

# Введение и классификация рекомендательных систем

Мурат Апишев  
NLP Team Lead



# План занятия

- 1 Что такое RS?
- 2 Типы RS
- 3 Типы данных для RS
- 4 Предсказания и рекомендации
- 5 Примеры RS



# Терминология

**Пользователь (user)** — сущность, для которой строятся рекомендации (пользователь сервиса, PR-агентство)

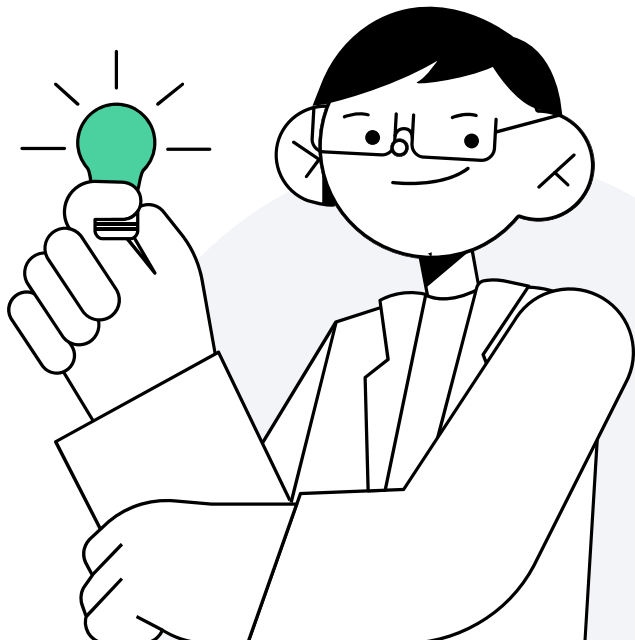
**Объект (item)** — сущность, которую рекомендуют (фильмы, книги, туристический маршрут)

**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 объём информации для ознакомления (мы никогда не изучаем все доступные YouTube ролики)
- Помогает пользователю принять решение



## Recommended



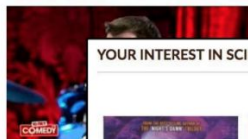
**British Airways safety video - director's cut**  
British Airways  
4,923,360 views • 1 week ago



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



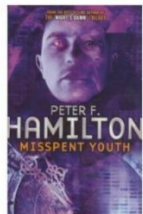
**Man City vs Tottenham Extended Highlights**  
TheFinalWhistle HD  
894,623 views • 1 day ago



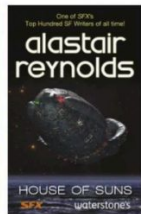
**Garik Харин Батрутдин**  
Old Spice Russia  
23,393,866 views



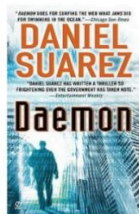
## YOUR INTEREST IN SCIENCE FICTION



Want to Read  
★★★★★  
Not interested



Want to Read  
★★★★★  
Not interested

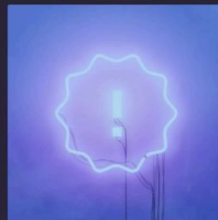


Want to Read  
★★★★★  
Not interested



Want to Read  
★★★★★  
Not interested

## Recommended new releases



New Release Radio



Stony Hill



Add Violence



Ritual

Images saved in Drive

Background...

Layout...

Theme...

Transition...

Present

Comments

Share

Explore

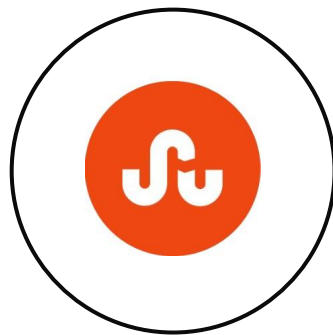
Search your docs and the web



Click to add title  
Click to add text

# Функции RS

- **Рекомендации:** объекты, пакеты объектов, последовательности объектов
- **Аннотации:** указать пользователю передачи в ТВ- программе, которые ему могут быть интересны
- **Браузинг:** StumbleUpon, Surfingbird



