УДК 004.738.5:004.438R
ББК 32.971.353

Х89

Храмов Д. А.
Х89 Сбор данных в Интернете на языке R. – М.: ДМК Пресс, 2017. – 280 с.: ил.

ISBN 978-5-97060-459-5

Всё, что регистрирует человек и созданные им машины, может считаться данными. Фиксируя новое и переводя архивы в цифровую форму, мы с каждым днём производим всё больше данных. Но гораздо чаще случается так, что данные разбросаны по всемирной сети на многочисленных страницах онлайновых магазинов, заметках в социальных сетях, логах серверов и т. п. Прежде чем начать работать с такими данными, их необходимо собрать и сохранить в пригодном для анализа виде. Решению этих вопросов и посвящена данная книга.

Основной материал книги разделён на две части. В первой части дано краткое введение в R – описание среды разработки, языка и основных пакетов-расширений. Вторая часть посвящена непосредственно сбору данных: работе с открытыми данными, извлечению данных из веб-страниц и из социальных сетей. Также рассмотрены необходимые технические вопросы: протокол HTTP, функции импорта данных различных форматов и регулярные выражения. Завершается рассказ созданием карт на основе собранных данных.

Издание предназначено специалистам по анализу данных, а также программистам, интересующихся сбором данных в Интернете.

УДК 004.738.5:004.438R
ББК 32.971.353

ISBN 978-5-97060-459-5

ŋqdqуеɏ рɆ dеdɆрɏзə р ɐɔɑрɔɏ Ɉqdqɏqуè

Содержание

ŋqdqɏeɎ bɆ dedɆbɎɜe d ɛɑɑbɔɏ ɟqddɎqɏè

ŋqdqɐɘ⅄ bɆ dɘdɆb⅄ɜɘ d ɕɑɑbɔɐ Ɉqdd⅄qɐè