

id-utils

COLLABORATORS

	<i>TITLE :</i> id-utils		
<i>ACTION</i>	<i>NAME</i>	<i>DATE</i>	<i>SIGNATURE</i>
WRITTEN BY		December 7, 2024	

REVISION HISTORY

NUMBER	DATE	DESCRIPTION	NAME

Contents

1	id-utils	1
1.1	id-utils.guide	1
1.2	id-utils.guide/Introduction	1
1.3	id-utils.guide/Quick start	2
1.4	id-utils.guide/Common options	3
1.5	id-utils.guide/Universal options	3
1.6	id-utils.guide/Reading options	4
1.7	id-utils.guide/Writing options	4
1.8	id-utils.guide/Walker options	4
1.9	id-utils.guide/File listing options	5
1.10	id-utils.guide/Extraction options	5
1.11	id-utils.guide/Language map	6
1.12	id-utils.guide/C-C++ scanner	8
1.13	id-utils.guide/Assembler scanner	9
1.14	id-utils.guide/Text scanner	10
1.15	id-utils.guide/Defining scanners	10
1.16	id-utils.guide/mkid invocation	11
1.17	id-utils.guide/lid invocation	11
1.18	id-utils.guide/lid aliases	14
1.19	id-utils.guide/Emacs gid interface	14
1.20	id-utils.guide/eid invocation	15
1.21	id-utils.guide/fid invocation	16
1.22	id-utils.guide/fnid invocation	16
1.23	id-utils.guide/xtokid invocation	17
1.24	id-utils.guide/Past and Future	17
1.25	id-utils.guide/Index	18

Chapter 1

id-utils

1.1 id-utils.guide

ID utilities

This manual documents version 3.2 of the ID utilities.

Introduction	Overview of the tools with tutorial.
Quick start	Quick start procedure.
Common options	Common command-line options.
mkid invocation	Creating an ID database.
lid invocation	Querying an ID database by token.
fid invocation	Listing a file's tokens.
fnid invocation	Looking up file names.
xtokid invocation	Testing language scanners.
Past and Future	History and future directions.
Index	General index.

1.2 id-utils.guide/Introduction

Introduction

An "ID database" is a binary file containing a list of file names, a list of tokens, and a sparse matrix indicating which tokens appear in which files.

With this database and some tools to query it (described in this manual), many text-searching tasks become simpler and faster. For example, you can list all files that reference a particular `'#include'` file throughout a huge source hierarchy, search for all the memos containing references to a project, or automatically invoke an editor on all files containing references to some function or variable. Anyone with a large software project to maintain, or a large set of text files to organize, can benefit from the ID utilities.

Although the name 'ID' is short for 'identifier', the ID utilities handle more than just identifiers; they also treat other kinds of tokens, most notably numeric constants, and the contents of certain character strings. Thus, this manual will use the word "token" as a term that is inclusive of identifiers, numbers and strings.

There are several programs in the ID utilities family:

`'mkid'`

scans files for tokens and builds the ID database file.

`'lid'`

queries the ID database for tokens, then reports matching file names or matching lines.

`'fid'`

lists all tokens recorded in the database for given files, or tokens common to two files.

`'fnid'`

matches the file names in the database, rather than the tokens.

`'xtokid'`

extracts raw tokens--helps with testing of new 'mkid' scanners.

In addition, the ID utilities have historically provided several query programs which are specializations of 'lid':

`'gid'`

(alias for 'lid -R grep') lists all lines containing the requested pattern.

`'eid'`

(alias for 'lid -R edit') invokes an editor on all files containing the requested pattern, and if possible, initiates a text search for that pattern.

`'aid'`

(alias for 'lid -ils') treats the requested pattern as a case-insensitive literal substring.

Please report bugs to 'bug-gnu-utils@gnu.ai.mit.edu'. Remember to include the version number, machine architecture, input files, and any other information needed to reproduce the bug: your input, what you expected, what you got, and why it is wrong. Diffs are welcome, but please include a description of the problem as well, since this is sometimes difficult to infer. See Bugs.

1.3 id-utils.guide/Quick start

Quick Start Procedure

Unpack the distribution.

Type `./configure`

Type `make`

Type `make install` as a user with the appropriate privileges (e.g., `bin` or perhaps even `root`).

Type `cd /usr/include; mkid` to build an ID database covering all of the system header files.

Type `lid FILE`, then `gid strtok`, then `aid stdout`.

