



# Системы рекомендаций

Коноплёва Екатерина Лицей НИУ ВШЭ

Цель работы: Выявить механизм работы систем рекомендаций для грамотного взаимодействия с различными Интернет ресурсами, сформировать представления о возможных ошибках в их работе, представив возможные решения.

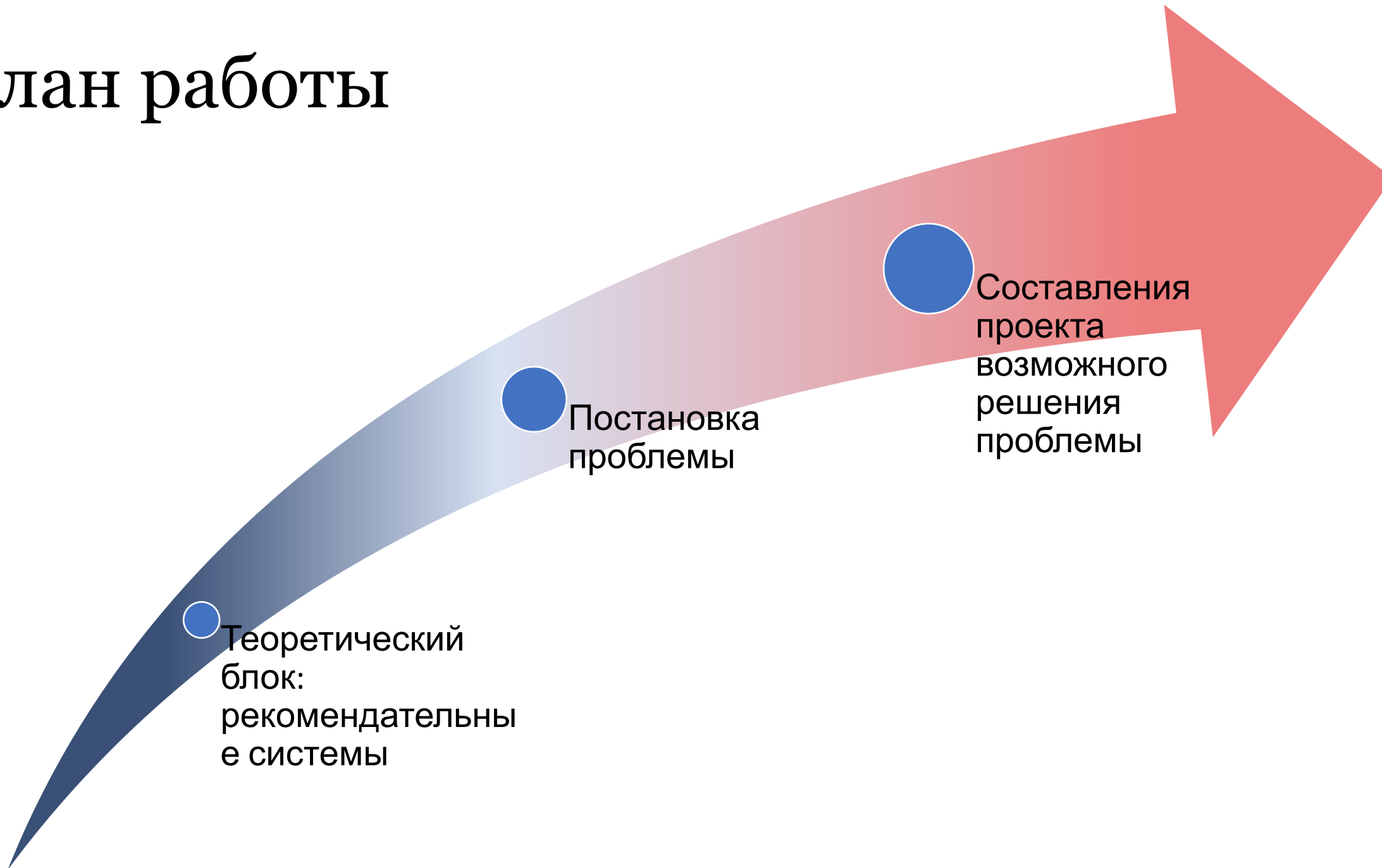
Объект исследования: системы рекомендаций



