

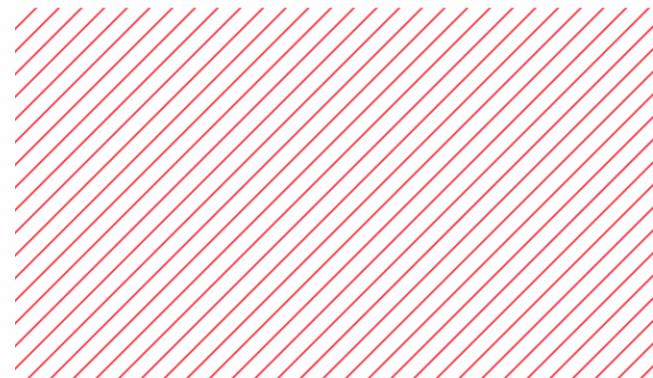
академия
больших
данных



Information retrieval intro

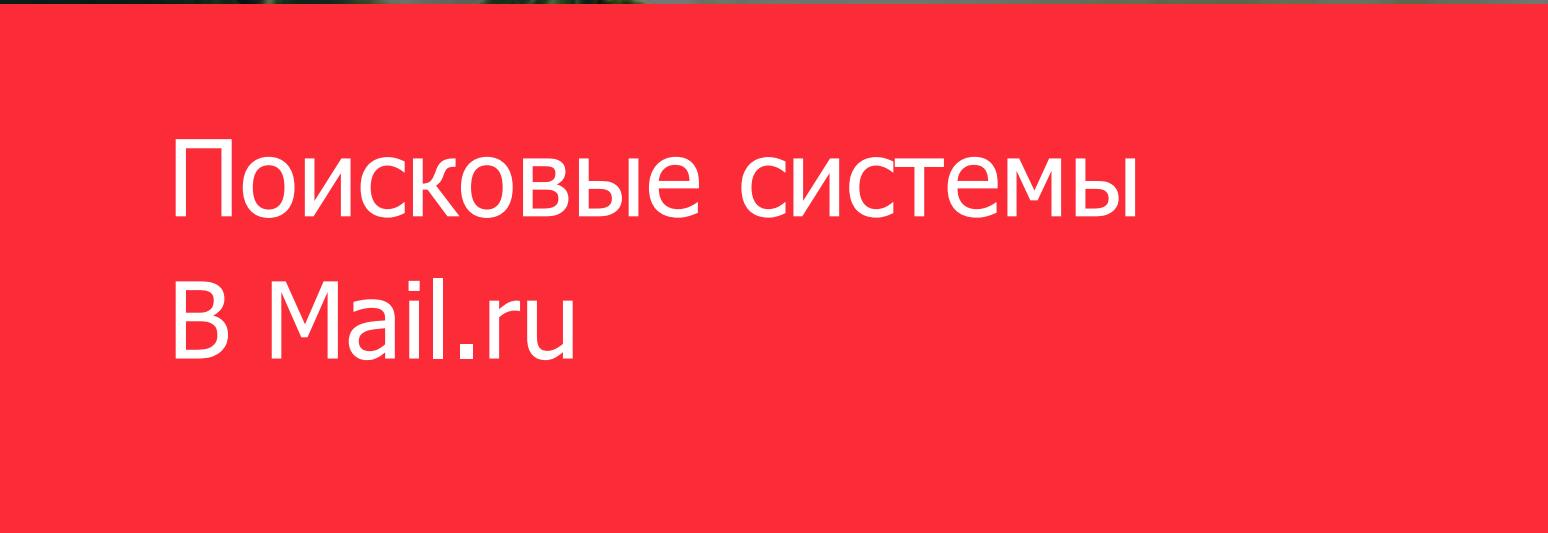
Владимир Гулин

Руководитель направления поисковых и
рекомендательных технологий

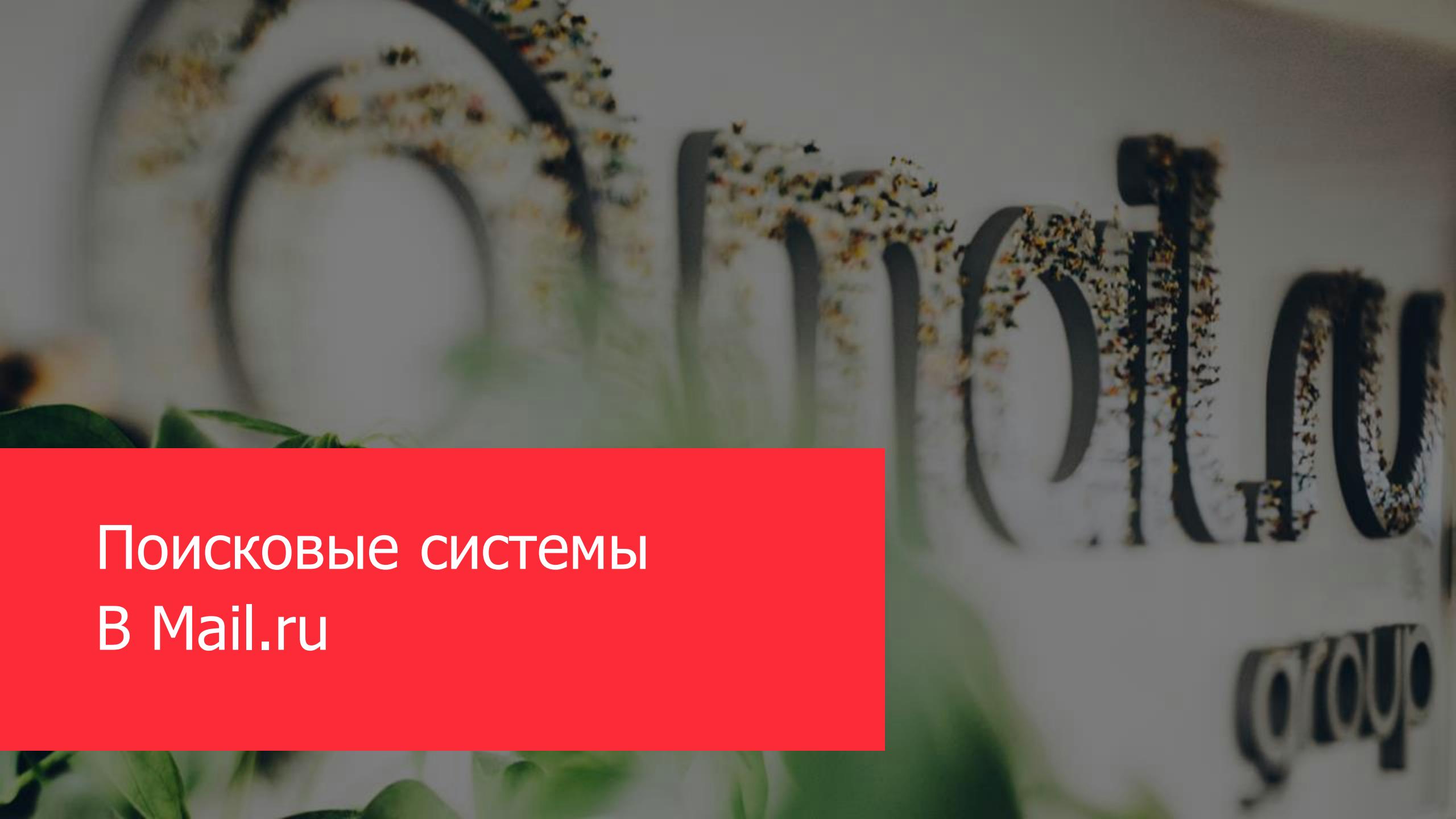


План

1. Поисковые и рекомендательные системы в Mail.ru
2. Задачи ML в информационном поиске
3. Смежные задачи информационного поиска
4. Задача обучения ранжированию (семинар)



Поисковые системы
В Mail.ru



mail.ru
group

Продукты

Развиваем поиски в продуктах компании

- Портальный поиск Mail.ru
- Поиск в социальных сетях
- Поиск в e-сом приложениях
- ...

Портальный поиск



поиск джокер фильм 2019 Найти

Интернет Соцсети beta Картинки Видео Новости Ответы

• **Джокер (фильм, 2019) смотреть онлайн в hd 720 качестве — 7 книг**
livelib.ru/selection/1475535...
Фильм ДЖОКЕР 2019 смотреть на русском с субтитрами. Последний трейлер Джокера был только что выпущен, и вы можете проверить, что у вас нет раннего...

