

RECOMMENDER SYSTEMS



АЛЕКСАНДР КОНДРАШКИН

CTO & Co-founder

@LoyaltyLab



alkondr@loyaltylab.co



@alexkondr

ЦЕЛИ БЛОКА

ЦЕЛИ БЛОКА

- Научиться создавать рекомендательные системы
- Разобраться в математике рекомендаций
- Получить представления о том, где применять рекомендации

ОПИСАНИЕ БЛОКА

ЧТО БУДЕМ ИЗУЧАТЬ

- Введение и классификация рекомендательных систем
- Рекомендации на основе содержания
- Коллаборативная фильтрация
- Рекомендации на основе скрытых факторов
- Гибридные рекомендательные системы

МАТЕРИАЛЫ ПО БЛОКУ

ЛИТЕРАТУРА

- “Recommender Systems: The Textbook”, Charu C. Aggarwal, 2016, Springer Publishing Company, Incorporated
- “Recommender Systems Handbook”, Francesco Ricci, Lior Rokach, Bracha Shapira (editors), 2015, Springer US
- Google Scholar, Semantic Scholar, ResearchGate, Mendeley, ...

ПЛАН ЗАНЯТИЯ

ROADMAP НА СЕГОДНЯ

1

Что такое RS?

2

Типы RS

3

Типы данных
для RS

4

Предсказания и
рекомендации

5

Примеры RS

ВВЕДЕНИЕ

ТЕРМИНОЛОГИЯ

- **Пользователь (user)** — сущность, для которой строятся рекомендации (пользователь сервиса, PR-агентство)
- **Объект (item)** — сущность, которую рекомендуют (фильмы, книги, туристический маршрут)
- **Транзакция** — запись о взаимодействии пользователя с сервисом (лог-файл)
- **Сервис** — приложение, веб-сайт, и т.д.

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА

- Набор инструментов и технологий
- Рекомендует (советует) объекты, наиболее интересные для пользователя
- Помогает пользователю сузить объем информации для ознакомления
(мы никогда не изучаем все доступные YouTube ролики)
- Помогает пользователю принять решение

RS ВЕЗДЕ

Recommended



British Airways safety video - director's cut
British Airways ✓
4,923,360 views • 1 week ago



Trevor Noah Stand Up Comedy Special: Trevor Noah...
Kin Jin
169,152 views • 2 months ago



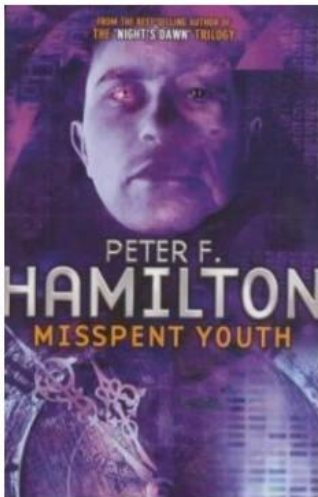
Man City vs Tottenham Extended Highlights
TheFinalWhistle HD
894,623 views • 1 day ago



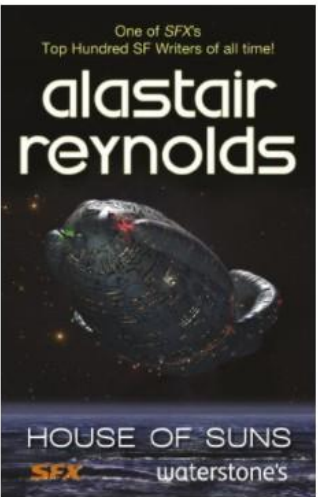
Гарик Харбатрутдин
OldSpiceRus
23,393,866 views