# Для чего они нужны

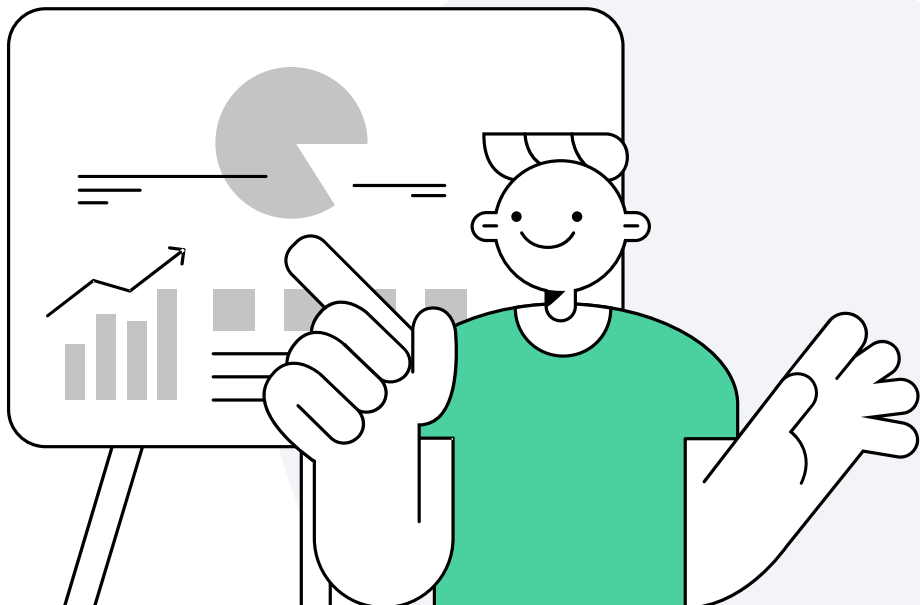
- Увеличить количество продаж (просмотров/кликов, CTR, ...)
- Продавать различные объекты (не только самые популярные)
- Увеличить удовлетворение пользователя от использования сервиса
- Увеличить лояльность пользователя к сервису
- Лучше понимать пользователя (лучше понимать, что нужно пользователю)

# 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

# Пользователь

- Характеристики пользователя
- Социо-демографические данные
- Личностные (personality)
- Информация о взаимодействиях с объектами
- Поведенческие паттерны

## 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

- Рейтинги (бинарные, численные)
- Теги (свойства, которые пользователи присваивают объектам)
- Действия/события
- Время
- Контекст



Like



Love



Haha



Yay



Wow



Sad



Angry

# Когда собирать?

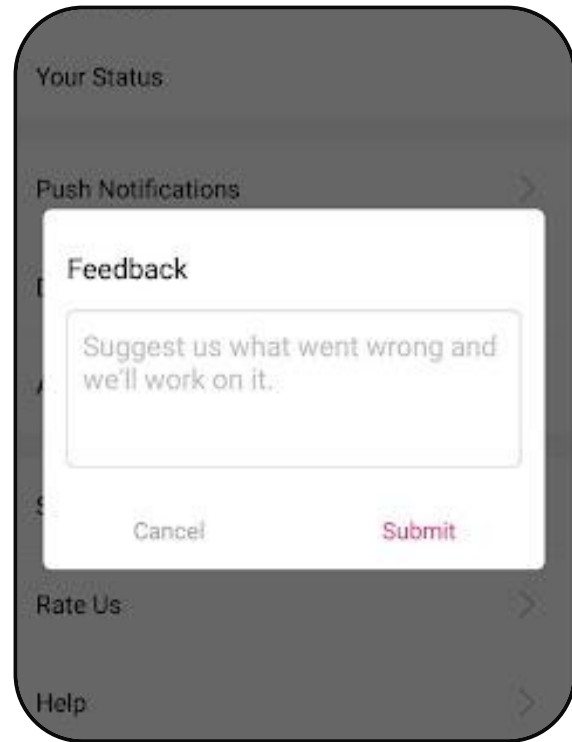
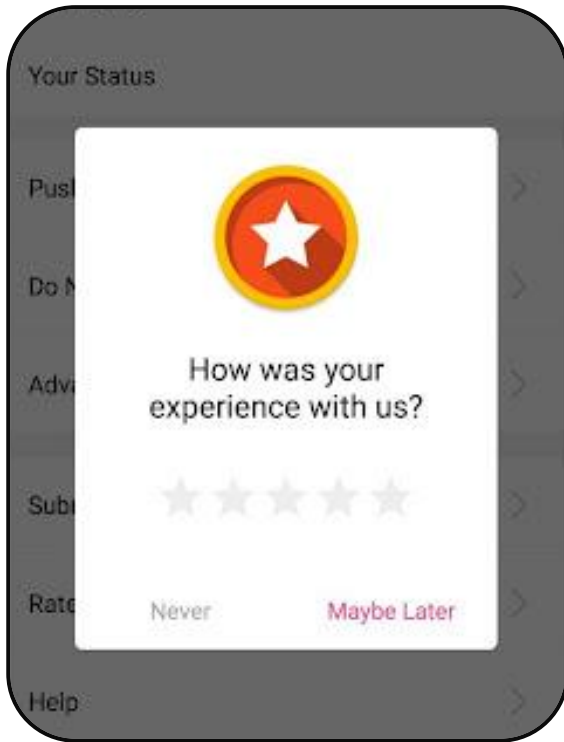
- До взаимодействия пользователя с объектом (ожидаемое мнение)
- В момент знакомства пользователя с объектом
- В момент взаимодействия пользователя с объектом
- Сразу после взаимодействия пользователя с объектом
- Через некоторое время после взаимодействия пользователя с объектом



