

Немного книг и ссылок,  
которые помогут  
подготовиться к  
хакатону. Или... решить,  
что это точно не для вас



...матром сверх фуражки, бма одет в женскую шубенку  
видано, захмелев, обнявши рукой солдата, сидевшего  
често, пел хрипавым, перерывающимся голосом  
несно. Солдаты держались за бока, глядя

Л.Н. Толстой  
**ВОЙНА  
И МИР**



Дарелл Хафф (Huff Darrell). «Как лгать при помощи статистики».

Научно-популярная книга, которой несколько десятилетий и она только сейчас вышла на русском языке.

Так как одна из основ анализа данных – математическая статистика, неплохо будет к ней немного прикоснуться. Пусть в книге не очень много научной глубины, зато интересные и остроумные примеры.

Книга показывает, что надо осторожно делать выводы из своих исследований. Чтобы не получилось так: «Онлайн-опрос показал, что 100% людей пользуются интернетом».



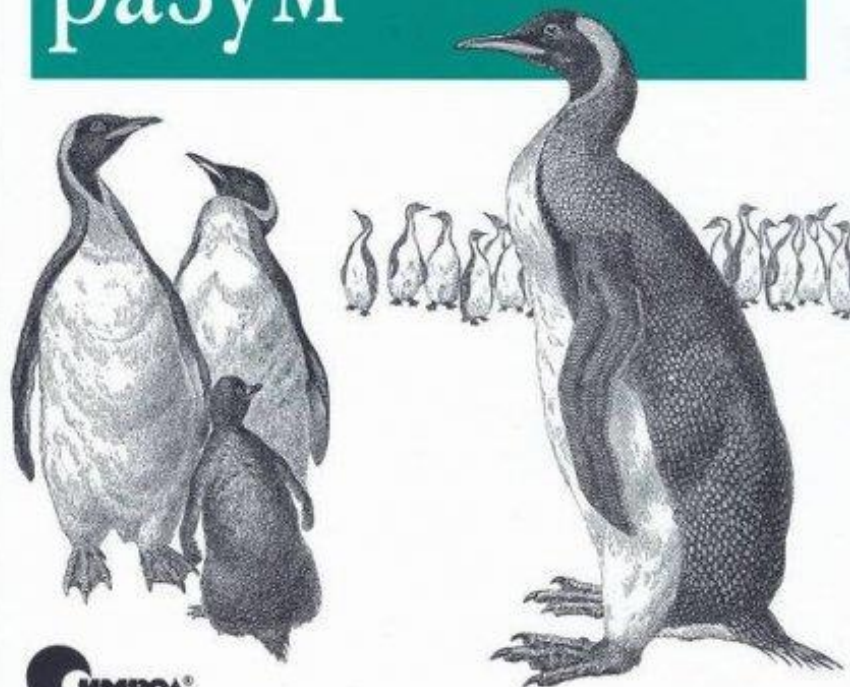
«Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим».

Немного пафоса, но вы точно поймете, почему про большие данные так много говорят и все пытаются их анализировать.

Создание интеллектуальных приложений для Web 2.0

*Программируем*

# КОЛЛЕКТИВНЫЙ РАЗУМ



**С**ИМВОЛ®  
O'REILLY®

*Тоби Сегаран*

«Программируем коллективный разум».

Практические примеры, минимум воды. Питон и программирование. Книга 2008 года. Не потеряла своей актуальности, может служить для сравнения того, что думали 8 лет назад и что думают об этой теме теперь.

PDF очень легко находится в интернетах (чуть ли не первая ссылка в поиске).





Национальный  
исследовательский университет  
«Высшая школа экономики»

АВТОРСКИЙ  
УЧЕБНИК

Б.Г. Миркин

# Введение в анализ данных

Учебник и практикум



УМО ВО рекомендует

ЮРАЙТ  
biblio-online.ru

Б.Г. Миркин. «Введение в анализ данных. Учебник и практикум».