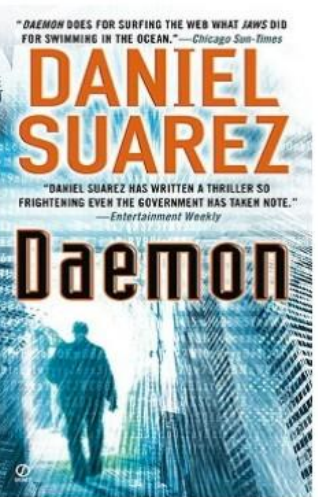
YOUR INTEREST IN SCIENCE FICTION



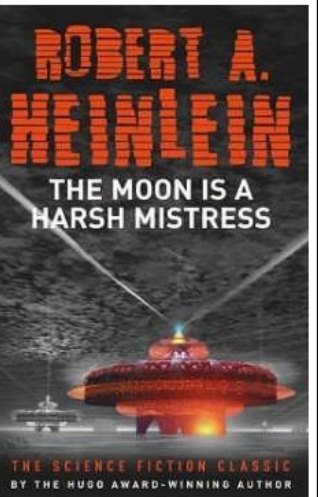
Want to Read
★★★★★
Not interested



Want to Read
★★★★★
Not interested



Want to Read
★★★★★
Not interested



Want to Read
★★★★★
Not interested

Recommended new releases



New Release Radio



Stony Hill



Add Violence



Ritual

Changes saved in Drive

Background... Layout Theme... Transition...

Present Comments Share

Explore

Search your docs and the web

Click to add title

Click to add text

Click to add title

Click to add text

ФУНКЦИИ RS

Рекомендации

объекты,
пакеты объектов,
последовательности
объектов

Аннотация

указать
пользователю
передачи в ТВ-
программе,
которые ему могут
быть интересны

Браузинг

StumbleUpon,
Surfingbird



—

RS ИСПОЛЬЗУЮТ, ЧТОБЫ...

RS ИСПОЛЬЗУЮТ, ЧТОБЫ...

- Увеличить количество продаж (*просмотров/кликов, CTR, ...*)
- Продавать различные объекты (*не только самые популярные*)
- Увеличить удовлетворение пользователя от использования сервиса
- Увеличить лояльность пользователя к сервису
- Лучше понимать пользователя (*лучше понимать, что нужно пользователю*)

СОСТАВЛЯЮЩИЕ ОПРЕДЕЛЕНИЯ

Академическая

1. Рекомендовать то, что пользователь сам не нашел бы
2. Рекомендовать то, что будет нравиться пользователю
3. Сделать пользователя счастливым (очень сложно измерить)

Коммерческая

1. Оптимизировать какие-нибудь показатели (легко измерить)
2. Минимизировать какие-нибудь затраты

Решение: баланс между
академией и промышленностью

ДАННЫЕ И ИСТОЧНИКИ ЗНАНИЙ



НЕТОЛОГИЯ
групп

ДАННЫЕ

RS — система обработки информации

Для построения рекомендаций собираются и анализируются
данные о пользователях/объектах/транзакциях

Разные модели — разные данные

ОБЪЕКТ

- Структурная информация
- Свойства
- Информация о взаимодействиях с объектом
- Стоимость (в том числе когнитивная) использования/приобретения
- Описания, оставленные пользователями

СЛОЖНОСТЬ ОПИСАНИЯ

1. Простые объекты:

Новости, книги, фильмы

2. Посложнее:

Автомобили, компьютеры

3. Сложные:

Финансовые вложения, рабочие вакансии, путешествия

Product details

Hardcover: 504 pages

Publisher: Knopf; First Edition edition (October 8, 2013)

Language: English

ISBN-10: 0385351399

ISBN-13: 978-0385351393

Product Dimensions: 6.4 x 1.5 x 8.8 inches

Shipping Weight: 1.7 pounds ([View shipping rates and policies](#))

Average Customer Review: ★★★★★ 3,397 customer reviews

Amazon Best Sellers Rank: #231,934 in Books ([See Top 100 in Books](#))

#1044 in [Books](#) > [Mystery, Thriller & Suspense](#) > [Thrillers & Suspense](#) > **Technothrillers**

#1567 in [Books](#) > [Science Fiction & Fantasy](#) > [Science Fiction](#) > **Dystopian**