# Когда собирать?



# Когда собирать?



# Когда собирать?

Притяжение

2017, Россия, 12+



## Оцените этот фильм

Оценки улучшают персональные рекомендации





# Мнение пользователя

① **Explicit feedback** — пользователь явно (explicitly) сообщает системе свое мнение об объекте

- Рейтинг
- Like/dislike
- Upvote/Downvote

② **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

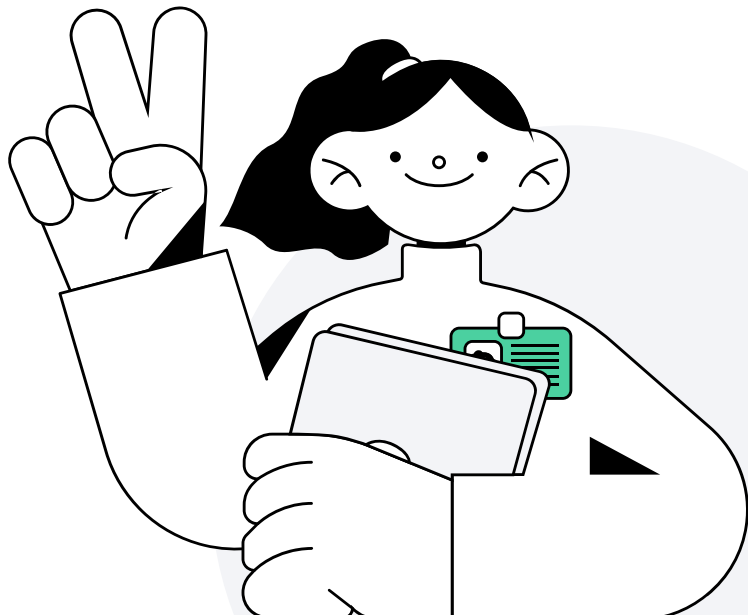
- Самый информативный feedback
- Действие != отзыв
- Кликнул != понравилось
- Не кликнул != не понравилось
- Действие = потенциальный отзыв
- Легко собрать информацию о действиях пользователя (много информации)
- Не требует от пользователя когнитивных усилий

# Типы RS

## ① Неперсонализированные

- Не популярны в академической среде
- Популярны в продакшене

## ② Персонализированные



# 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

## 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: In the Company of Men  
DVD  
\$15.12 ✓prime



Jesse Stone: Lost in Translation  
DVD  
\$9.52 ✓prime



Jesse Stone: Night Passage  
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

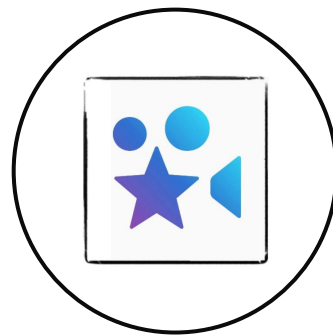
- |    |  |                     |   |   |
|----|--|---------------------|---|---|
| 1. | <a href="#">Побег из Шоушенка (1994)</a><br>The Shawshank Redemption | 9.197 (427 853)     | ★ | ▼ |
| 2. | <a href="#">Зеленая миля (1999)</a><br>The Green Mile                | 9.148 (407 274)     | ★ | ▼ |
| 3.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

- Content-based
- Collaborative Filtering
- Demographic
- Knowledge-based
- Community-based
- Hybrid





# Введение и классификация рекомендательных систем

Мурат Апишев  
NLP Team Lead

