NAME

diction – print wordy and commonly misused phrases in sentences

SYNOPSIS

diction [**−b**] [**−d**] [**−f** *file* [**−n**|**−L** *language*]] [*file...*]
diction [**−−beginner**] [**−−ignore-double-words**] [**−−file** *file* [**−−no-default-file**|**−−language** *language*]] [*file...*]
diction **−h**|**−−help**
diction **−−version**

DESCRIPTION

Diction finds all sentences in a document that contain phrases from a database of frequently misused, bad or wordy diction. It further checks for double words. If no files are given, the document is read from standard input. Each found phrase is enclosed in **[ ]** (brackets). Suggestions and advice, if any and if asked for, are printed headed by a right arrow **->**. A sentence is a sequence of words, that starts with a capitalised word and ends with a full stop, double colon, question mark or exclaimation mark. A single letter followed by a dot is considered an abbreviation, so it does not terminate a sentence. Various multi-letter abbreviations are recognized, they do not terminate a sentence as well, neither do fractional numbers.

Diction understands *cpp*(1) **#line** lines for being able to give precise locations when printing sentences.

OPTIONS

−b, **−−beginner**
> Complain about mistakes typically made by beginners.

−d, **−−ignore-double-words**
> Ignore double words and do not complain about them.

−s, **−−suggest**
> Suggest better wording, if any.

−f *file*, **−−file** *file*
> Read the user specified database from the specified *file* in addition to the default database.

−n, **−−no-default-file**
> Do not read the default database, so only the user-specified database is used.

−L *language*, **−−language** *language*
> Set the phrase file language (de, **en**, **nl**).

−h, **−−help**
> Print a short usage message.

−−version
> Print the version.

ERRORS

On usage errors, 1 is returned. Termination caused by lack of memory is signalled by exit code 2.

EXAMPLE

The following example first removes all roff constructs and headers from a document and feeds the result to diction with a German database:

    deroff -s file.mm | diction -L de | fmt

ENVIRONMENT

LC_MESSAGES=de|en|nl
> specifies the message language and is also used as default for the phrase language. The default language is **en**.

FILES

${prefix}/share/diction/*   databases for various languages

AUTHOR

This program is GNU software, copyright 1997–2007 Michael Haardt <michael@moria.de>.

The english phrase file contains contributions by Greg Lindahl <lindahl@pbm.com>, Wil Baden, Gary D. Kline, Kimberly Hanks and Beth Morris. The dutch phrase file was contributed by Hans Lodder.

HISTORY

There has been a diction command on old UNIX systems, which is now part of the AT&T DWB package. The original version was bound to roff by enforcing a call to deroff. This version is a reimplementation and must run in a pipe with **deroff**(1) if you want to process roff documents. Similarly, you can run it in a pipe with **dehtml**(1) or **detex**(1) to process HTML or TeX documents.

SEE ALSO

deroff(1), fmt(1), style(1)

Cherry, L.L.; Vesterman, W.: *Writing Tools—The STYLE and DICTION programs*, Computer Science Technical Report 91, Bell Laboratories, Murray Hill, N.J. (1981), republished as part of the 4.4BSD User's Supplementary Documents by O'Reilly.

Strunk, William: *The elements of style*, Ithaca, N.Y.: Priv. print., 1918, http://coba.shsu.edu/help/strunk/

NAME

style − analyse surface characteristics of a document

SYNOPSIS

style [**−L** *language*] [**−l** *length*] [**−r** *ari*] [*file*...]

style [**−−language** *language*] [**−−print-long** *length*] [**−−print-ari** *ari*] [*file*...]

style −h|−−help

style −−version

DESCRIPTION

Style analyses the surface characteristics of the writing style of a document. It prints various readability grades, length of words, sentences and paragraphs. It can further locate sentences with certain characteristics. If no files are given, the document is read from standard input.

Numbers are counted as words with one syllable. A sentence is a sequence of words, that starts with a capitalised word and ends with a full stop, double colon, question mark or exclamation mark. A single letter followed by a dot is considered an abbreviation, so it does not end a sentence. Various multi-letter abbreviations are recognized, they do not end a sentence as well. A paragraph consists of two or more new line characters.

Readability grades

Style understands *cpp*(1) **#line** lines for being able to give precise locations when printing sentences.

Kincaid formula

The Kincaid Formula has been developed for Navy training manuals, that ranged in difficulty from 5.5 to 16.3. It is probably best applied to technical documents, because it is based on adult training manuals rather than school book text. Dialogs (often found in fictional texts) are usually a series of short sentences, which lowers the score. On the other hand, scientific texts with many long scientific terms are rated higher, although they are not necessarily harder to read for people who are familiar with those terms.

$$Kincaid = 11.8 * \frac{syllables}{words} + 0.39 * \frac{words}{sentences} - 15.59$$

Automated Readability Index

The Automated Readability Index is typically higher than Kincaid and Coleman-Liau, but lower than Flesch.

$$ARI = 4.71 * \frac{letters}{words} + 0.5 * \frac{words}{sentences} - 21.43$$

Coleman-Liau Formula

The Coleman-Liau Formula usually gives a lower grade than Kincaid, ARI and Flesch when applied to technical documents.

