*. If *list-file* is **-**, files will be read from the standard input stream. The **-f** option can be given at most once.

-h *min-hash-table-size*

Use the first prime number greater than or equal to the argument for the size of the hash table. Larger values will usually make searching faster, but will make the index file larger and cause *indxbib* to use more memory. The default hash table size is 997.

-i *excluded-fields*

Don't index the contents of fields whose names are in *excluded-fields*. Field names are one character each. If this option is not present, *indxbib* excludes fields **X**, **Y**, and **Z**.

-k *max-keys-per-record*

Use no more keys per input record than specified in the argument. If this option is not present, the maximum is 100.

-l *min-key-length*

Discard any key whose length in characters is shorter than the value of the argument. If this option is not present, the minimum key length is 3.

-n *threshold*

Discard the *threshold* most common words from the common words file. If this option is not present, the 100 most common words are discarded.

-o *basename*

Name the index *basename.i*.

-t *max-key-length*

Truncate keys to *max-key-length* in characters. If this option is not present, keys are truncated to 6 characters.

-w Index whole files. Each file is a separate record.**Files**

file.i index for *file*

Ind.i default index name

/usr/local/share/groff/1.23.0/eign

contains the list of common words. The traditional name, “*eign*”, is an abbreviation of “English ignored [word list]”.

indxbibXXXXXX

temporary file

See also

“Some Applications of Inverted Indexes on the Unix System”, by M. E. Lesk, 1978, AT&T Bell Laboratories Computing Science Technical Report No. 69.

refer(1), *lkbib(1)*, *lookbib(1)*