You have just built, installed and used the most common commands of the GNU ID utilities. If you ever need help remembering which system header files contain a particular declaration, or reference a particular symbol, you'll want to keep the ID file you built in `/usr/include` for later use. If your working directory is elsewhere at the time, simply provide the `-f /usr/include` option to `lid` (see Reading options).

1.4 id-utils.guide/Common options

Common command-line options

Certain options, and regular expression syntax, are shared by various groupings of the ID utilities. We describe these in the sections below, rather than repeating them for each program.

Universal options	Options common to all programs.
Extraction options	Options for programs that extract tokens from source ↔ files.
Walker options	Options for programs that walk file and directory ↔ trees.
Reading options	Options for programs that read ID databases.
Writing options	Options for programs that write ID databases.
File listing options	Options for programs that list file names.

1.5 id-utils.guide/Universal options

Options Common to All Programs

=====

`--help`

Print a usage message listing all available options, then exit successfully.

'--version'

Print the version number, then exit successfully.

1.6 id-utils.guide/Reading options

Options for Programs that Read ID Databases

=====

'-f FILENAME'

'--file=FILENAME'

FILENAME is the ID database to read when processing queries. At present, only a single '--file' option is processed, but in future releases, more than one ID database may be named on the command line.

'\$IDPATH'

'IDPATH' is an environment variable that contains a colon-separated list of ID database names. If this variable is present, and no '--file' options are presented on the command line, the ID databases named in 'IDPATH' are implied.(1)

If no ID databases are specified either on the command line or via the 'IDPATH' environment variable, then the ID utilities search for a file named 'ID' in the current working directory, and then in successive parent directories.

----- Footnotes -----

(1) At present, this feature is fully implemented, since only the first of a list of ID database names is processed.

1.7 id-utils.guide/Writing options

Options for Programs that Write ID Databases

=====

'-o FILENAME'

'--output=FILENAME'

The '--output' option names the file in which to write a new ID database. If no '--output' (or '--file') option is present, an output file named 'ID' is implied.

'-f FILENAME'

'--file=FILENAME'

This is a synonym for '--output'

1.8 id-utils.guide/Walker options

Options for Programs that Walk File and Directory Trees.

=====

The programs 'mkid' and 'xtokid' accept the names of files and directories on the command line. Files are scanned if there is a scanner available and enabled for the file's source language. Directories are recursively descended, searching for files whose names match the rules listed in the *language map* file (see Language map).

The following option controls the file tree walker:

'-p NAMES'

'--prune=NAMES'

One or more file or directory names may appear in NAMES. The file tree walker will stop short at these files and directories and their contents will not be scanned.

1.9 id-utils.guide/File listing options

Options for Programs that List File Names

=====

The programs 'lid' and 'fnid' can print lists of file names as the result of queries. The following option controls how these lists are formatted:

'-S STYLE'

'--separator=STYLE'

STYLE may be one of 'braces', 'space' or 'newline'.

The STYLE of 'braces' means that file names with common directory prefix and common suffix are printed using the shell's brace notation in order to compress the output. For example, './src/foo.c ./src/bar.c' can be printed in brace notation as './src/{foo,bar}.c'.

The STYLEs of 'space' and 'newline' mean that file names are separated spaces or by newlines, respectively.

If the list of files is being printed on a terminal, brace notation is the default. If not, file names are separated by spaces if the KEY is included in the output, and by newlines the KEY STYLE is 'none' (see lid invocation).

1.10 id-utils.guide/Extraction options

Options for Programs that Scan Source Files

=====

'mkid' and 'xtokid' walk file trees, select source files by name, and extract tokens from source files. They accept the following options:

```
'-m MAPFILE'
'--lang-map=MAPFILE'
    MAPFILE contains rules for determining the source languages from
    file names. See Language map

'-i LANGUAGES'
'--include=LANGUAGES'
    The '--include' option names LANGUAGES whose source files should
    be scanned and incorporated into the ID database. By default, all
    languages known to the ID utilities are enabled.

'-x LANGUAGES'
'--exclude=LANGUAGES'
    The '--exclude' option names LANGUAGES whose source files should
    NOT be scanned. The default list of excluded languages is empty.
    Note that only one of '--include' or '--exclude' may be specified
    on the command line for a single run.

'-l LANGUAGE:OPTIONS'
'--lang-option=LANGUAGE:OPTIONS'
    Language-specific scanners also accept options. LANGUAGE denotes
    the desired scanner, and OPTION are the command-line options that
    should be passed through to it. For example, to pass the -X
    -COKE-BOTTLE options to the scanner for the language SWIZZLE, pass
    this: -L SWIZZLE:"-X -COKE-BOTTLE", or this:
    -LANG-OPTION=SWIZZLE:"-X -COKE-BOTTLE", or this: -L SWIZZLE-X -L
    SWIZZLE:-COKE-BOTTLE. Use the '--help' option to see the
    command-line option summary for
```

To determine which tokens to extract from a file and store in the database, 'mkid' calls a "scanner"; we say a scanner "recognizes" a particular language. Scanners for several languages are built-in to 'mkid'; you can add your own scanners as well, as explained in See Defining scanners.