• **Джокер (2019) смотреть онлайн или скачать торрентом: утечка фильма...**
soskidkami.ru/news/dzhoker-2019...
Самая горячая новинка 2019 года появилась на торрент-треккерах., а значит нет смысла терпеть экранку — можно посмотреть фильм Джокер с удовольствием.

Джокер (фильм 2019) смотреть онлайн в хорошем качестве бесплатно
film-dj.online/news/dzhoker_film...
Джокер (фильм 2019). ПОДЕЛИСЬ С ДРУЗЬЯМИ, ЕСЛИ ПОНРАВИЛСЯ ФИЛЬМ! Фильм просто шикарный, очень эмоциональный.

■ **Видео по запросу «джокер фильм 20...** 8 тыс. результатов

Джокер (2019) - Безымянный youtube 12 октября 2019
Джокер — Русский трейлер (2019) youtube 3 апреля 2019
Джокер 2019 фильм TS mail.ru 26 ноября 2019
Джокер войны youtube 15 декабря 2019

• **Джокер (2019) — КиноПоиск**
kinopoisk.ru › Джокер
Джокер (2019) — Joker. Всё о фильме: дата выхода, трейлеры, фото, актеры. Отзывы зрителей и профессиональные рецензии. Рейтинг. Общие сборы и бюджет фильма. Интересные факты и ошибки в фильме.
★★★★★ 8 из 10 | 145 тыс. голосов

w **Джокер (фильм, 2019) — Википедия**
ru.wikipedia.org/wiki/Джокер_...
«Джокер» — американский психологический триллер режиссёра Тодда Филлипса по сценарию, написанному Филлипсом совместно со Скоттом Сильвером. Главную роль исполняет Хоакин Феникс. Сюжет фильма представляет собой версию истории происхождения суперзлодея Джокера, одного из ключевых антагонистов Бэтмена в комиксах компании DC Comics. Википедия

Джокер
Joker, США, Канада, 2019, 18+

★ 8,3 Кино Mail.Ru ★ 8,3 КиноПоиск ★ 8,8 IMDb
«Джокер» — американский психологический триллер режиссёра Тодда Филлипса по сценарию, написанному Филлипсом совместно со Скоттом Сильвером. Главную роль исполняет Хоакин Феникс. Сюжет фильма представляет собой версию истории происхождения суперзлодея Джокера, одного из ключевых антагонистов Бэтмена в комиксах компании DC Comics. Википедия
Жанр: драма, триллер, криминал
Режиссёр: Todd Phillips
Продюсеры: Bradley Cooper, Todd Phillips, Michael E. Uslan, Bruce Berman, Aaron L. Gilbert, +
Длительность: 2 ч. 3 мин.
Актеры
Роберт Де Ниро Хоакин Феникс Брайан Тайри Генри Зази Битц Фрэнсис Конрой Ещё