$$Coleman\text{-}Liau = 5.88 * \frac{letters}{words} - \frac{29.5 * sentences}{words} - 15.8$$

Flesh reading easy formula

The Flesh reading easy formula has been developed by Flesh in 1948 and it is based on school text covering grade 3 to 12. It is wide spread, especially in the USA, because of good results and simple computation. The index is usually between 0 (hard) and 100 (easy), standard English documents averages approximately 60 to 70. Applying it to German documents does not deliver good results because of the different language structure.

$$Flesch\ Index = 206.835 - 84.6 * \frac{syllables}{words} - 1.015 * \frac{words}{sentences}$$

Fog Index

The Fog index has been developed by Robert Gunning. Its value is a school grade. The "ideal" Fog Index level is 7 or 8. A level above 12 indicates the writing sample is too hard for most people to read. Only use it on texts of at least hundred words to get meaningful results. Note that a correct implementation would not count words of three or more syllables that are proper names, combinations of easy words, or made three syllables by suffixes such as −ed, −es, or −ing.

$$Fog\ Index = 0.4 * \left(\frac{words}{sentences} + 100 * \frac{words \geq 3\ syllables}{words} \right)$$

Lix formula

> The Lix formula developed by Björnsson from Sweden is very simple and employs a mapping table as well:

$$Lix = \frac{words}{sentences} + 100 * \frac{words > 6\ characters}{words}$$

Index	34	38	41	44	48	51	54	57
School year	5	6	7	8	9	10	11	

SMOG-Grading

> The SMOG-Grading for English texts has been developed by McLaughlin in 1969. Its result is a school grade.

$$SMOG\text{-}Grading = \sqrt{\frac{(words \geq 3\ syllables)}{sentences} * 30} + 3$$

It has been adapted to German by Bamberger & Vanecek in 1984, who changed the constant +3 to −2.

Word usage

The word usage counts are intended to help identify excessive use of particular parts of speech.

Verb Phrases

> The category of verbs labeled "to be" identifies phrases using the passive voice. Use the passive voice sparingly, in favor of more direct verb forms. The flag **-p** causes **style** to list all occurrences of the passive voice.

The verb category "aux" measures the use of modal auxiliary verbs, such as "can", "could", and "should". Modal auxiliary verbs modify the mood of a verb.

Conjunctions

> The conjunctions counted by style are coordinating and subordinating. Coordinating conjunctions join grammatically equal sentence fragments, such as a noun with a noun, a phrase with a phrase, or a clause to a clause. Coordinating conjunctions are "and," "but," "or," "yet," and "nor."

Subordinating conjunctions connect clauses of unequal status. A subordinating conjunction links a subordinate clause, which is unable to stand alone, to an independent clause. Examples of subordinating conjunctions are "because," "although," and "even if."

Pronouns

> Pronouns are contextual references to nouns and noun phrases. Documents with few pronouns generally lack cohesiveness and fluidity. Too many pronouns may indicate ambiguity.

Nominalizations

> Nominalizations are verbs that are changed to nouns. Style recognizes words that end in "ment," "ance," "ence," or "ion" as nominalizations. Examples are "endowment," "admittance," and "nominalization." Too much nominalization in a document can sound abstract and be difficult to understand. The flag **-N** causes **style** to list all nominalizations. The flag **-n** prints all sentences with either the passive voice or a nominalization.

OPTIONS

−L *language*, **−−language** *language*
> set the document language (**de**, **en**, **nl**).

−l *length*, **−−print-long** *length*
> print all sentences longer than *length* words.

−r *ari*, **−−print-ari** *ari*
> print all sentences whose readability index (ARI) is greater than *ari*.

−p *passive*, **−−print-passive**
    print all sentences phrased in the passive voice.

−N *nominalizations*, **−−print-nom**
    print all sentences containing nominalizations.

−n *nominalizations-passive*, **−−print-nom-passive**
    print all sentences phrased in the passive voice or containing nominalizations.

−h, **−−help**
    Print a short usage message.

−−version
    Print the version.

ERRORS
On usage errors, 1 is returned. Termination caused by lack of memory is signalled by exit code 2.

ENVIRONMENT
LC_MESSAGES=de|en|nl
    specifies the default document language. The default language is **en**.

LC_CTYPE=iso-8859-1
    specifies the document character set. The default character set is ASCII.

AUTHOR
This program is GNU software, copyright 1997–2007 Michael Haardt <michael@moria.de>.

It contains contributions by Jason Petrone <jpetrone@acm.org>, Uschi Stegemeier <uschi@morwain.de> and Hans Lodder.

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

HISTORY
There has been a style command on old UNIX systems, which is now part of the AT&T DWB package. The original version was bound to roff by enforcing a call to deroff.

SEE ALSO
deroff(1), diction(1)

Cherry, L.L.; Vesterman, W.: *Writing Tools—The STYLE and DICTION programs*, Computer Science Technical Report 91, Bell Laboratories, Murray Hill, N.J. (1981), republished as part of the 4.4BSD User's Supplementary Documents by O'Reilly.

Coleman, M. and Liau,T.L. (1975). 'A computer readability formula designed for machine scoring', Journal of Applied Psychology, 60(2), 283-284.