# Задачи:

1. Изучить методы работы рекомендательных систем, их свойства и особенности.
2. Определить проблемы возникающие при работе с рекомендательными системами
3. Выделить проблему и рассмотреть её на конкретном примере
4. Определить возможный путь практического решения проблемы
5. Сделать презентацию по итогам изучения темы и выступить с ней на школьной конференции проектных работ

# План работы



# Актуальность





Более 4,8  
миллиарда  
используют  
Интернет, это 59%  
всего населения  
Земли (данные на  
момент начало  
2020 года)



[https://techblog.sdstudio.top/107-internet-statistika-2020-okonchatelnyj-spisok-faktov-i-tendencij/#1\\_48\\_milliarda\\_celovek\\_ispolzuyut\\_internet\\_azednevno](https://techblog.sdstudio.top/107-internet-statistika-2020-okonchatelnyj-spisok-faktov-i-tendencij/#1_48_milliarda_celovek_ispolzuyut_internet_azednevno)



Машинное обучение - подобласть искусственного интеллекта. Подход, полагающийся на выучивание закономерностей из имеющихся размеченных или неразмеченных данных -



Используется не прописанный алгоритм решения, а данные на основе, которых система обучается.

Задача:

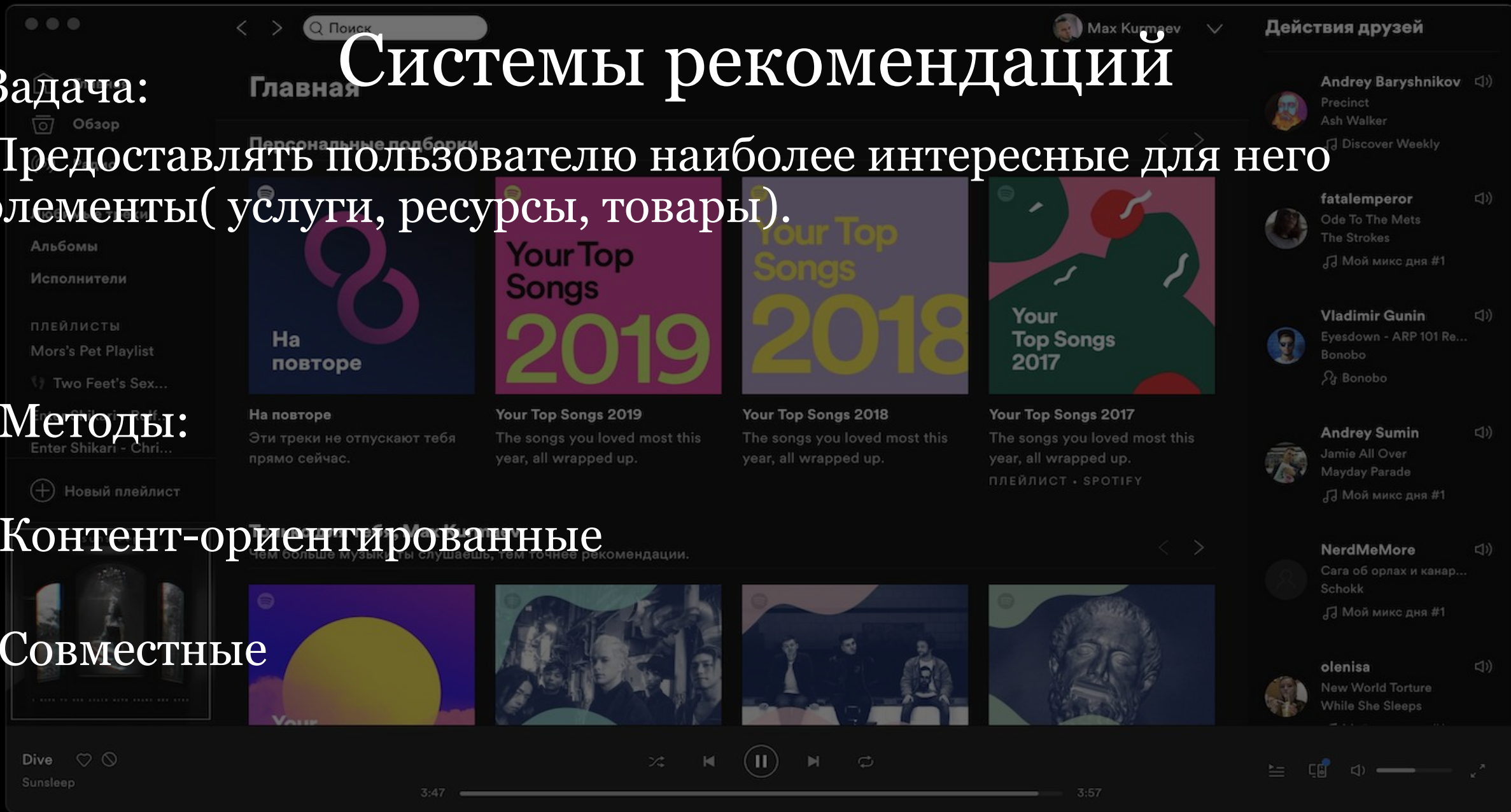
Предоставлять пользователю наиболее интересные для него элементы (услуги, ресурсы, товары).