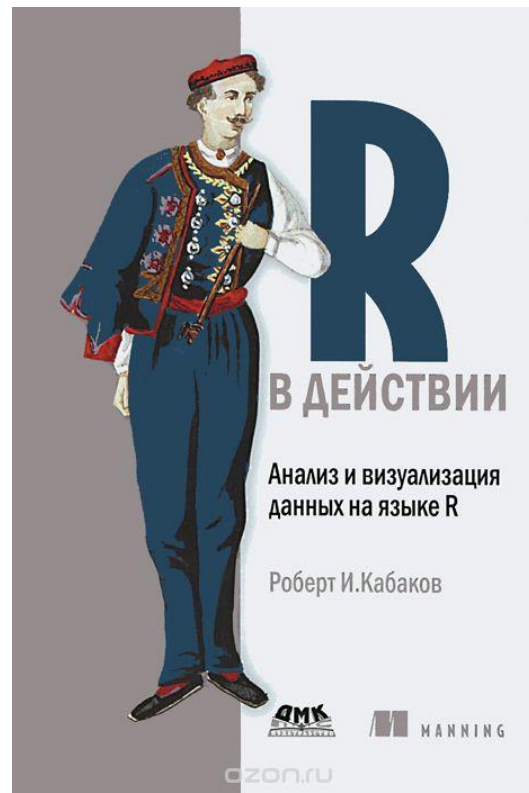
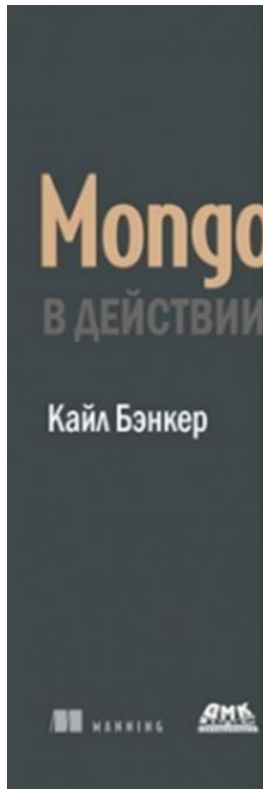
Возможно, стоит быстро просмотреть (это академический учебник).

Автор интересен тем, что занимался анализом данных еще в те времена, когда это не было модно. Поэтому присутствует систематичность изложение и немного кажущейся скукоты.



Для тех, кто дошел до этого слайда и не умер от скуки. Да, да, есть целая серия «Занимательная манга», часть книг из которых посвящена анализу данных.

Там есть сюжет, юмор, учеба. Надеемся, что вы получите несколько часов удовольствия. Скажем же огромное спасибо Сину Такахаси!



От теории к практике. Пара книг из серии «в действии». Данные для нашего Хакатона будут лежать в базе MongoDB, а обрабатывать эти данные можно, например, Питоном или специальным языком программирования R. Про Питон многие догадываются, а вот R известен меньше, но заточен именно на анализ данных.



Немного видео...



## Ссылки и видео:

- Классический курс по машинному обучению от Константина Воронцова. Он тяжелый, но попробуйте все-таки посмотреть [хотя бы первую лекцию](#).
- Андрей Себрант [про то, что такое big data и почему это интересно](#).
- Андрей Райгородский. [Лекция](#) о теории графов и интернете. Очень советуем в факультативном режиме посмотреть вообще лекции Райгородского, даже не относящиеся к нашей теме.
- Курс «[Анализ данных в R](#)».
- [Специализация BigData](#) на Coursera.

**Ссылки...**

- На Хабре есть [целый хаб](#), который посвящен анализу данных.
- [Kaggle](#). Сайт с задачами по машинному обучению.
- [Zooniverse](#). Проект по исследованию разнообразных космических данных.
- [Mlclass](#) – ресурс про Data Science.
  
- А еще...
- Воспользуйтесь поиском. Читайте, смотрите, интересуйтесь. ;)