Internationalisation of R help pages

Eio Campitelli and Renata Hirota

2023-09-26

Signatories

Project team

- Elio Campitelli: Maintainer of several packages published on CRAN. Spanish translator of rOpenSci Packages: Development, Maintenance, and Peer Review.
- Renata Hirota

Contributors

- Maëlle Salmon
- Yanina Bellini Saibene
- Michael Chirico?
- Gergely Daroczi?
- Hadley?

Consulted

-
- Yanina Bellini Saibene

The Problem

R's help system is one of the great features of the language. Usage, description and examples are available right there in the console. It is important that this documentation is correct and useful both in the core packages and in the general contributed packages ecosystem. To that aim, CRAN enforces a suite of policies on documentation, such as all exported object having documentation, all arguments being documented, etc.

Besides meeting those technical requirements, documentation needs to also be accessible to end users to be useful. This involves clear and consise language, consistent style, useful examples and many other characteristics not tested by CRAN. Not trivially, it also requires being written in a language the user actually understands.

English is currently the de-facto international language and this is reflected in R function and variable names, like how the `mean()` function is not called `promedio()` or `Mittelwert()` and documentation language. And while contributed packages have a "Language" field and can be documented in other languages, the vast majority of contributed packages are documented in English.

There is a small number of packages documented in other languages, presumably in accordance to their target audience. For example, labstatR is a companion package for the italian book Laboratorio Di Statistica Con

R and is documented in Italian. The chilemapas package provides various simplified maps for Chile and it's documentation and function names are in Spanish.

Packages documented in non-English languages can be more accessible for their intended populations, but they are much less accessible to the wider community. Useful functions/algorithms implemented in those packages would be hard to use for the rest of the community. So package authors are faced with the decision of making their package inaccessible to their target demographic or isolated from the wider ecosystem.

Real cases of this issue exists. For example, the utilsIPEA package is a package for the Brazilian Instituto de Pesquisa Economica Aplicada is documented in English and it's functions are a mix of English and Portuguese. The author publicly expressed his need for bilingual documentation:

> I am writing a package to facilitate importing Brazilian socio-economic microdata sets (Census, PNAD, etc). I foresee two distinct groups of users of the package:
>
> - Users in Brazil, who may feel more at ease with the documentation in Portuguese. The probably can understand English to some extent, but a foreign language would probably make the package feel less "ergonomic".
> - The broader international users community, from whom English documentation may be a necessary condition.
>
> Is it possible to write a package in a way that the documentation is "bilingual" (English and Portuguese), and that the language shown to the user will depend on their country/language settings?

Moreover, CRAN hosts at least two packages that have a secondary package version with documentation in another language. The ExpDes package has the companion package ExpDes.pt with documentation in Portuguese. Similarly, the orloca package is documented in Spanish in the orloca.es package.

Needless to say, this method of bilingual documentation is not recommended, as it's very hard to maintain and doesn't scale to other languages. A better alternative would be for R to allow packages to have multilingual documentation.

The proposal

Overview

SDe habló con R core en

We propose to extend the R help system to allow for multiple help pages for the same function in different languages. By default, `help(function)` would show the documentation in the preferred language of the user or fall-back to the canonical documentation otherwise (most likely, in English). It would also include a link to the canonical documentation and warnings if translations are out of date.

Detail

A possible implementation would be as follows:

- Original packages have their "canonical" help pages written in the original languages. (e.g. `mein_paket` is documented in German).
- Translations are hosted in translation modules that the user can install to get the documentation in that language. (e.g. `mein_paket.en` would provide English documentation for `mein_paket`). These modules are R packages that use the PackageType field to indicate that they are a translation module (e.g. `PackageType: translation`). The Depends field is used to indicate the package being translated and the minimum version supported.
- Translation modules would store .po files with the translated strings.

- When a translation module is installed, the .Rd files of the original package are parsed and translated using `gettext()` and the .po files in the translation module and serialised into binary help pages (like regular packages have).
- When loading a package, R will also search for installed translations and load them too.
- `help()` gains an new "language" argument which defaults to `Sys.getenv("LANGUAGE")`.
- `help()` searches for the loaded topics. If any translation is available, then it would use the `language` argument to disambiguate.
- Help pages should include a link to the original (canonical) documentation.

More details can be found in the rhelpi18n repository.

Mapa conceptual

Project plan

Start-up phase

Technical delivery

Other aspects

Requirements

People

- Elio Campitelli: Maintainer of several packages published on CRAN. Spanish translator of rOpenSci Packages: Development, Maintenance, and Peer Review.
- Maëlle Salmon:
- Michael Lawrence?
- Yanina
- Renata
- Michael Chirico?
- Gergely Daroczi?
  - https://daroczig.github.io/slides/2022-06-22-useR_2022_poster-Gergely_Daroczi-Internationalization_of_R_packages_with_r18r.pdf

Processes

Tools & Tech

Funding

Summary

Success

Definition of done

Measuring success

Future work

Key risks