Методы:

Контент-ориентированные

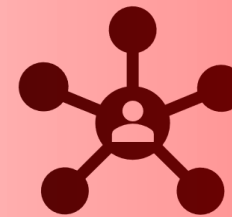
Совместные

# Системы рекомендаций





# Методы совместной фильтрации



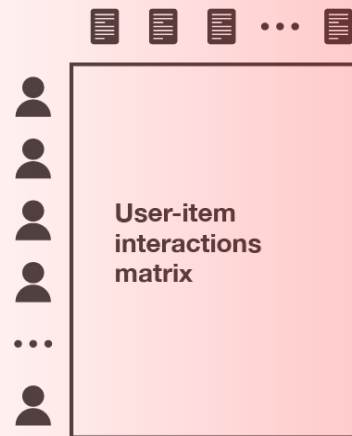
Используемые данные:

Прошлые взаимодействия между пользователями и элементами

Общая схема работы:

1. Формирования множества пользователей с аналогичными предпочтениями

2. Предсказания на основе степени близости пользователей и популярности наиболее подходящих элементов

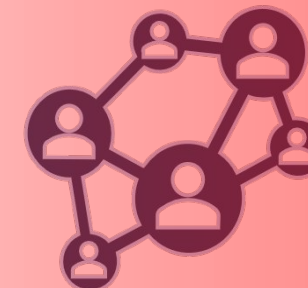


| Users       | User-item interactions matrix                   | Items    |
|-------------|---|----------|
| subscribers | rating given by a user to a movie (integer)     | movies   |
| readers     | time spent by a reader on an article (float)    | articles |
| buyers      | product clicked or not when suggested (boolean) | products |
|             | ...   |          |

Существует деление на

Основанными на памяти подходы  
непосредственный поиск ближайших соседей

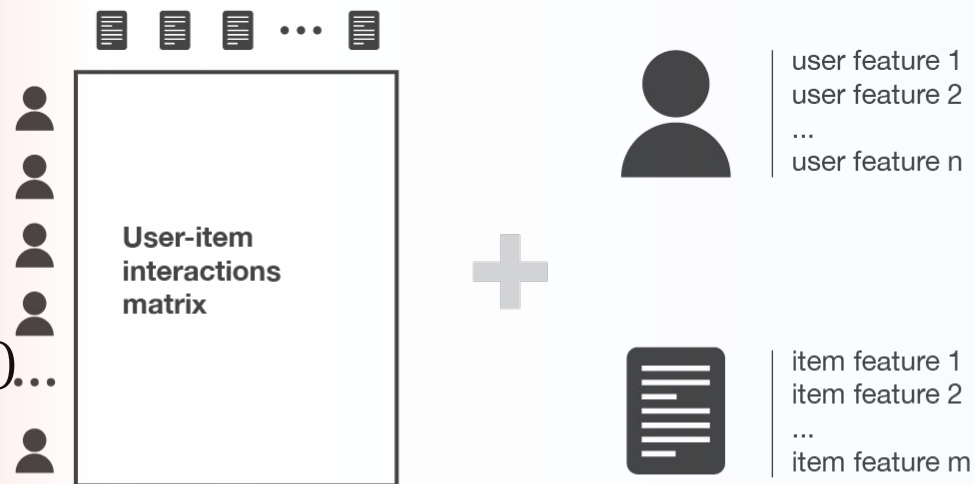
Модельные подходы  
нахождения рейтинга элемента у пользователей