#17644 in [Books](#) > [Literature & Fiction](#) > **Literary**

ПОЛЬЗОВАТЕЛЬ

1. Характеристики пользователя
2. Социо-демографические данные
3. Личностные (personality)
4. Информация о взаимодействиях с объектами
5. Поведенческие паттерны

The screenshot displays a user profile interface. At the top, there's a header 'Featured Skills & Endorsements' with a link 'Add a new skill'. Below this, three featured skills are listed: 'Python' (20 endorsements), 'Software Develop...' (17 endorsements), and 'Git' (12 endorsements). Each skill entry includes a user profile picture and text indicating endorsements from specific individuals or groups. Below the featured skills, a section titled 'Tural is also good at...' displays a grid of other skills: 'SQL' (13), 'Web Development' (10), 'Recommender Sy...' (10), 'Parsing' (8), 'Data Mining' (6), 'Regular Expressions' (7), 'Sphinx' (5), 'Django' (5), 'C#' (4), 'Scrum' (4), 'IR' (2), 'English', 'Computer Science', 'Machine Learning', 'Data Analysis', and 'Feature Extraction'.

Featured Skills & Endorsements [Add a new skill](#)

Python · 20
Endorsed by Petr Ermakov and 3 others who are highly skilled at this
Endorsed by 5 of Tural's colleagues at ivi.ru

Software Develop... · 17
Endorsed by 4 of Tural's colleagues at ivi.ru

Git · 12
Endorsed by Sergey Orlov and 1 other who is highly skilled at this
Endorsed by 4 of Tural's colleagues at ivi.ru

Tural is also good at...

SQL · 13
Web Development · 10
Recommender Sy... · 10
Parsing · 8
Data Mining · 6
Regular Expressions · 7
Sphinx · 5
Django · 5
C# · 4
Scrum · 4
IR · 2
English
Computer Science
Machine Learning
Data Analysis
Feature Extraction

ТРАНЗАКЦИЯ

1. Рейтинги (бинарные, численные)
2. Теги (свойства, которые пользователи присваивают объектам)
3. Действия/события
4. Время
5. Контекст