The ID utilities determine which scanner to use for a particular file by consulting the language-map file. Scanners for several are already built-in to the ID utilities. You can see which languages have built-in scanners, and examine their language-specific options by invoking 'mkid --help' or 'xtokid --help'.

Language map	Mapping file names to source languages.
C-C++ scanner	For the C and C++ programming language.
Assembler scanner	For assembly language.
Text scanner	For documents or other non-source code.
Defining scanners	Defining new scanners in the source code.

1.11 id-utils.guide/Language map

Mapping file names to source languages

The file `'id-lang.map'`, installed by default in `'$(prefix)/share/id-lang.map'`, contains rules for mapping file names to source languages. Each rule comprises three parts: a shell GLOB pattern, a language name, and language-specific scanner options.

The special pattern `'**'` denotes the default source language. This is the language that's assigned to file names that don't match any other pattern.

The special pattern `'***'` should be followed by a file name. The named file should contain more language-map rules and is included at this point.

The order in which rules are presented in a language-map file is significant. This order influences the order in which files are displayed as the result of queries. For example, the distributed language-map file places all rules for C `.H` files ahead of `.C` files, so that in general, declarations will precede definitions in query output. The same thing is done for C++ and its many different source file name extensions.

Here is a pared-down version of the `'id-lang.map'` file distributed with the ID utilities:

```
# Default language
**      IGNORE # Although this is listed first,
           # the default language pattern is
           # logically matched last.

# Backup files
*~      IGNORE
*.bak    IGNORE
*.bk[0-9] IGNORE

# SCCS files
[sp].*   IGNORE

# list header files before code files
*.h      C
*.h.in    C
*.H      C++
*.hh     C++
*.hpp    C++
*.hxx    C++

# list C 'meta' files next
*.l      C
*.lex    C
*.y      C
*.yacc   C

# list C code files after header files
```

```

*.c      C
*.C      C++
*.cc     C++
*.cpp    C++
*.cxx    C++

# list assembly language after C
*.[sS]    asm --comment=;
*.asm     asm --comment=;

# [nt]roff
*.[0-9]    roff
*.ms       roff
*.me       roff
*.mm       roff

# TeX and friends
*.tex      TeX
*.ltx      TeX
*.texi     texinfo
*.texinfo  texinfo

```

1.12 id-utils.guide/C-C++ scanner

C/C++ Language Scanner

The C scanner is the most commonly used. Files that match the glob pattern `*.h`, `*.c`, as well as `yacc` files that match `*.y` or `*.yacc`, and `lex` files that match `*.l` or `*.lex`, are processed with this scanner.

Scanner-specific options (Note, these options are presented WITHOUT the required `-l` or `--lang-option=` prefix):

`-k CHARACTER-CLASS`

`--keep=CHARACTER-CLASS`

Consider the characters in CHARACTER-CLASS as valid constituents of identifier names. For example, if you are indexing C code that contains `$` in some of its identifiers, you can include these by using `--lang-option=C:--keep=$`, or `-l C:"-k $"` (if you don't like to type so much).

`-i CHARACTER-CLASS`

`--ignore=CHARACTER-CLASS`

Consider the characters in CHARACTER-CLASS as valid constituents of identifier names, but discard all tokens containing these characters. For example, if some C code has identifiers containing `$`, but you don't want these cluttering up your ID database, use `--lang-option=C:--ignore=$`, or the terser equivalent `-l C:"-i $"`.

`-u`

`--strip-underscore`

Strip one leading underscore from C identifiers encapsulated as character strings. This option is useful if you are indexing C code that contains symbol-table name strings for systems that prepend an underscore to external symbols. By default, the leading underscore is retained.

1.13 id-utils.guide/Assembler scanner

Assembly Language Scanner

Assembly languages use a variety of commenting conventions, and allow a variety of special characters to *dirty up* local symbols, preventing name space conflicts with symbols defined by higher-level languages. Also, some compilation systems prepend an underscore to external symbols. The options listed below are designed to address these differences.

`'-c CHARACTER-CLASS'`

`'--comment=CHARACTER-CLASS'`

The characters in CHARACTER-CLASS are considered left delimiters for comments that extend until the end of the current line.