# Контент-ориентированные методы

Используемые данные:

Свойства пользователя и элементов, с которыми пользователь уже взаимодействовал (дополнительная информация)...



Общая схема работы

1. Выявления свойств пользователя и элементов
2. Построения моделей, выявления зависимости
3. Предсказания на основе модели

**Collaborative information**

(The user-item interactions matrix)

**Content information**

Can be users or/and items features

**Model**

Takes user or/and items features and returns predicted interactions

Оценка TF-IDF характеризует важность слова в конкретном документе  
Дается на основе частоты употреблений слова в конкретном документе и в других документах



# Гибридные

Методы, которые сочетают в себе методы совместной фильтрации и контент-ориентированного подхода

Очень распространены

Две формы:

Две модели независимо (одну модель совместной фильтрации и одну модель на основе контента) обучаются, затем их предложения объединяются

Единая модель (часто нейронная сеть), объединяющая оба подхода используя в качестве входных данных предварительную информацию (о пользователе и / или элементе) и информацию о взаимодействии.



# Проблемы рекомендательных систем: «область ограниченной информации»

Рекомендательная система работающая слишком хорошо может в результате оказать негативный эффект.

Рекомендуя, только однотипные элементы и только усиливая привязку именно к такому типу элементов.

Методы совместной фильтрации

степень близости с пользователями используя реакции которых делаются прогнозы будет увеличиваться, вероятность, что в следующем прогнозе будет использоваться то же множества растет

Контент-ориентированные методы

В начальный момент времени, когда пользователь выбирает первый элемент, следующая рекомендация будет соответствовать одному элементу и выбор любого из рекомендованных элементов ещё больше закрепит свойства первого элемента ( важно, что когда количество элементов имеющих один общий признак он же окажется в подборку)



# Проблемы рекомендательных систем: Популярно – популярнее – ещё популярнее

В рекомендательных системах, где используются методы совместной фильтрации важным критерием является популярность элемента

Популярные элементы это те, что вырыли большое количество пользователей.



Рекомендательная система находит начинает рекомендовать всё более и более популярные элементы. Остальные признаки не могут закрепиться из-за слишком большого количества разных предпочтений

# Проблемы рекомендательных систем: Вывод

Элементы, которые вы выбираете всегда влияют на дальнейшие рекомендации



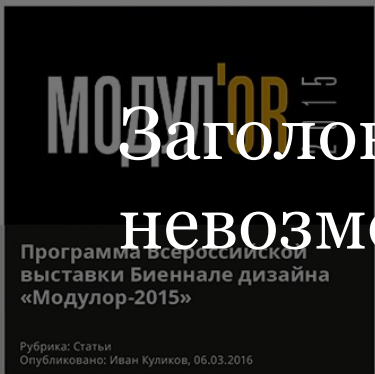
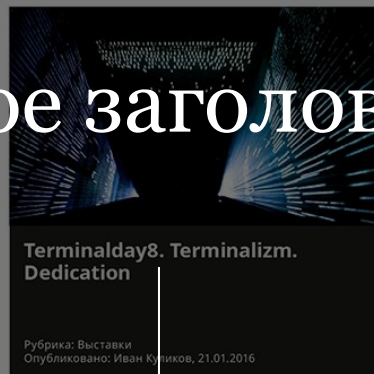
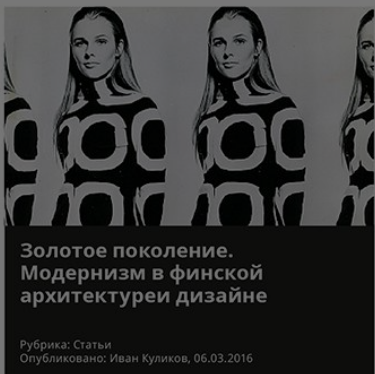
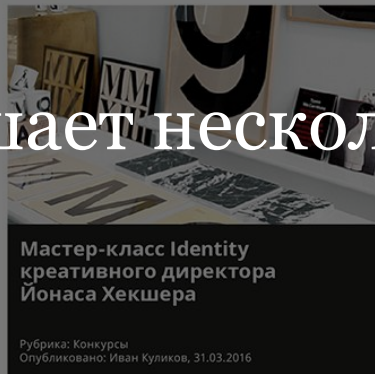
Выбор элементов из рекомендованного важен, для последующих рекомендаций