Like



Love



Haha



Yay



Wow



Sad

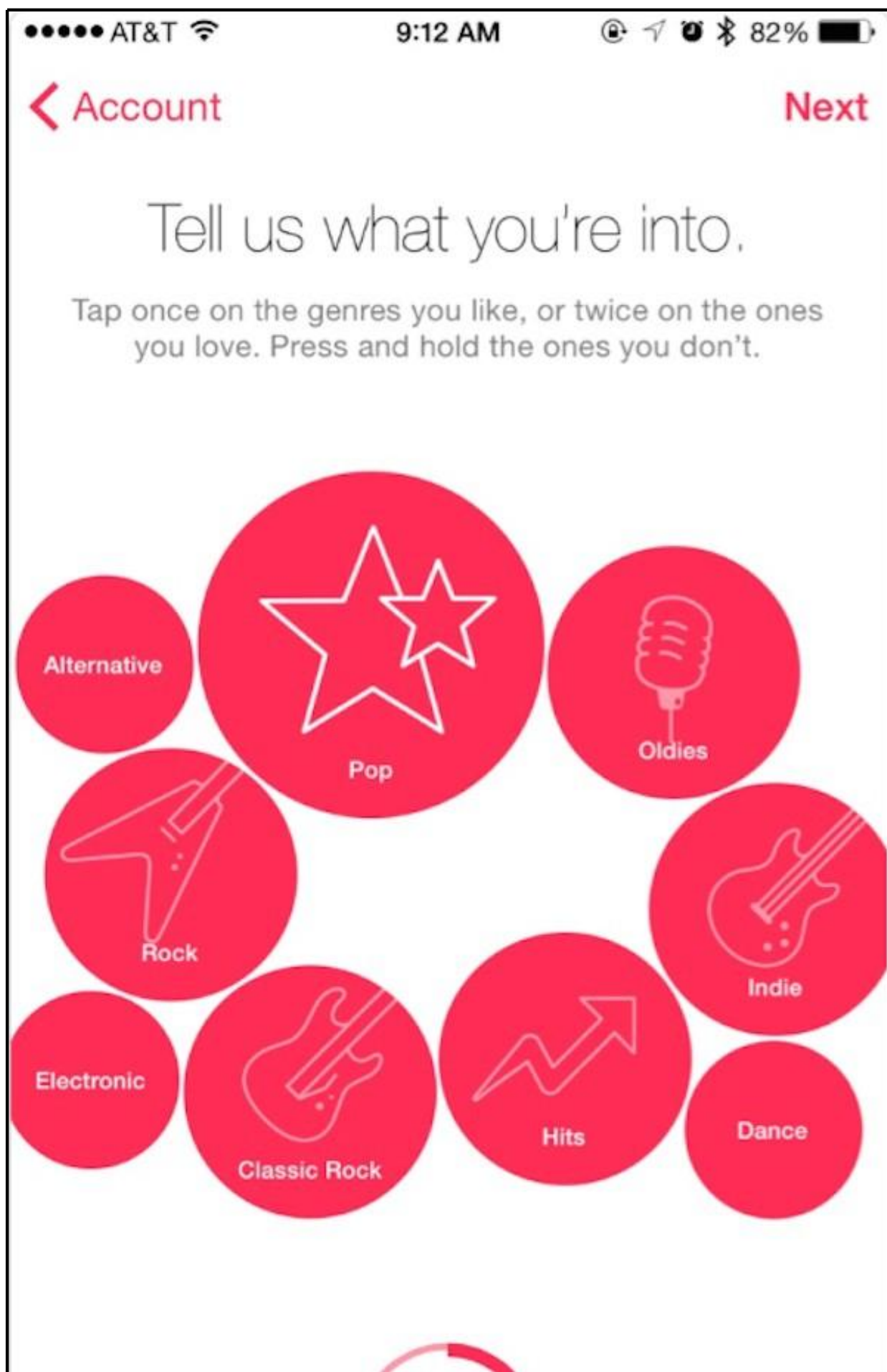


Angry

КОГДА СОБИРАТЬ?

1. До взаимодействия пользователя с объектом (*ожидаемое мнение*)
2. В момент знакомства пользователя с объектом
3. В момент взаимодействия пользователя с объектом
4. Сразу после взаимодействия пользователя с объектом
5. Через некоторое время после взаимодействия пользователя с объектом

КОГДА СОБИРАТЬ? (2)



< Account

Ne

Tell us what you're into.

Tap once on the genres you like, or twice on the ones you love. Press and hold the ones you don't.



Settings

Edit Profile

Your Status

Push

Do N

Adv

Sub

Rate

Help

About



How was your experience with us?



Never

Maybe Later

Settings

Edit Profile

Your Status

Push Notifications

Feedback

Suggest us what went wrong and we'll work on it.

