

# Package ‘cld3’

January 29, 2023

**Type** Package

**Title** Google's Compact Language Detector 3

**Version** 1.5.0

**Description** Google's Compact Language Detector 3 is a neural network model for language identification and the successor of 'cld2' (available from CRAN). The algorithm is still experimental and takes a novel approach to language detection with different properties and outcomes. It can be useful to combine this with the Bayesian classifier results from 'cld2'. See <<https://github.com/google/cld3#readme>> for more information.

**License** Apache License 2.0

**Encoding** UTF-8

**URL** <https://docs.ropensci.org/cld3/> <https://github.com/ropensci/cld3>  
<https://ropensci.r-universe.dev/cld3>

**Imports** Rcpp

**LinkingTo** Rcpp

**RoxygenNote** 6.0.1.9000

**SystemRequirements** libprotobuf and protobuf-compiler

**Suggests** testthat, cld2

**NeedsCompilation** yes

**Author** Jeroen Ooms [aut, cre] (<<https://orcid.org/0000-0002-4035-0289>>),  
Google Inc [cph] (CLD3 C++ library)

**Maintainer** Jeroen Ooms <[jeroen@berkeley.edu](mailto:jeroen@berkeley.edu)>

**Repository** CRAN

**Date/Publication** 2023-01-29 22:20:06 UTC

## R topics documented:

cld3 . . . . .	2
<b>Index</b>	<b>3</b>

**Description**

The function `detect_language()` is vectorised and guesses the the language of each string in text or returns NA if the language could not reliably be determined. The function `detect_language_multi()` is not vectorised and detects all languages inside the entire character vector as a whole.

**Usage**

```
detect_language(text)
```

```
detect_language_mixed(text, size = 3)
```

**Arguments**

<code>text</code>	a string with text to classify or a connection to read from
<code>size</code>	number of languages to detect

**Examples**

```
# Vectorized best guess
text <- c("To be or not to be?", "Ce n'est pas grave.",
         "Hij heeft de klok horen luiden maar weet niet waar de klepel hangt.")
detect_language(text)

# Multiple languages in one text (doesn't seem to work well)
detect_language_mixed(text)
```

# Index

`cld3`, [2](#)

`detect_language(cld3)`, [2](#)

`detect_language()`, [2](#)

`detect_language_mixed(cld3)`, [2](#)

`detect_language_multi(cld3)`, [2](#)

`detect_language_multi()`, [2](#)