Сайт — новостная лента, информирующая о самых важных и интересных событиях в сфере дизайна: конкурсах и новых проектах, лекциях и творческих встречах, выставках и публикации.

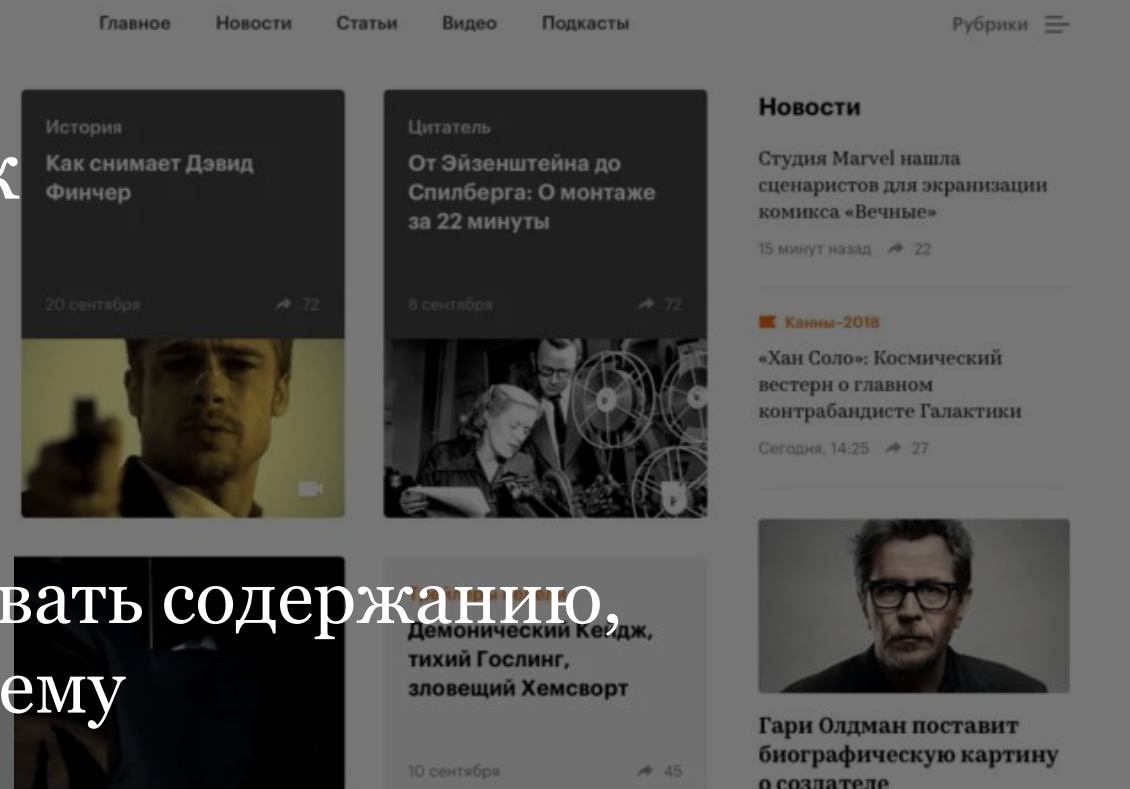
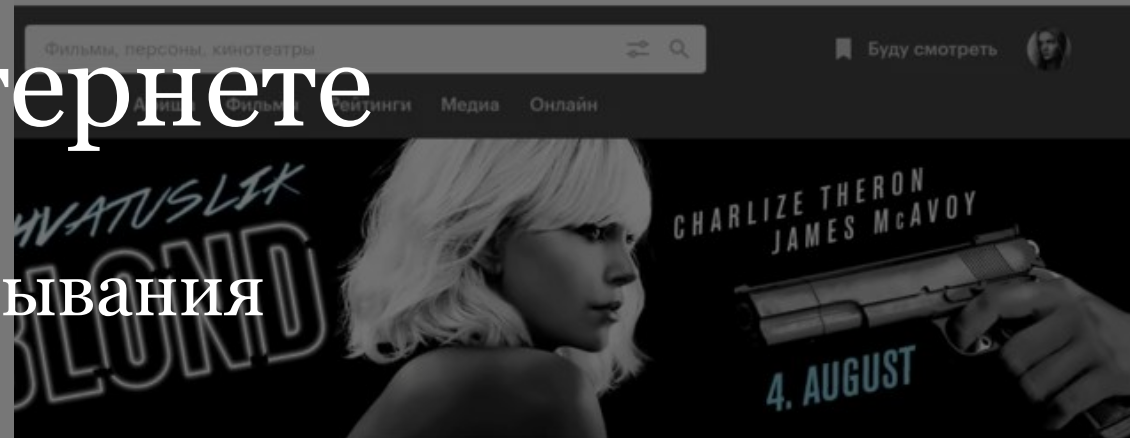
# Статьи и новости в Интернете