Cancel

Submit

Rate Us

Help

About

< Account

Ne

Притяжение

2017, Россия, 12+

Tap once
you

Alternative

Roc

Electronic



Оцените этот фильм

Оценки улучшают персональные рекомендации



< Account

Притяжение

2017, Россия, 12+

Tap once
you

Alternative

Roc

Electronic

Оцените

Оценки улучшают пер

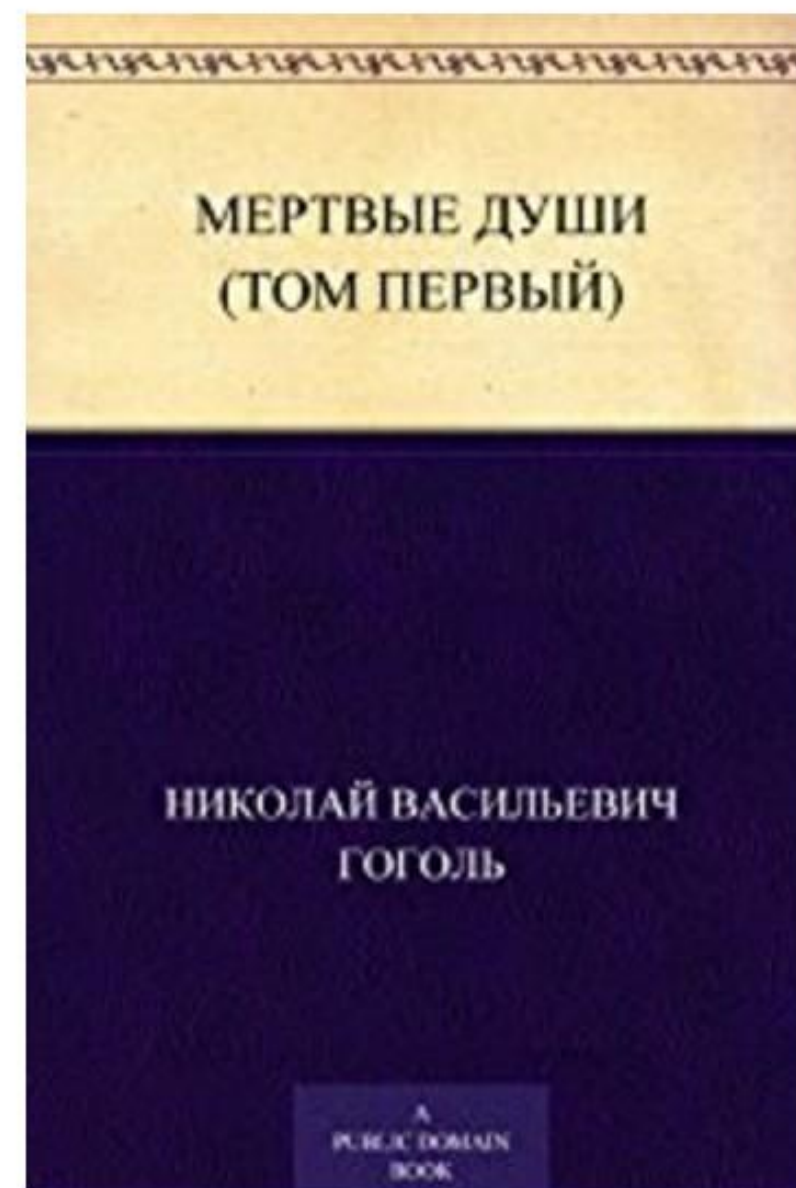


About

Settings

amazon.com.

Tural, what do you think? Please share your opinion with others on Amazon.com.



Мертвые души (Том первый)
(Russian Edition)

Николай Васильевич Гоголь

Start by rating it



About

—

FEEDBACK



НЕТОЛОГИЯ
групп

МНЕНИЕ ПОЛЬЗОВАТЕЛЯ

- Explicit feedback
- Пользователь явно (explicitly) сообщает системе свое мнение об объекте
- Рейтинг, Like/dislike, Upvote/Downvote
- Implicit feedback
- Система делает вывод о мнении пользователя по действиям, совершенным пользователем
- Клики, покупки, просмотры

EXPLICIT FEEDBACK

- Самый информативный feedback
- Пользователи редко ставят рейтинги (требуется от пользователя когнитивных усилий)
- Доступен не во всех областях (новостные порталы)
- Актуален только в контексте в котором он оставлен

IMPLICIT FEEDBACK

- Действие != отзыв
- Кликнул != понравилось
- Не кликнул != не понравилось
- Действие = потенциальный отзыв
- Легко собрать информацию о действиях пользователя (много информации)
- Не требует от пользователя когнитивных усилий

ТИПЫ RS

ГЛОБАЛЬНО

1. Неперсонализированные

- Не популярны в академической среде
- Популярны в продакшине

2. Персонализированные

НЕПЕРСОНАЛИЗИРОВАННЫЕ

1. Строятся по данным от всех пользователей (агрегированные данные)