`'-k CHARACTER-CLASS'`

`'--keep=CHARACTER-CLASS'`

Consider the characters of CHARACTER-CLASS as valid constituents of identifier names. For example, if you are indexing assembly code that prepends `'.'` to assembler directives, and prepends `'%'` to register names, you can keep these characters in the tokens by specifying `'--lang-option=asm:--keep=.%'`, or `'-l asm:"-k .%"'`.

`'-i CHARACTER-CLASS'`

`'--ignore=CHARACTER-CLASS'`

Consider the characters of CHARACTER-CLASS as valid constituents of identifier names, but discard all tokens containing these characters. For example, if you don't want to clutter your ID database with assembler directives that begin with a leading `'.'` or with assembler labels that contain `'@'`, use `'--lang-option=asm:--ignore=.%'`, or `'-l asm:"-i .@"'`.

`'-u'`

`'--strip-underscore'`

Strip one leading underscore from identifiers. This option is useful if your compilation system prepends an underscore to external symbols. By stripping the underscore, you can canonicalize such names and bring them into conformance the way they are expressed in the C language. By default, the leading underscore is retained.

`'-n'`

`'--no-cpp'`

Do not recognize C preprocessor directives. By default, such lines are handled in the same way as they are by the C language scanner.

1.14 id-utils.guide/Text scanner

Text Scanner

The plain text scanner is intended for human-language documents, or as the scanner of last resort for files that have no scanner that is more specific. It is customizable to the extent that character classes can be designated as token constituents or as token delimiters. The default token constituents are the alpha-numeric; all other characters are considered token delimiters.

```
'-i CHARACTER-CLASS'
'--include=CHARACTER-CLASS'
    Include characters belonging to CHARACTER-CLASS in tokens.

'-x CHARACTER-CLASS'
'--exclude=CHARACTER-CLASS'
    Exclude characters belonging to CHARACTER-CLASS from tokens, i.e.,
    treat them as token delimiters.
```

1.15 id-utils.guide/Defining scanners

Defining New Scanners in the Source Code

To add a new scanner in source code, you should add a new section to the file 'scanners.c'. It might be easiest to clone one of the existing scanners and modify it as necessary. For the hypothetical language FOO, you must define the functions 'get_token_foo', 'parse_args_foo', 'help_me_foo', as well as the tables 'long_options_foo' and 'args_foo'. If your scanner is modelled after one of the existing scanners, you'll also need a character-attribute table 'ctype_foo'.

This is not a terribly difficult programming task, but it requires recompiling and installing the new version of 'mkid' and 'xtokid'. You should use 'xtokid' to test the operation of the new scanner.

Once these functions and tables are ready, add function prototypes and an entry to to the 'languages_0' table near the beginning of the file.

Be warned that the existing scanners are built for speed, not elegance or readability. You might wish to create a new scanner that's easier to read and understand if you don't feel that speed is so important.

1.16 id-utils.guide/mkid invocation

`'mkid'`: Creating an ID Database

`'mkid'` builds an ID database. It accepts the names of files and/or directories on the command line, selects files that have an enabled scanner, then extracts and stores tokens from those files. The resulting ID database is architecture- and byte-order-independent so it can be shared among all systems.

The primary virtues of `'mkid'` are speed and high capacity. The size of the source trees it can index is limited only by available system memory. `'mkid'`'s indexing algorithm is very space-efficient and exhibits excellent locality-of-reference, and so is capable of operating with a working-set size that is only half the size of its virtual address space. A typical UNIX-like operating system with 16 megabytes of system memory should be able to build an ID database covering approximately 12,000-14,000 source files totalling approximately 50-100 Megabytes. A 66 Mhz 486 computer can build such a large ID database in approximately 10-15 minutes.

In a future release, `'mkid'` will be able to incrementally update an ID database much faster than it can build one from scratch. Until this feature becomes available, it might be a good idea to schedule a `'cron'` job to regularly update large ID databases during off-hours.

`'mkid'` writes the ID file, therefore it accepts the `'--output'` (and `'--file'`) options as described in See Writing options. `'mkid'` extracts tokens from source files, therefore it accepts the `'--lang-map'`, `'--include'`, `'--exclude'`, and `'--lang-option'` options, as well as the language-specific scanner options, all of which are described in See Extraction options. `'mkid'` walks file trees, therefore it handles file and directory names on its command line and the `'--prune'` option as described in See Walker options.

In addition, `'mkid'` accepts the following command-line options:

`'-s'`

`'--statistics'`

`'mkid'` reports statistics about resource usage at the end of its run.