Всё решает несколько строк названия статьи



Главное заголовков

Заголовков может не соответствовать содержанию, невозможно точно определить тему



# Информационные ресурсы

Дзен  
Публикации на основе ваших интересов

Яндекс.Браузер  
со встроенным Дзеном

Лента, составляется системой рекомендаций

3DNEWS  
HI-NEWS.RU  
Домашний Очаг  
АВТОВЗГЛЯД  
леди@mail.ru  
woman's day

Общего каталога может не быть



Для комфортного использования необходимо следить за выбором

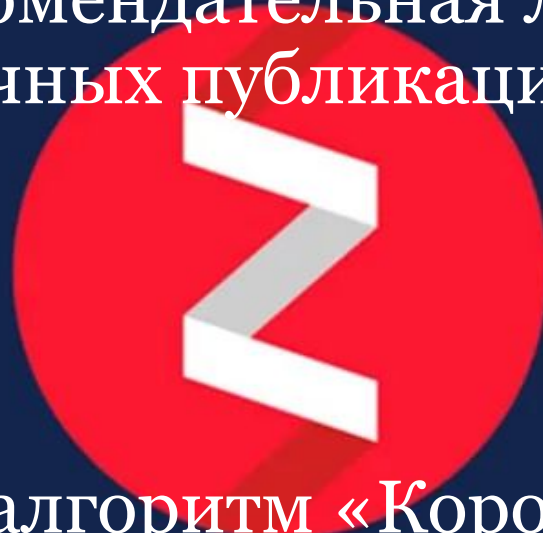
Используются на постоянной основе, история сохраняется

Видео  
Популярные ролики

03:24  
04:16 HD  
01:22  
17:00  
17:30  
17:00  
00:44  
00:38  
17:25  
17:30  
16:30

## Конкретный пример: Яндекс Дзен

Интернет-сервис, рекомендательная лента контента и платформа для различных публикаций.



Частично использует алгоритм «Королёв», который использует нейронную сеть

# Яндекс Дзен



# Теоретическое решение

При просмотре ленты видны только названия



Нужна дополнительная информация, максимально емко выраженная о соответствии статьи указанным пользователем предпочтений



Создание дополнительной системы «маркировки» новостей в соответствии с заранее указанными требованиями пользователя

# Теоретическое решение

1. Несколько статей выделяются как статьи образцы ( по каждой интересующей теме, как минимум)
2. При просмотре ленты считывается полный текст статей
3. Используется алгоритм контент-ориентированного метода для сравнения текстов образцов и нового
4. На статья отображается цветовой маркер соответствия ( зеленый – обнаружено соответствие, красный – нет)