Средний рейтинг фильма

2. Строятся для среднего пользователя

НЕПЕРСОНАЛИЗИРОВАННЫЕ

1. Строятся по данным от всех пользователей (агрегированные данные)

Средний рейтинг фильма

2. Строятся для среднего пользователя

3. **Одинаковые** для всех пользователей

Возможны рандомизация и фильтрация объектов

4. Гарантируют, что пользователь получит рекомендации

НЕПЕРСОНАЛИЗИРОВАННЫЕ (2)

New Releases [See Top 100](#)



Lego Batman...
Blu-ray
\$9.99 ✓prime



Wonder...
Blu-ray
\$24.99 ✓prime



Shin Godzilla...
Blu-ray
\$24.99 ✓prime

Movers & Shakers [See Top 100](#)



Jesse Stone:...
DVD
\$15.12 ✓prime



Jesse Stone:...
DVD
\$9.52 ✓prime



Jesse Stone:...
DVD
\$11.96 ✓prime

250 лучших фильмов + ещё немного

Рейтинг составлен по результатам голосования посетителей сайта. Любой желающий может принять в нем участие, проголосовав за свой любимый фильм. [Как рассчитывается Топ-250?](#)

[← предыдущий день](#)

фильм				
1.	Побег из Шоушенка (1994) The Shawshank Redemption	9.197 (427 853)	★	▼
2.	Зеленая миля (1999) The Green Mile	9.148 (407 274)	★	▼
3.	Форрест Гамп (1994) Forrest Gump	9.001 (400 415)	★	▼
4.	Список Шиндлера (1993) Schindler's List	8.895 (215 122)	★	▼
5.	1+1 (2011) Intouchables	8.865 (415 971)	★	▼
6.	Начало (2010) Inception	▲ 1 8.768 (454 567)	★	▼
7.	Король Лев (1994) The Lion King	▲ 1 8.767 (279 331)	★	▼
8.	Леон (1994) Léon	▼ 2 8.763 (333 657)	★	▼

ПЕРСОНАЛИЗИРОВАННЫЕ

1. Персональны для каждого пользователя
2. Для построения необходима информация о предпочтениях пользователя
3. Нет гарантий, что пользователь получит рекомендации

КЛАССЫ ПОДХОДОВ

КЛАССЫ ПОДХОДОВ

1. Content-based



КЛАССЫ ПОДХОДОВ

1. Content-based
2. Collaborative Filtering



КЛАССЫ ПОДХОДОВ

1. Content-based
2. Collaborative Filtering
3. Demographic



КЛАССЫ ПОДХОДОВ

1. Content-based
2. Collaborative Filtering
3. Demographic
4. Knowledge-based



КЛАССЫ ПОДХОДОВ

1. Content-based
2. Collaborative Filtering
3. Demographic
4. Knowledge-based
5. Community-based



КЛАССЫ ПОДХОДОВ

1. Content-based
2. Collaborative Filtering
3. Demographic
4. Knowledge-based
5. Community-based
6. Hybrid



ПОСТРОЕНИЕ РЕКОМЕНДАЦИЙ

РЕКОМЕНДАЦИИ VS. ПРЕДСКАЗАНИЯ

- Рекомендательные алгоритмы предсказывают (рейтинги, действия)
- Предсказание != рекомендации
- Предсказанные объекты — кандидаты на рекомендацию

SCORE VS. RANK

Scoring — вычисление значения для оценки объекта

(предсказание рейтинга)

