

NAME

myspell — report spelling exceptions

SYNOPSIS

```
myspell [ -? ] [ -dictionary dictfile ] [ -help ] [ -locale name ] [ -privatedictionary dictfile ]
[ -suffixrules rulefile ] [ -verbose ] [ -version ] [ +dictfile ] [ =rulefile ] [ -- ] [ file(s) ]
```

DESCRIPTION

myspell (pronounced *misspell*) reports spelling exceptions (i.e., words from files, or *stdin*, that are not found in the combined system and private dictionaries, possibly after stripping word suffixes) as a sorted list of unique words, or locators and words, on *stdout*.

A word to be spell checked may contain any ASCII letter or apostrophe, or any character in the range 128..255. All other characters are silently ignored.

With suitable locale-specific dictionaries, and optionally, suffix rules, **myspell** can check spelling for files in any human language that can be encoded in ASCII, or any of the ISO 8859-n code pages, or Unicode in UTF-8 encoding, provided that whitespace separates words. [Languages that lack word separators, such as Lao and Thai, require sophisticated grammatical analysis to identify words.] For Unicode, some prefiltering may be needed to remove Unicode punctuation (otherwise, it will simply be reported as spelling exceptions).

If the files to be spell checked contain document markup, that markup should usually be stripped by a suitable initial filter step; see the **EXAMPLES** section below.

OPTIONS

myspell options can be prefixed with either one or two hyphens, and can be abbreviated to any unique prefix. Thus, **-h**, **-hel**, and **--help** are equivalent.

To avoid confusion with options, if a filename begins with a hyphen, it must be disguised by a leading absolute or relative directory path, e.g., */tmp/-foo* or *./-foo*. Alternatively, precede the file list with the **--** option.

-- Everything following on the command line is a filename, even if it looks like an option.

-? Same as **-help**.

-dictionary *dictfile*

Add *dictfile* to the list of system dictionaries. If a locale is set, there may be locale-dependent dictionaries already in the list. This option may be used any number of times.

If this option is not specified, and no locale is set, then **myspell** will use a built-in list of system dictionaries.

-help Display a brief help message on *stderr*, giving a usage description, and then terminate immediately with a success return code.

-locale *name*

Set the locale temporarily to *name*, which must be an ISO country code and name of a directory in the **myspell** installation tree. This selects the language of the documents to be spell checked.

-privatedictionary *dictfile*

Add the private dictionary *dictfile* to the list of private dictionaries that augment the system dictionary. Typically, this is a document-specific list of exceptional words known to be correctly spelled. This option may be used any number of times.

For **spell**(1) compatibility, this option may be abbreviated to **+dictfile**.

-strip Strip word suffixes according to user-defined or locale-specific rules. This usually reduces the number of false reports.

-suffixrules *rulefile*

Supply additional suffix rules in *rulefile*.

- This option may be abbreviated to `=rulefile`.
- verbose** Include a location report of the form *filename:linenumber:* as a prefix of every spelling exception, and sort the report by location. This option may be abbreviated to a single letter.
 - version** Display the program version number and release date on *stderr*, and then terminate immediately with a success return code.
 - +dictfile** Shorthand for **-privatedictionary** *dictfile*. This option may be used any number of times (unlike in **spell**(1)).
 - =rulefile** Shorthand for **-suffixrules** *rulefile*. This option may be used any number of times.

EXAMPLES

In these examples, we use file suffixes of `.ser` (spelling errors) for exception lists, and `.sok` (spelling okay) for private dictionaries, but these are merely conventions, without significance for **myspell**.

```
myspell report.txt > report.ser
myspell +report.sok report.txt > report.ser
deroff *.rno | myspell -s french.sfx > temp.ser
detex *.tex | myspell -p mydict.sok > temp.ser
dehtml *.html | myspell -l fr =french.sfx > temp.ser
dexml *.xml | myspell -locale da > temp.ser
dexml *.xml | myspell -l da =danish.sfx > temp.ser
```

DICTIONARIES

Dictionaries are simply lists of words known to be correctly spelled, stored one word per line, without any leading or trailing whitespace. Unlike dictionaries for other spell checkers, those for **myspell** need not be sorted. However, if dictionaries are to be shared between spell checkers, they should be kept sorted, and for each language, the locale used for the sort must be consistent.

Once the input is free of spelling errors, the output of **myspell** will be a list of exceptional words that are not in the current dictionaries, but are known to be correct. They can be added to a private dictionary that is used on subsequent runs, thereby reducing the size of future reports.

There are numerous sources of word lists for various languages on the Internet (search for *word list* with your favorite search engine), e.g.,

```
ftp://ftp.ox.ac.uk/pub/wordlists/
ftp://ibiblio.org/pub/docs/books/gutenberg/etext96/pgw*
ftp://qiclab.scn.rain.com/pub/wordlists/
http://www.phreak.org/html/wordlists.shtml
```

Dictionaries for other spell checkers can usually be trivially adapted for use with **myspell**.

In addition, any corpus of text in a single language that is known to be relatively free of errors can be easily filtered with **tr**(1) and **sort**(1) to produce a candidate spelling dictionary for any language that is not yet supported by **myspell**. Internet archives of articles, books, reports, theses, and even news stories can often be readily located by Web search engines.

SUFFIX RULES

Suffix rules guide reduction of the input word lists to reduce dictionary sizes and reduce false reports. As such, they are *entirely optional*.

Suffix rules are defined in simple text files that contain one rule per line, beginning with a suffix regular expression, and followed by a possibly-empty list of replacement suffixes, one of which may be the empty string, indicated by adjacent quotation marks. Comments run from sharp (#) to end of line, and blank lines are ignored.

Here is a short example for English:

```
'$                # Jones' -> Jones
's$              # it's -> it
ed$             " " e    # breaded -> bread, flamed -> flame
ied$            ie y     # died -> die, cried -> cry
```

```

ly$      " "      # acutely -> acute
s$       # cats -> cat

```

While suffix rules suffice for many Indo-European languages, others don't need them at all, and still others have more complex changes in spelling as words change in case, number, or tense. For such languages, the simplest solution seems to be a larger dictionary that incorporates at least all of the common word forms.

FILES

/usr/local/share/myspell/myspell-x.y.z/locale/XX/.dict*

Default dictionaries for locale XX.

/usr/local/share/myspell/myspell-x.y.z/locale/XX/.sfx*

Default suffix rules for locale XX.

/usr/local/share/myspell/myspell-x.y.z/spell.awk

Spell checker source program.

SEE ALSO

aspell(1), **dehtml(1)**, **deroff(1)**, **desgml(1)**, **detex(1)**, **dexml(1)**, **ispell(1)**, **locale(1)**, **sort(1)**, **spell(1)**, **tr(1)**.

BUGS

No options are provided to select language variants, such as American, Canadian, and British English. These can still be handled with supplemental dictionaries specified with the **-dictionary** or **-privatedictionary** options.

Much more work needs to be done to provide language-specific suffix-rule files, and to collect dictionaries for many more languages.

AUTHOR

Nelson H. F. Beebe
 Center for Scientific Computing
 University of Utah
 Department of Mathematics, 110 LCB
 155 S 1400 E RM 233
 Salt Lake City, UT 84112-0090
 Tel: +1 801 581 5254
 FAX: +1 801 581 4148
 Email: beebe@math.utah.edu, beebe@acm.org,
 beebe@ieee.org, beebe@computer.org (Internet)
 WWW URL: <http://www.math.utah.edu/~beebe>