

СТРУКТУРА ДОПОЛНИТЕЛЬНЫХ ПАКЕТОВ В R¹

Кузнецова Л.В., Зарядов И.С.

Российский университет дружбы народов, izaryadov@sci.pfu.edu.ru

Рассматривается структура дополнительных пакетов, разрабатываемых для программного комплекса R.

Ключевые слова: статистический комплекс R, структура, пакет.

Введение

Статистический программный комплекс R в последнее время становится всё более и более популярным не только на Западе, но и России. Это связано с тем, что R можно использовать при решении различных задач, связанных с обработкой и анализом данных, в следующих областях: медицина, экология, финансы, страхование, социальные науки, статистическая генетика, решение дифференциальных уравнений, кластерный анализ, высокопроизводительные технологии и параллельные вычисления, математическое моделирование, решение оптимизационных задач.

В базовую комплектацию R входит около 30 пакетов, отвечающие за графику и основные математико-статистические функции. Помимо базовых пакетов на текущий момент разработано еще 5321 дополнительный пакет, находящийся в свободном доступе.

Основные принципы создания пакетов в R

Основная идея дополнительного (вспомогательного) пакета в R – это механизм, отвечающий за дополнительные коды, данные и документацию.

Коды могут быть написаны на языке программирования R, C, C++ или Fortran, могут быть предназначены для различных операционных систем (Unix, Windows, MacOS), но при этом структура создаваемого пакета остается неизменной [1]. Создаваемый для решения конкретной задачи вспомогательный пакет обязательно содержит в себе:

- файл **DESCRIPTION** (полное описание пакета – название, версия, дата создания, предназначение, лицензия и др.);
- файл **NAMESPACE** (файл, содержащий названия импортируемых и экспортируемых переменных, классов создаваемых и/или используемых объектов)

и следующие поддиректории (некоторые из них могут отсутствовать, но не могут быть пустыми):

- **R** (скрипты разработанных функций на языке R или S, с расширениями .R, .r, .S или .s);
- **data** (файлы с данными (к примеру, статистические данные, выборки) с расширениями .csv, .txt, .R);
- **demo** (R скрипты, используемые для демонстрации работы некоторых функций пакета, с помощью функции demo());
- **man** (документация по пакету);
- **src** (содержит файлы с кодами, которые должны быть откомпилированы);
- **inst** (содержимое этой поддиректории при загрузке и запуске вспомогательного пакета устанавливается в рабочую директорию R);
- **tests** (тесты для проверки правильности созданных кодов, правильности установки пакета);
- **exec** (дополнительные исполняемые скрипты для интерпретаторов shell, Perl или Tcl);

¹ Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научного проекта № 14-07-00090-а.

– **po** (отвечает за языковую локализацию пакета – перевод предупреждений и сообщений об ошибках), а также дополнительные файлы: **INDEX** (содержит описание наиболее важных с точки зрения разработчика объектов в пакете), **configure** (используется в UNIX-подобных системах при установке пакета), **cleanup** используется в UNIX-подобных системах при установке пакета), **LICENSE** и **NEWS**.

Простейшим дополнительным пакетом в R является пакет, состоящий только из двух файлов (**DESCRIPTION** и **NAMESPACE**) и поддиректории **R** со скриптами разработанных функций.

Выводы

Разработка дополнительных пакетов для пакета R, как показано, не является особо сложной процедурой.

Литература

1. Writing R Extensions. Version 3.0.3 (2014-03-06) – <http://cran.r-project.org/>
2. R Installation and Administration. Version 3.0.3 (2014-03-06) – <http://cran.r-project.org/>

R PACKAGE STRUCTURE

Kuznetcova L.V., Zaryadov I.S

Peoples' Friendship University of Russia, izaryadov@sci.pfu.edu.ru

The structure of packages in statistical complex R is presented.

Key words: R statistic complex, package development.