O'REILLY®

Прикладной анализ текстовых данных на Python

МАШИННОЕ ОБУЧЕНИЕ И СОЗДАНИЕ ПРИЛОЖЕНИЙ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Бенджамин Бенгфорт
Ребекка Билбро и Тони Охеда

ПИТЕР®

Applied Text Analysis with Python

Enabling Language-Aware Data Products with Machine Learning

Benjamin Bengfort, Rebecca Bilbro, and Tony Ojeda

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Бенджамин Бенгфорт
Ребекка Билбро и Тони Охеда

Прикладной анализ текстовых данных на Python

МАШИННОЕ ОБУЧЕНИЕ И СОЗДАНИЕ ПРИЛОЖЕНИЙ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

ПИТЕР®

Санкт-Петербург • Москва • Екатеринбург • Воронеж
Нижний Новгород • Ростов-на-Дону
Самара • Минск

2019

ББК 32.973.233-018
УДК 004.62
Б46

Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони

Б46 Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. — СПб.: Питер, 2019. — 368 с.: ил. — (Серия «Бестселлеры O'Reilly»).

ISBN 978-5-4461-1153-4

Технологии анализа текстовой информации стремительно меняются под влиянием машинного обучения. Нейронные сети из теоретических научных исследований перешли в реальную жизнь, и анализ текста активно интегрируется в программные решения. Нейронные сети способны решать самые сложные задачи обработки естественного языка, никого не удивляет машинный перевод, «беседа» с роботом в интернет-магазине, перефразирование, ответы на вопросы и поддержание диалога. Почему же Сири, Алекса и Алиса не хотят нас понимать, Google находит не то, что мы ищем, а машинные переводчики веселят нас примерами «трудностей перевода» с китайского на албанский? Ответ кроется в мелочах — в алгоритмах, которые правильно работают в теории, но сложно реализуются на практике. Научитесь применять методы машинного обучения для анализа текста в реальных задачах, используя возможности и библиотеки Python. От поиска модели и предварительной обработки данных вы перейдете к приемам классификации и кластеризации текстов, затем приступите к визуальной интерпретации, анализу графов, а после знакомства с приемами масштабирования научитесь использовать глубокое обучение для анализа текста.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.973.233-018
УДК 004.62

ISBN 978-1491963043 англ.

ISBN 978-5-4461-1153-4

Краткое содержание