`'-v'`

`'--verbose'`

`'mkid'` reports statistics about each file as it is scanned, and about the resource usage of its indexing algorithm at regular intervals.

1.17 id-utils.guide/lid invocation

`'lid'`: Querying an ID Database by Token

The 'lid' program accepts PATTERNS on the command line which it matches against the tokens stored in an ID database. The interpretation of a PATTERN is determined by the makeup of the PATTERN string itself, or can be overridden by command-line options. If a PATTERN contains regular expression meta-characters, it is used to perform a regular-expression substring search. If no such meta-characters are present, PATTERN is used to perform a literal word search. (By default, all searches are sensitive to alphabetic case.) If no PATTERN is supplied on the command line, 'lid' lists every entry in the ID database.

'lid' reads the ID database, therefore it accepts the '--file' option, and consults the 'IDPATH' environment variable, as described in See Reading options. 'lid' lists file names, therefore it accepts the '--separator' option, as described in See File listing options.

In addition, 'lid' accepts the following command-line options:

'-i'

'--ignore-case'

Ignoring differences in alphabetic case between the PATTERN and the tokens in the ID database.

'-l'

'--literal'

Match PATTERN as a literal string. Use this option if PATTERN contains regular-expression meta-characters, but you don't wish to perform a regular-expression search.

'-r'

'--regexp'

Match PATTERN as an *extended* regular expression(1). Use this option if no regular-expression meta-characters are present in PATTERN, but you wish to force a regular-expression search (note: in this case, a *literal substring* search might be faster).

'-w'

'--word'

Match PATTERN using a word-delimited (non substring) search. This is the default for literal searches.

'-s'

'--substring'

Match PATTERN using a substring (non word-delimited) search. This is the default for regular expression searches.

'-k STYLE'

'--key=STYLE'

STYLE can be one of 'token', 'pattern' or 'none'. This option controls how the subject of the query is presented. This is best illustrated by example:

```
$ lid --key=token '^dest.'
```

destaddr	libsys/memcpy.c
destination	libsys/regex.c

```

destlst      libsys/rx.c
destpos      libsys/rx.c
destset      libsys/rx.h libsys/rx.c

$ lid --key=pattern '^dest.'
^dest.       libsys/rx.h libsys/{memcpy,regex,rx}.c

$ lid --key=none '^dest.'
libsys/rx.h libsys/{memcpy,regex,rx}.c

```

When `--key` is either `'token'` or `'pattern'`, the first column of output is a TOKEN or PATTERN, respectively. When `--key` is `'none'`, neither of these is printed, and the file name list begins immediately. The default is `'token'`.

`'-R STYLE'`

`'--result=STYLE'`

STYLE can be one of `'filenames'`, `'grep'`, `'edit'` or `'none'`. This option controls how the value associated with the query's KEY is presented. When STYLE is `'filenames'`, a list of file names is printed (this is the default). When STYLE is `'grep'`, the lines that match PATTERN are printed in the same format as `'egrep -n'`. When STYLE is `'edit'`, the file names are passed to an editor, and if possible PATTERN is passed as an initial search string (see `eid` invocation). When STYLE is `'none'`, the file names are not processed in any way. This can be useful if you wish to see what tokens match a PATTERN, but don't care about where they reside.

`'-d'`

`'-o'`

`'-x'`

These options may be used in any combination to specify the radix of numeric matches. `'-d'` allows matching on decimal numbers, `'-o'` on octal numbers, and `'-x'` on hexadecimal numbers. Any combination of these options may be used. The default is to match all three radices.

`'-F RANGE'`

`'--frequency=RANGE'`

Match tokens whose occurrence count falls in RANGE. RANGE may be expressed as a single number N, or as a range N`'..'`M. Either limit of the range may be omitted (e.g., `'..'M`, or N`'..'`). If the lower limit N is omitted, it defaults to `'1'`. If the upper limit is omitted, it defaults in the present implementation to `'65535'`, the maximum value of an unsigned 16-bit integer.

Particularly useful queries are `'lid -F1'`, which helps locate identifiers that are defined but never used, or are used but never defined. Similarly, `'lid -F2'` can help find functions that possess a prototype declaration and a definition, but are never called.

`'-a NUMBER'`

`'--ambiguous=NUMBER'`

List identifiers (not numbers) that are ambiguous for the first NUMBER characters. This feature might be useful when porting programs to ancient pea-brained compilers that don't support long identifier names. However, the best long-term option is to set

such systems on fire.

lid aliases	Aliases for specialized lid queries
Emacs gid interface	GNU Emacs query interface
eid invocation	Invoking an editor on query results

----- Footnotes -----

(1) Extended regular expressions are the same as those accepted by 'egrep'.

