en\_US Hunspell Dictionary

Version 2020.12.07

Mon Dec 7 20:14:35 2020 -0500 [5ef55f9]

http://wordlist.sourceforge.net

README file for English Hunspell dictionaries derived from SCOWL.

These dictionaries are created using the speller/make-hunspell-dict

script in SCOWL.

The following dictionaries are available:

en\_US (American)

en\_CA (Canadian)

en\_GB-ise (British with "ise" spelling)

en\_GB-ize (British with "ize" spelling)

en\_AU (Australian)

en\_US-large

en\_CA-large

en\_GB-large (with both "ise" and "ize" spelling)

en\_AU-large

The normal (non-large) dictionaries correspond to SCOWL size 60 and,

to encourage consistent spelling, generally only include one spelling

variant for a word. The large dictionaries correspond to SCOWL size

70 and may include multiple spelling for a word when both variants are

considered almost equal. The larger dictionaries however (1) have not

been as carefully checked for errors as the normal dictionaries and

thus may contain misspelled or invalid words; and (2) contain

uncommon, yet valid, words that might cause problems as they are

likely to be misspellings of more common words (for example, "ort" and

"calender").

To get an idea of the difference in size, here are 25 random words

only found in the large dictionary for American English:

Bermejo Freyr's Guenevere Hatshepsut Nottinghamshire arrestment

crassitudes crural dogwatches errorless fetial flaxseeds godroon

incretion jalapeño's kelpie kishkes neuroglias pietisms pullulation

stemwinder stenoses syce thalassic zees

The en\_US, en\_CA and en\_AU are the official dictionaries for Hunspell.

The en\_GB and large dictionaries are made available on an experimental

basis. If you find them useful please send me a quick email at

kevina@gnu.org.

If none of these dictionaries suite you (for example, maybe you want

the normal dictionary that also includes common variants) additional

dictionaries can be generated at http://app.aspell.net/create or by

modifying speller/make-hunspell-dict in SCOWL. Please do let me know

if you end up publishing a customized dictionary.

If a word is not found in the dictionary or a word is there you think

shouldn't be, you can lookup the word up at http://app.aspell.net/lookup

to help determine why that is.

General comments on these list can be sent directly to me at

kevina@gnu.org or to the wordlist-devel mailing lists

(https://lists.sourceforge.net/lists/listinfo/wordlist-devel). If you

have specific issues with any of these dictionaries please file a bug

report at https://github.com/kevina/wordlist/issues.

IMPORTANT CHANGES INTRODUCED In 2016.11.20:

New Australian dictionaries thanks to the work of Benjamin Titze

(btitze@protonmail.ch).

IMPORTANT CHANGES INTRODUCED IN 2016.04.24:

The dictionaries are now in UTF-8 format instead of ISO-8859-1. This

was required to handle smart quotes correctly.

IMPORTANT CHANGES INTRODUCED IN 2016.01.19:

"SET UTF8" was changes to "SET UTF-8" in the affix file as some

versions of Hunspell do not recognize "UTF8".

ADDITIONAL NOTES:

The NOSUGGEST flag was added to certain taboo words. While I made an

honest attempt to flag the strongest taboo words with the NOSUGGEST

flag, I MAKE NO GUARANTEE THAT I FLAGGED EVERY POSSIBLE TABOO WORD.

The list was originally derived from Németh László, however I removed

some words which, while being considered taboo by some dictionaries,

are not really considered swear words in today's society.

COPYRIGHT, SOURCES, and CREDITS:

The English dictionaries come directly from SCOWL

and is thus under the same copyright of SCOWL. The affix file is

a heavily modified version of the original english.aff file which was

released as part of Geoff Kuenning's Ispell and as such is covered by

his BSD license. Part of SCOWL is also based on Ispell thus the

Ispell copyright is included with the SCOWL copyright.

The collective work is Copyright 2000-2018 by Kevin Atkinson as well

as any of the copyrights mentioned below:

Copyright 2000-2018 by Kevin Atkinson

Permission to use, copy, modify, distribute and sell these word

lists, the associated scripts, the output created from the scripts,

and its documentation for any purpose is hereby granted without fee,

provided that the above copyright notice appears in all copies and

that both that copyright notice and this permission notice appear in

supporting documentation. Kevin Atkinson makes no representations

about the suitability of this array for any purpose. It is provided

"as is" without express or implied warranty.