Спасибо за внимание



# Библиография

- <https://texterra.ru/blog/kak-rabotaet-umnaya-lenta-razdel-rekomendatsii-i-algoritm-prometey-unikalnaya-informatsiya-ot-komand.html>
- <https://cmsmagazine.ru/journal/research-7-algoritmov-tovarnyh-rekomendacij/ps://esputnik.com/blog/tovarnye-rekomendacii-na-sajte-instrument-povysheniya-konversii>
- <https://ux.pub/ux-principy-dlya-personalnyh-rekomendacij-kontenta/>
- <https://neurohive.io/ru/osnovy-data-science/rekomendatelnye-sistemy-modeli-i-ocenka/>
- <https://vc.ru/ml/132779-mashinnoe-obuchenie-rekomendatelnye-sistemy>
- [https://www.ispras.ru/proceedings/docs/2012/22/isp\\_22\\_2012\\_401.pdf](https://www.ispras.ru/proceedings/docs/2012/22/isp_22_2012_401.pdf)
- <https://lpgenerator.ru/blog/2015/12/25/rekomendatelnye-sistemy-cto-eto/>
- <https://www.hse.ru/data/2012/05/29/1252471276/5.pdf>
- <http://intsysjournal.ru/pdfs/21-4/18-30-Kutanin.pdf>
- <https://www.machinelearningmastery.ru/introduction-to-recommender-systems-6c66cf15ada/>
- [https://logic.pdmi.ras.ru/~sergey/slides/N14\\_AIUkraine.pdf](https://logic.pdmi.ras.ru/~sergey/slides/N14_AIUkraine.pdf)
- <http://datareview.info/article/sistemyi-rekomendatsiy-zadachi-podhodyi-algoritmyi/>

# Библиография

[http://web.iis.nsk.su/files/dzyuba\\_m\\_thesis.pdf](http://web.iis.nsk.su/files/dzyuba_m_thesis.pdf)

<https://cyberleninka.ru/article/n/algorithmy-dlya-rekomendatelnoy-sistemy-tehnologiya-lenkor>

<https://www.dissercat.com/content/issledovanie-i-razrabotka-algoritmov-rekomendatelnikh-sistem-na-osnove-grafovykh-modelei-dan>

<https://www.pvsm.ru/algorithmy/289907>

<https://zen.yandex.ru/media/unclemod/algorithmy-i-adeks-dzen-ili-kak-povliat-avtoru-na-pokazy-5efb4924b0a97609b1ead09f>

<https://texterra.ru/blog/kak-rabotayut-algoritmy-yandeks-dzena-intervyu-s-media-direktorom-servisa-daniilom-trabunom.html>

[https://yandex.ru/q/question/algorithmy\\_iadeksa\\_menaiut\\_stranitsy\\_v\\_2f3b9544/?utm\\_source=yandex&utm\\_medium=wizard&answer\\_id=f135d94a-437c-4783-b950-13b5ff5c6977#f135d94a-437c-4783-b950-13b5ff5c6977](https://yandex.ru/q/question/algorithmy_iadeksa_menaiut_stranitsy_v_2f3b9544/?utm_source=yandex&utm_medium=wizard&answer_id=f135d94a-437c-4783-b950-13b5ff5c6977#f135d94a-437c-4783-b950-13b5ff5c6977)

<https://batareika.media/zen/algorithm.html>

[https://sf.ru/local/templates/new/assets/files/Posobie\\_Poznat\\_JaZen.pdf](https://sf.ru/local/templates/new/assets/files/Posobie_Poznat_JaZen.pdf)

<https://uspei.com/servisy-programmy/algorithmy-i-iskusstvennyj-intellekt-yandeks-dzena/>

<https://internet--technologies-ru.turbopages.org/internet-technologies.ru/s/news/obnovlenie-algoritma-rekomendaciy-ot-yandeks-dzen.html>

<https://spyserp.com/ru/blog/yandex-dzen>

<https://tass.ru/obschestvo/6401818>

<https://www.hse.ru/data/2017/05/14/1171296413/Григорий%20Сапунов%20—%20Введение%20в%20машинное%20обучение.pdf>

<https://texterra.ru/blog/kak-rabotayut-algoritmy-yandeks-dzena-intervyu-s-media-direktorom-servisa-daniilom-trabunom.html>