1.18 id-utils.guide/lid aliases

Aliases for Specialized 'lid' Queries
=====

Historically, the ID utilities have provided several query interfaces which are specializations of 'lid' (see lid invocation).

'gid'
(alias for 'lid -R grep') lists all lines containing the requested pattern.

'eid'
(alias for 'lid -R edit') invokes an editor on all files containing the requested pattern, and optionally initiates a text search for that pattern.

'aid'
(alias for 'lid -ils') treats the requested pattern as a case-insensitive literal substring.

1.19 id-utils.guide/Emacs gid interface

GNU Emacs query interface
=====

The 'id-utils' source distribution comes with a file 'id-utils.el', which defines a GNU Emacs interface to 'gid'. To install it, put 'id-utils.el' somewhere that Emacs will find it (i.e., in your 'load-path') and put

```
(autoload 'gid "gid" nil t)
```

in one of Emacs' initialization files, e.g., '~/.emacs'. You will then be able to use 'M-x gid' to run the command.

The 'gid' function prompts you with the word around point. If you want to search for something else, simply delete the line and type the

pattern of interest.

The function then runs the 'gid' program in a '*compilation*' buffer, so the normal 'next-error' function can be used to visit all the places the identifier is found (see Compilation).

1.20 id-utils.guide/eid invocation

'eid': Invoking an Editor on Query Results

=====

'lid -R edit' is an editing interface for the ID utilities that is most commonly used with 'vi'. Emacs users should use the interface defined in 'id-utils.el' (see Emacs gid interface). The ID utilities include an alias called 'eid', and for the sake of brevity, we'll use this alias for the remainder of this section. 'eid' performs a 'lid'-style, then asks if you wish to edit the files. If your query yields more than one line of output, you will be prompted after each line. This is the prompt you'll see:

```
Edit? [y1-9^S/nq]
```

You may respond with:

'y'

Edit all files listed.

'1...9'

Edit all files starting at the N + 1'st file.

'/STRING or 'CTRL-S'REGEXP'

Search into the file list, and begin editing with the first file name that matches the regular expression REGEXP.

'n'

Don't edit any files. If another line of query output is pending, advance to that line, for which another 'Edit?' prompt will appear.

'q'

Quit--don't edit any files, and don't process any more lines of query output.

Here is an example:

```
prompt$ eid FILE \^print
FILE          {ansi2knr,fid,filenames,idfile,idx,lid,misc,...}.c
Edit? [y1-9^S/nq] n
^print        {ansi2knr,fid,getopt,getopt1,lid,mkid,regex,scanners}.c
Edit? [y1-9^S/nq] 2
```

This will start editing at 'getopt'.c.

'eid' invokes the editor defined by the environment variable 'VISUAL'. If 'VISUAL' is undefined, it uses the environment variable

'EDITOR' instead. If 'EDITOR' is undefined, it defaults to 'vi'. It is possible for 'eid' to pass the editor an initial search pattern so that your cursor will immediately alight on the token of interest. This feature is controlled by the following environment variables:

'EIDARG'

A printf(3) format string for the editor argument to search for the matching token. For 'vi', this should be '+/%s/'.

'EIDLDEL'

The regular-expression meta-character(s) for delimiting the beginning of a word (the 'eid' Left DELimiter'). 'eid' inserts this in front of the matching token when a word-search is desired. For 'vi', this should be '\<'.

'EIDRDEL'

The regular-expression meta-character(s) for delimiting the end of a word (the 'eid' Right DELimiter'). 'eid' inserts this in end of the matching token when a word-search is desired. For 'vi', this should be '\>'.

1.21 id-utils.guide/fid invocation

'fid': Listing a file's tokens

'fid' prints the tokens found in a given file. If two file names are passed on the command line, 'fid' prints the tokens that are common to both files (i.e., the *set intersection* of the two token sets).

'lid' reads the ID database, therefore it accepts the '--file' option, and consults the 'IDPATH' environment variable, as described in See Reading options.

If the standard output is attached to a terminal, the printed tokens are separated by spaces. Otherwise, the tokens are printed one per line.

1.22 id-utils.guide/fnid invocation

'fnid': Looking up filenames

'fnid' queries the list of file names stored in the ID database. It accepts shell *wildcard* patterns on the command line. If no pattern is supplied, '*' is implied. 'fnid' prints the file names that match the given patterns.