Ranking — ранжирование объектов для построения списка рекомендаций

Для ranking не обязателен scoring

COLD-START PROBLEMS

Возникают когда нет достаточного объема знаний для построения рекомендаций

Cold-start — новая система, нет никаких данных

Cold user — новый пользователь

Cold item — новый объект

—

EVALUATION

ОЦЕНКА RS

Цель: узнать насколько хорошо работает RS

1. Оффлайн оценка

Оценивает качество работы алгоритма

Оценивает насколько хорошо алгоритм имитирует исходные данные

ОЦЕНКА RS

Цель: узнать насколько хорошо работает RS

1. Оффлайн оценка

Оценивает качество работы алгоритма

Оценивает насколько хорошо алгоритм имитирует исходные данные

2. Онлайн оценка

Оценка качества работы системы

Оценка рекомендаций

ТРАДИЦИОННЫЕ МЕТРИКИ

1. Accuracy and error metrics (оценка предсказанного значения)

MAE, MSE, RMSE

ТРАДИЦИОННЫЕ МЕТРИКИ

2. Decision support metrics

оценка выбора

Precision, Recall, F1, MAP, ROC

ТРАДИЦИОННЫЕ МЕТРИКИ

3. Rank metrics

оценка ранжирования

nDSG, MRR, MPR

ТРАДИЦИОННЫЕ МЕТРИКИ

4. User-centered metrics

разное

Интервью, опросы, A/B

ТРАДИЦИОННЫЕ МЕТРИКИ

1. Accuracy and error metrics (оценка предсказанного значения)
MAE, MSE, RMSE
2. Decision support metrics (оценка выбора)
Precision, Recall, F1, MAP, ROC
3. Rank metrics (оценка ранжирования)
nDSG, MRR, MPR
4. User-centered metrics (разное)
Интервью, опросы, A/B

OFFLINE СРАВНЕНИЕ

1. Нельзя сравнивать абсолютные значения метрик, полученные:
 - на разных датасетах
 - в разных условиях
 - для разных областей
2. Сравнение проводится на одном и том же датасете при одинаковых условиях

КОММЕРЧЕСКАЯ СОСТАВЛЯЮЩАЯ

! Бизнесу не интересна оффлайн точность рекомендательного алгоритма

Важны такие показатели как:

- конверсия
- количество продаж
- покрытие каталога
- и т.д.

ДРУГИЕ МЕТРИКИ

- Coverage — покрытие каталога
- Novelty — новизна объектов в рекомендациях
- Diversity — разнообразие объектов в рекомендациях
- Serendipity — эффект неожиданности

ПРЕДСТАВЛЕНИЕ РЕКОМЕНДАЦИЙ

ПРЕЗЕНТАЦИЯ

1. **Explicit** — “яркое” указание на рекомендацию (только сегодня и только для вас)

пользователь может посчитать, что его пытаются обмануть

ПРЕЗЕНТАЦИЯ

1. **Explicit** — “яркое” указание на рекомендацию (только сегодня и только для вас)

пользователь может посчитать, что его пытаются обмануть

2. **Organic** — не “кричащая” презентация рекомендаций


пользователь может не заметить рекомендации

На практике пытаются комбинировать оба метода

EXPLICIT VS. ORGANIC

EXPLICIT VS. ORGANIC

-51%
TODAY!



Eastland Suites Hotel & Conf...
Urbana, Champaign

Last booked: 8 hours ago

.genius10% DISCOUNT

Today's Value Price

See more

Double Room

In high demand!

Risk free: You can cancel later, so lock in this great price today.

Very good 8.5
131 reviews

Price for 4 nights

-51%

€ 428


€ 208

FREE cancellation

No prepayment needed

View property & book >

49



**-51%
TODAY!**

Eastland Suites Hotel & Conf...

Urbana, Champaign

★★★★★

Very good 8.5

131 reviews

The Circle

by Dave Eggers

★★★★★ 3.46

Rating Details · 126,395 Ratings · 17,672 Reviews

When Mae Holland is hired to work for the Circle, the world's most powerful internet company, she feels she's been given the opportunity of a lifetime. The Circle, run out of a sprawling California campus, links users' personal emails, social media, banking, and purchasing with their universal operating system, resulting in one online identity and a new age of civility and ...more

Read

My rating: ★★★★★

Listen

Hardcover, First edition, 493 pages
Published October 8th 2013 by Knopf

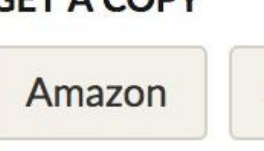
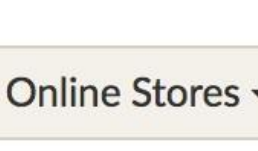


More Details...