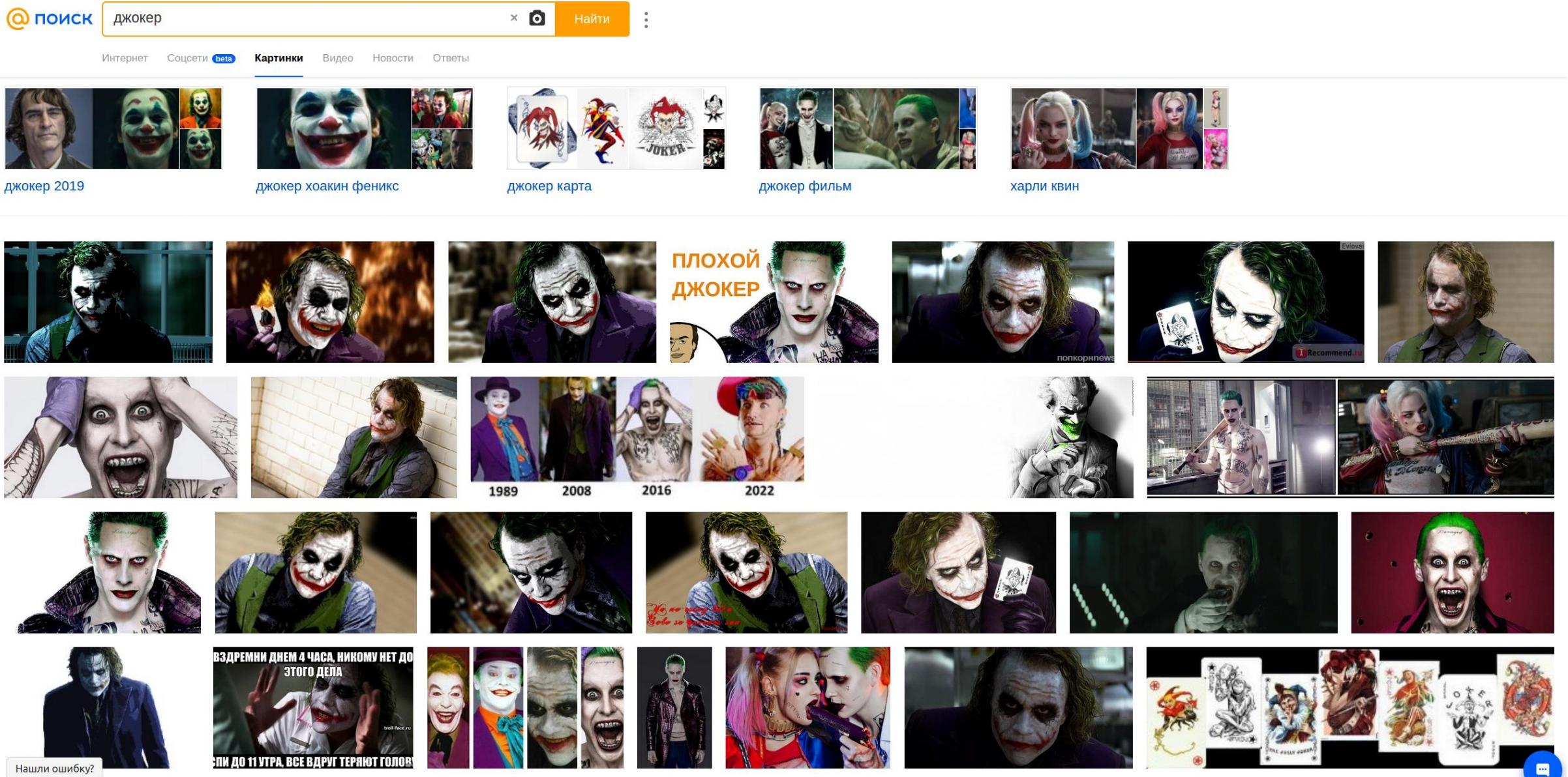
Смотрите также
 Ещё

Поиск по картинкам

поиск джокер

Интернет Соцсети **Картинки** Видео Новости Ответы

джокер 2019 джокер хоакин феникс джокер карта джокер фильм харли квин



Нашли ошибку?

ВЗДРЖНИ ДНЕМ 4 ЧАСА, НИКОМУ НЕТ ДО ЭТОГО ДЕЛА.
СПИ ДО 11 УТРА, ВСЕ ВДРУГ ТЕРЯЮТ ГОЛОВЫ!