'fnid' prints file names, and as such accepts the '--separator' option as described in See File listing options.

For example, the command:

```
fnid \*.c
```

lists all the `\.c` files in the database. (The `\` here protects the `*` from being expanded by the shell.)

1.23 id-utils.guide/xtokid invocation

`'xtokid'`: Testing Language Scanners

`'xtokid'` accepts the names of files and/or directories on the command line, then extracts and prints a stream of tokens from those files for which it has a valid, enabled scanner. This is useful primarily for debugging new `'mkid'` scanners (see Defining scanners).

`'xtokid'` extracts tokens from source files, therefore it accepts the `'--lang-map'`, `'--include'`, `'--exclude'`, and `'--lang-option'` options, as well as the language-specific scanner options, all of which are described in See Extraction options. `'xtokid'` walks file trees, therefore it handles file and directory names on its command line and the `'--prune'` option as described in See Walker options.

The name `'xtokid'` indicates that it is the "eXtract TOKEns ID utility".

1.24 id-utils.guide/Past and Future

Past and Future

Greg McGary conceived of the ideas behind the ID utilities when he began working on the Unix kernel in 1984. He needed a navigation tool to help him find his way around the expansive, unfamiliar landscape. The first `'id-utils'`-like tools were shell scripts, and produced an ASCII database that looks much like the output of `'lid ".*"'`. It took over an hour on a VAX 11/750 to build a database for a 4.1BSD derived kernel. The first version of `'lid'` used the UNIX system utility `'look'`, modified to handle very long lines.

In 1986, Greg rewrote the shell scripts in C to improve performance. Build times for the ID file were shortened by an order of magnitude. The ID utilities were first posted to `'comp.sources.unix'` in September 1987 under the name `'id'`.

Over the next few years, several versions diverged from the original source. Tom Horsley at Harris Computer Systems Division stepped forward to take over maintenance and integrated some of the fixes from divergent

versions. A first release of the renamed 'mkid' version 2 was posted to 'alt.sources' near the end of 1990. At that time, Tom wrote a Texinfo manual with the encouragement the net community. (Tom especially thanks Doug Scofield and Bill Leonard whom he dragooned into helping poorfraed and edit--they found several problems in the initial version.) Karl Berry revamped the manual for Texinfo style, indexing, and organization in 1995.

In January 1995, Greg McGary reemerged as the primary maintainer and launched development of 'mkid' version 3, whose primary new feature is an efficient algorithm for building databases that is linear in both time and space over the size of the input text. (The old algorithm was quadratic in space so it was incapable of handling very large source trees.) For the first time, the code was released under the GNU Public License.

In June 1996, the package was renamed again to 'id-utils' and was released for the first time under FSF copyright as part of the GNU system. All programs had their command-line arguments completely revised. The 'mkid' and 'xtokid' programs also gained a file-tree walker, so that directory names can be passed on the command line instead of the names of every individual file. Greg reorganized and rewrote most of the Texinfo manual to reflect these changes.

Future releases of 'id-utils' might include:

an optional coupling with GNU 'grep', so that 'grep' can use an ID database for hints

a 'cscope' work-alike query interface

incremental update of the ID database.

1.25 id-utils.guide/Index

Index

compilation	Emacs buffer	Emacs gid interface
-ambiguous		lid invocation
-comment		Assembler scanner
-exclude <1>		Extraction options
-exclude		Text scanner
-file <1>		Writing options
-file		Reading options
-frequency		lid invocation
-help		Universal options
-ignore <1>		Assembler scanner
-ignore		C-C++ scanner
-ignore-case		lid invocation
-include <1>		Extraction options
-include		Text scanner
-keep <1>		Assembler scanner
-keep		C-C++ scanner