Alan Beale <biljir@pobox.com> also deserves special credit as he has,

in addition to providing the 12Dicts package and being a major

contributor to the ENABLE word list, given me an incredible amount of

feedback and created a number of special lists (those found in the

Supplement) in order to help improve the overall quality of SCOWL.

The 10 level includes the 1000 most common English words (according to

the Moby (TM) Words II [MWords] package), a subset of the 1000 most

common words on the Internet (again, according to Moby Words II), and

frequently class 16 from Brian Kelk's "UK English Wordlist

with Frequency Classification".

The MWords package was explicitly placed in the public domain:

The Moby lexicon project is complete and has

been place into the public domain. Use, sell,

rework, excerpt and use in any way on any platform.

Placing this material on internal or public servers is

also encouraged. The compiler is not aware of any

export restrictions so freely distribute world-wide.

You can verify the public domain status by contacting

Grady Ward

3449 Martha Ct.

Arcata, CA 95521-4884

grady@netcom.com

grady@northcoast.com

The "UK English Wordlist With Frequency Classification" is also in the

Public Domain:

Date: Sat, 08 Jul 2000 20:27:21 +0100

From: Brian Kelk <Brian.Kelk@cl.cam.ac.uk>

> I was wondering what the copyright status of your "UK English

> Wordlist With Frequency Classification" word list as it seems to

> be lacking any copyright notice.

There were many many sources in total, but any text marked

"copyright" was avoided. Locally-written documentation was one

source. An earlier version of the list resided in a filespace called

PUBLIC on the University mainframe, because it was considered public

domain.

Date: Tue, 11 Jul 2000 19:31:34 +0100

> So are you saying your word list is also in the public domain?

That is the intention.

The 20 level includes frequency classes 7-15 from Brian's word list.

The 35 level includes frequency classes 2-6 and words appearing in at

least 11 of 12 dictionaries as indicated in the 12Dicts package. All

words from the 12Dicts package have had likely inflections added via

my inflection database.

The 12Dicts package and Supplement is in the Public Domain.

The WordNet database, which was used in the creation of the

Inflections database, is under the following copyright:

This software and database is being provided to you, the LICENSEE,

by Princeton University under the following license. By obtaining,

using and/or copying this software and database, you agree that you

have read, understood, and will comply with these terms and

conditions.:

Permission to use, copy, modify and distribute this software and

database and its documentation for any purpose and without fee or

royalty is hereby granted, provided that you agree to comply with

the following copyright notice and statements, including the

disclaimer, and that the same appear on ALL copies of the software,

database and documentation, including modifications that you make

for internal use or for distribution.

WordNet 1.6 Copyright 1997 by Princeton University. All rights

reserved.

THIS SOFTWARE AND DATABASE IS PROVIDED "AS IS" AND PRINCETON

UNIVERSITY MAKES NO REPRESENTATIONS OR WARRANTIES, EXPRESS OR

IMPLIED. BY WAY OF EXAMPLE, BUT NOT LIMITATION, PRINCETON

UNIVERSITY MAKES NO REPRESENTATIONS OR WARRANTIES OF MERCHANT-

ABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OR THAT THE USE OF THE

LICENSED SOFTWARE, DATABASE OR DOCUMENTATION WILL NOT INFRINGE ANY

THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

The name of Princeton University or Princeton may not be used in

advertising or publicity pertaining to distribution of the software

and/or database. Title to copyright in this software, database and

any associated documentation shall at all times remain with

Princeton University and LICENSEE agrees to preserve same.

The 40 level includes words from Alan's 3esl list found in version 4.0

of his 12dicts package. Like his other stuff the 3esl list is also in the

public domain.

The 50 level includes Brian's frequency class 1, words appearing

in at least 5 of 12 of the dictionaries as indicated in the 12Dicts

package, and uppercase words in at least 4 of the previous 12

dictionaries. A decent number of proper names is also included: The

top 1000 male, female, and Last names from the 1990 Census report; a

list of names sent to me by Alan Beale; and a few names that I added

myself. Finally a small list of abbreviations not commonly found in

other word lists is included.

The name files form the Census report is a government document which I

don't think can be copyrighted.

The file special-jargon.50 uses common.lst and word.lst from the

"Unofficial Jargon File Word Lists" which is derived from "The Jargon

File". All of which is in the Public Domain. This file also contain

a few extra UNIX terms which are found in the file "unix-terms" in the

special/ directory.

The 55 level includes words from Alan's 2of4brif list found in version

4.0 of his 12dicts package. Like his other stuff the 2of4brif is also

in the public domain.