GET A COPY

Amazon Online Stores Libraries

Recommend It | Stats | Recent Status Updates

READERS ALSO ENJOYED

GENRES

Fiction	1,067 users
Science Fiction	712 users
Science Fiction > Dystopia	547 users

ОБЪЯСНЕНИЕ

Пользователь не всегда понимает, почему ему был рекомендован объект

Объяснение должно быть простым и наглядным
*объект приобретают с тем, что у вас в корзине
потому что вам понравились объекты A и B*

Объяснение не должно требовать дополнительных знаний
не надо показывать математические выкладки

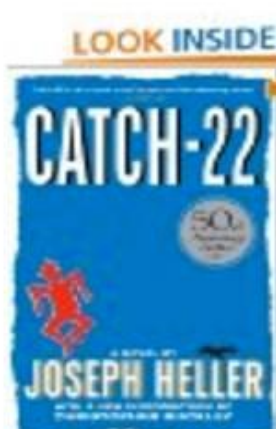
EXPLICIT VS. ORGANIC

ПОТОМУ ЧТО



Animal farm: A Fairy Story
by George Orwell (April 6, 2004)
Average Customer Review: ★★★★★ (3,906)
In Stock
List Price: \$9.99
Price: \$6.83
[411 used & new](#) from \$0.25

☐ I own it ☐ Not interested ☒ ★★★★★ Rate this item
Recommended because you rated **Brave New World** ([Fix this](#))



Catch-22: 50th Anniversary Edition
by Joseph Heller (April 5, 2011)
Average Customer Review: ★★★★★ (1,742)
In Stock
List Price: \$18.00
Price: \$9.87
[281 used & new](#) from \$3.57

☐ I own it ☐ Not interested ☒ ★★★★★ Rate this item
Recommended because you rated **Brave New World** ([Fix this](#))

KASABIAN
GORILLAZ
DEMON DAYS

dodgy

Radio station
You've listened to this in the evening

GORILLAZ
Oxxxymiron
You've listened to this in the evening

Working to a Beat

Deep Techno concentration

Get into the universe of deep techno music which will surround you with the unbreakable wall of concentration.

Trip Hop Classics

Trip Hop emerged in the mid-1990s, when the Hip Hop & House music scene merged into an exciting and innovative hybrid. Listen to the...

—

МИРЫ RS

ЕЩЕ RS

- Sequence Mining RS
- Context-aware RS
- Group RS
- Social-based RS
- Personality-based RS

ЕЩЕ ЗАДАЧИ

- Active Learning
- Cross-domain
- Multi-domain
- Privacy
- Security

PRODUCTION RS

RS as a Service (черный ящик)

Yusp (<http://www.yusp.com/>)

! Не обязательно
строить свой
велосипед

Frameworks

Turi (<https://turi.com/>)

Oryx 2 (<http://oryx.io>)

Spark (<https://spark.apache.org/>)

Mahout (<https://mahout.apache.org>)

PRODUCTION RS

RS as a Service (черный ящик)

Yusp (<http://www.yusp.com/>)

! Не обязательно
строить свой
велосипед

Frameworks

Turbo (<http://turbo.build/>)

Oryx (<http://oryx.io>)

Spark (<https://spark.apache.org/>)

Mahout (<https://mahout.apache.org>)

RECSYS CHECKLIST

1. Определить что и кому рекомендуем
2. Определить цель построения рекомендаций
3. Определить метрики для достижения цели
4. Определить метрики оценки системы (оффлайн и онлайн)
5. Подготовить данные
6. Построить систему, оптимизирующую метрики
7. Определить способ отображения рекомендаций
8. Оценить качество рекомендаций

ПРАКТИКА

https://grouplens.org/datasets/movielens/latest
/

ПОМОГАТЕЛЬНЫЙ МАТЕРИАЛ

ВОПРОСЫ