-lang-map	Extraction options
-lang-option	Extraction options
-lang-option=asm:-comment	Assembler scanner
-lang-option=asm:-ignore	Assembler scanner
-lang-option=asm:-keep	Assembler scanner
-lang-option=asm:-no-cpp	Assembler scanner
-lang-option=asm:-strip-underscore	Assembler scanner
-lang-option=asm:-c	Assembler scanner
-lang-option=asm:-i	Assembler scanner
-lang-option=asm:-k	Assembler scanner
-lang-option=asm:-n	Assembler scanner
-lang-option=asm:-u	Assembler scanner
-lang-option=C:-ignore	C-C++ scanner
-lang-option=C:-keep	C-C++ scanner
-lang-option=C:-strip-underscore	C-C++ scanner
-lang-option=C:-i	C-C++ scanner
-lang-option=C:-k	C-C++ scanner
-lang-option=C:-u	C-C++ scanner
-lang-option=text:-exclude	Text scanner
-lang-option=text:-include	Text scanner
-lang-option=text:-i	Text scanner
-lang-option=text:-x	Text scanner
-literal	lid invocation
-no-cpp	Assembler scanner
-output	Writing options
-prune	Walker options
-regexp	lid invocation
-result	lid invocation
-separator	File listing options
-statistics	mkid invocation
-strip-underscore <1>	Assembler scanner
-strip-underscore	C-C++ scanner
-substring	lid invocation
-verbose	mkid invocation
-version	Universal options
-word	lid invocation
-a	lid invocation
-c	Assembler scanner
-d	lid invocation
-F	lid invocation
-f <1>	Reading options
-f	Writing options
-i <1>	Extraction options
-i <2>	lid invocation
-i <3>	Text scanner
-i <4>	Assembler scanner
-i	C-C++ scanner
-k <1>	Assembler scanner
-k <2>	lid invocation
-k	C-C++ scanner
-l <1>	Extraction options
-l	lid invocation
-l asm:-comment	Assembler scanner
-l asm:-ignore	Assembler scanner
-l asm:-keep	Assembler scanner
-l asm:-no-cpp	Assembler scanner
-l asm:-strip-underscore	Assembler scanner

-l asm:-c	Assembler scanner
-l asm:-i	Assembler scanner
-l asm:-k	Assembler scanner
-l asm:-n	Assembler scanner
-l asm:-u	Assembler scanner
-l C:-ignore	C-C++ scanner
-l C:-keep	C-C++ scanner
-l C:-strip-underscore	C-C++ scanner
-l C:-i	C-C++ scanner
-l C:-k	C-C++ scanner
-l C:-u	C-C++ scanner
-l text:-exclude	Text scanner
-l text:-include	Text scanner
-l text:-i	Text scanner
-l text:-x	Text scanner
-m	Extraction options
-n	Assembler scanner
-o <1>	lid invocation
-o	Writing options
-p	Walker options
-R	lid invocation
-s	lid invocation
-S	File listing options
-s	mkid invocation
-u <1>	C-C++ scanner
-u	Assembler scanner
-v	mkid invocation
-w	lid invocation
-x <1>	Extraction options
-x <2>	lid invocation
-x	Text scanner
mkid progress	mkid invocation
alphabetic case, ignoring differences in	lid invocation
ambiguous identifier names, finding	lid invocation
architecture-independence	mkid invocation
assembler scanner	Assembler scanner
assembly language scanner	Assembler scanner
beginning-of-word editor argument	eid invocation
Berry, Karl	Past and Future
bugs, reporting	Introduction
C scanner, predefined	C-C++ scanner
common command-line options	Common options
creating databases	mkid invocation
cron	mkid invocation
cscope	Past and Future
databases, creating	mkid invocation
eid	eid invocation
EIDARG	eid invocation
EIDLDEL	eid invocation
EIDRDEL	eid invocation
Emacs interface to gid	Emacs gid interface
end-of-word editor argument	eid invocation
exclude languages	Extraction options
fid	fid invocation
file name separator	File listing options
file tree pruning	Walker options
filenames, matching	fnid invocation

fnid	fnid invocation
future	Past and Future
gid Emacs function	Emacs gid interface
grep	Past and Future
help, online	Universal options
history	Past and Future
Horsley, Tom	Past and Future
ID database file name <1>	Reading options
ID database file name <2>	Writing options
ID database file name <3>	Reading options
ID database file name	Writing options
ID database, definition of	Introduction
ID file format	mkid invocation
id-utils.el interface to Emacs	Emacs gid interface
ignoring differences in alphabetic case	lid invocation
include languages	Extraction options
introduction	Introduction
language map file	Extraction options
language-specific option	Extraction options
languages_0	Defining scanners
left delimiter editor argument	eid invocation
Leonard, Bill	Past and Future
load-path	Emacs gid interface
look and mkid 1	Past and Future
matching filenames	fnid invocation
McGary, Greg	Past and Future
numeric matches, specifying radix of	lid invocation
overview	Introduction
radix of numeric matches, specifying	lid invocation
right delimiter editor argument	eid invocation
scanners	Extraction options
scanners, defining in source code	Defining scanners
scanners.c	Defining scanners
Scofield, Doug	Past and Future
search for token, initial	eid invocation
sharing ID files	mkid invocation
single matches, showing	lid invocation
statistics	mkid invocation
text scanner	Text scanner
tokens common to two files	fid invocation
tokens in a file	fid invocation
version number, finding	Universal options
