

La importancia del código fuente y el paquete *annotater* para añadir contexto sobre librerías adicionales

Luis D. Verde Arregoitia

Palabras clave: reproducibilidad, addin, script

Abstract

Motivación

En programación, el archivo de texto que contiene las instrucciones a ejecutar es una de las unidades fundamentales que se necesitan para interactuar con un programa o lenguaje de manera interactiva o automatizada¹. Así como en el cine y teatro, el script (por su nombre en inglés) es el libreto o guion que da contexto y relata lo que va sucediendo de manera secuencial. Al tener un registro de lo que se está haciendo, es posible reproducir una misma secuencia de instrucciones, modificarla, y compartirla². En R, uno de los componentes más importantes que deben aparecer en los scripts son los llamados a librerías adicionales que no se instalan junto con R base y que aportan una infinidad de funciones nuevas para diferentes fines³. Actualmente existen más de 19,000 paquetes disponibles para instalar en CRAN, el repositorio más grande de librerías para R. Ante la diversidad de paquetes existentes, puede ser de utilidad hacer anotaciones en un script con las características principales de estas librerías. El paquete *annotater* surge al reconocer que en ocasiones es posible encontrar scripts que llaman paquetes sin usar sus funciones (añadiendo dependencias innecesarias), o sin conocer el objetivo principal de cada uno.

Funcionamiento

El funcionamiento general se basa en una serie de funciones que identifican los llamados para cargar paquetes (las funciones *library()*, *require()*, o *load()*) y extraen los nombres de cada uno. Para cada nombre, se obtiene el resultado de *utils::packageDescription* que en sí procesa el archivo ‘DESCRIPTION’ de un paquete instalado y genera un objeto tabular. Los campos de este objeto se utilizan para construir anotaciones que serán agregadas como comentarios en el script. El campo *Title* generalmente describe de manera concisa el objetivo general de un paquete, por lo cual este campo está disponible como anotación. También es posible obtener el repositorio fuente del cual se instaló un paquete (ej: CRAN, GitHub, bioconductor) y la versión instalada. Este tipo de anotaciones pueden ser útiles al momento de compartir archivos con código e incluso contribuyen con la reproducibilidad de un análisis o proceso al tener un registro de las versiones utilizadas. Las funciones de *annotater* se pueden utilizar a través de *addins*

LA IMPORTANCIA DEL CODIGO FUENTE Y EL PAQUETE ANNOTATER PARA AÑADIR CONTEXO SOBRE LIBRERIAS ADICIONALES

interactivos en el entorno Rstudio y funcionan para archivos .R y .Rmd.

Futuro

Se pretende probar e incorporar la funcionalidad de *annotater* para archivos de Quarto, y potencialmente diseñar nuevas funciones para trabajar con metapaquetes que al cargarse a su vez llaman a otros paquetes (ej: *tidyverse*, *tidymodels*).

Referencias

1. Flujo de trabajo: *Scripts* en ‘R para Ciencia de Datos’ <https://es.r4ds.hadley.nz/flujo-de-trabajo-scripts.html>
2. Mendoza Vega, J.B. (2018) R para principiantes. <https://bookdown.org/jboscomendoza/r-principiantes4/scripts.html>
3. Wickham, H y J. Bryan. R Packages (2a edición) <https://r-pkgs.org/>

* *annotater* se encuentra disponible en CRAN y su sitio de referencia es <https://annotater.liomys.mx>

Nombre autor de contacto: Luis D. Verde Arregoitia

Instituciones del autor de contacto: Instituto de Ecología AC, INECOL, México

luis@liomys.mx