The 60 level includes all words appearing in at least 2 of the 12

dictionaries as indicated by the 12Dicts package.

The 70 level includes Brian's frequency class 0 and the 74,550 common

dictionary words from the MWords package. The common dictionary words,

like those from the 12Dicts package, have had all likely inflections

added. The 70 level also included the 5desk list from version 4.0 of

the 12Dics package which is in the public domain.

The 80 level includes the ENABLE word list, all the lists in the

ENABLE supplement package (except for ABLE), the "UK Advanced Cryptics

Dictionary" (UKACD), the list of signature words from the YAWL package,

and the 10,196 places list from the MWords package.

The ENABLE package, mainted by M\Cooper <thegrendel@theriver.com>,

is in the Public Domain:

The ENABLE master word list, WORD.LST, is herewith formally released

into the Public Domain. Anyone is free to use it or distribute it in

any manner they see fit. No fee or registration is required for its

use nor are "contributions" solicited (if you feel you absolutely

must contribute something for your own peace of mind, the authors of

the ENABLE list ask that you make a donation on their behalf to your

favorite charity). This word list is our gift to the Scrabble

community, as an alternate to "official" word lists. Game designers

may feel free to incorporate the WORD.LST into their games. Please

mention the source and credit us as originators of the list. Note

that if you, as a game designer, use the WORD.LST in your product,

you may still copyright and protect your product, but you may \*not\*

legally copyright or in any way restrict redistribution of the

WORD.LST portion of your product. This \*may\* under law restrict your

rights to restrict your users' rights, but that is only fair.

UKACD, by J Ross Beresford <ross@bryson.demon.co.uk>, is under the

following copyright:

Copyright (c) J Ross Beresford 1993-1999. All Rights Reserved.

The following restriction is placed on the use of this publication:

if The UK Advanced Cryptics Dictionary is used in a software package

or redistributed in any form, the copyright notice must be

prominently displayed and the text of this document must be included

verbatim.

There are no other restrictions: I would like to see the list

distributed as widely as possible.

The 95 level includes the 354,984 single words, 256,772 compound

words, 4,946 female names and the 3,897 male names, and 21,986 names

from the MWords package, ABLE.LST from the ENABLE Supplement, and some

additional words found in my part-of-speech database that were not

found anywhere else.

Accent information was taken from UKACD.

The VarCon package was used to create the American, British, Canadian,

and Australian word list. It is under the following copyright:

Copyright 2000-2016 by Kevin Atkinson

Permission to use, copy, modify, distribute and sell this array, the

associated software, and its documentation for any purpose is hereby

granted without fee, provided that the above copyright notice appears

in all copies and that both that copyright notice and this permission

notice appear in supporting documentation. Kevin Atkinson makes no

representations about the suitability of this array for any

purpose. It is provided "as is" without express or implied warranty.

Copyright 2016 by Benjamin Titze

Permission to use, copy, modify, distribute and sell this array, the

associated software, and its documentation for any purpose is hereby

granted without fee, provided that the above copyright notice appears

in all copies and that both that copyright notice and this permission

notice appear in supporting documentation. Benjamin Titze makes no

representations about the suitability of this array for any

purpose. It is provided "as is" without express or implied warranty.

Since the original words lists come from the Ispell distribution:

Copyright 1993, Geoff Kuenning, Granada Hills, CA

All rights reserved.

Redistribution and use in source and binary forms, with or without

modification, are permitted provided that the following conditions

are met:

1. Redistributions of source code must retain the above copyright

notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright

notice, this list of conditions and the following disclaimer in the

documentation and/or other materials provided with the distribution.

3. All modifications to the source code must be clearly marked as

such. Binary redistributions based on modified source code

must be clearly marked as modified versions in the documentation

and/or other materials provided with the distribution.

(clause 4 removed with permission from Geoff Kuenning)

5. The name of Geoff Kuenning may not be used to endorse or promote

products derived from this software without specific prior

written permission.

THIS SOFTWARE IS PROVIDED BY GEOFF KUENNING AND CONTRIBUTORS ``AS IS'' AND

ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE

IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE

ARE DISCLAIMED. IN NO EVENT SHALL GEOFF KUENNING OR CONTRIBUTORS BE LIABLE

FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL

DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS

OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION)

HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT

LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY

OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF

SUCH DAMAGE.

Build Date: Mon Dec 7 20:19:27 EST 2020

Wordlist Command: mk-list --accents=strip en\_US 60