7

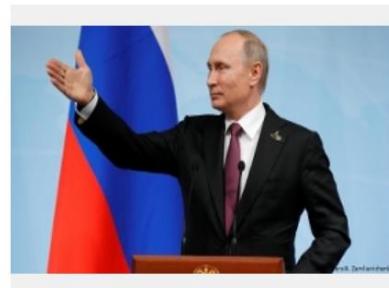
Поиск по картинке



@ ПОИСК

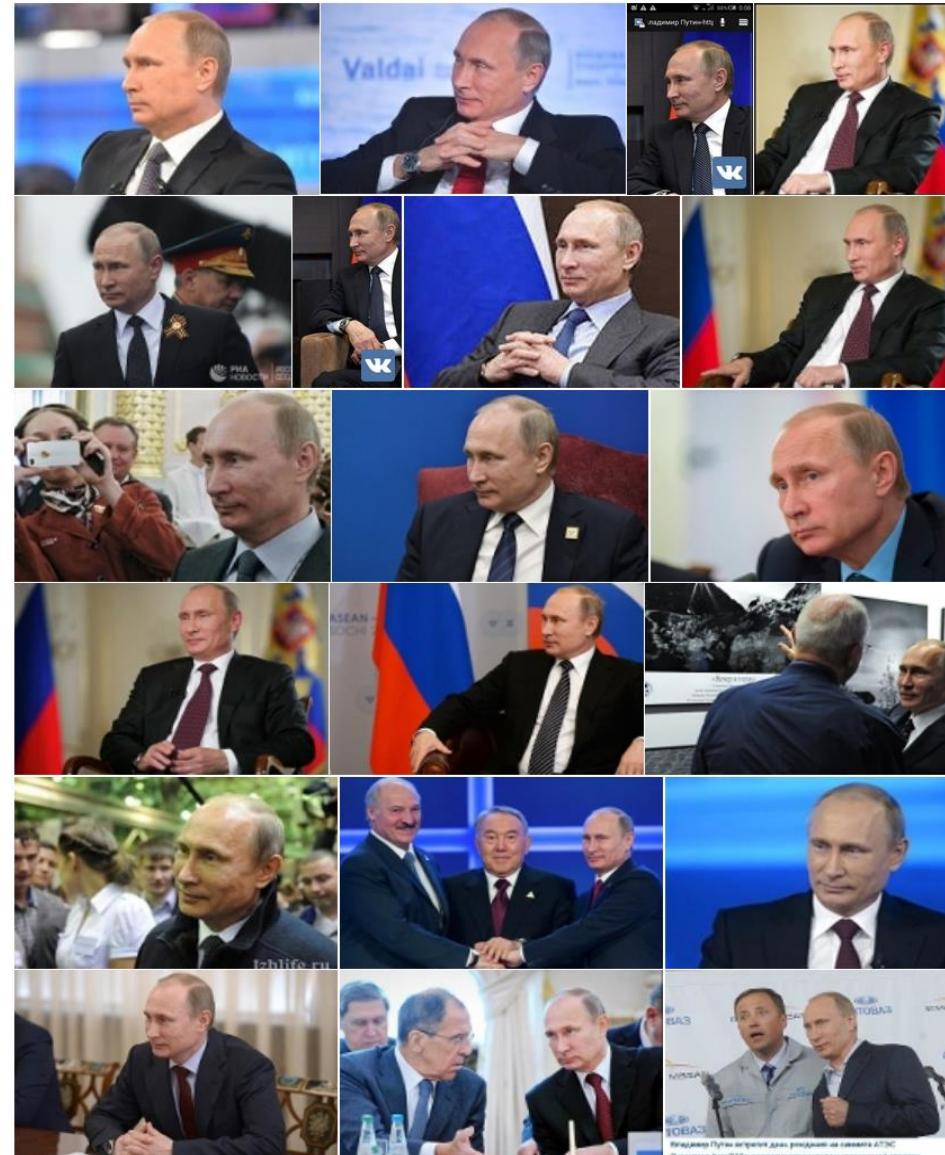
Загруженное изображение ×

Найти



Исходное изображение 700x394

Интернет Соцсети beta Картинки Видео Новости Ответы



Поиск по видео

@ mail.ru
group

Поиск по видео

Интернет Соцсети beta Картинки Видео Новости Ответы Фильтры

Игра престолов 7 сезон

Watch later Share

Игра престолов 7 сезон 3 серия

youtube 30.07.17 Игра престолов 7 сезон 3 серия

Также смотрят

Игра престолов

Сезон 1 2 3 4 5 6 7 8

1 2 3 4 5 6 7 8

▶ 59:35 Игра престолов 7 сезон 3 серия youtube 30.07.17

▶ 01:00:00 Игра престолов сезон 7 серия 1 mail.ru 17.10.19

▶ 55:40 ИГРА ПРЕСТОЛОВ 7 СЕЗОН 2 СЕРИЯ (ОЗВУЧКА LOSTFILM) ok.ru 25.07.17

▶ 54:32 Игра престолов 7 сезон 1 серия straxhelp.ru

▶ 50:20 ИГРА ПРЕСТОЛОВ 7 СЕЗОН 1 серия 2017г Серия в HD chros.tech

▶ 58:56 Игра Престолов 7 сезон 1 серия / Game of Thrones (2017) HD1080p | Amedia gelomarted.ru

Поиск по социальным сетям

Поиск по социальным сетям

Причины для поиска:

- Интернет
- Соцсети beta
- Картинки
- Видео
- Новости
- Ответы

Результаты поиска:

Onlinetours: горячие туры и акции
@onlinetoursru

Любите лыжи и сноуборды?
Самое время готовить инвентарь! Подборка супертуров на ...
2 часа назад

Комбинированный тур на #Мальдивы с захватом Нового года Мечты должны сбываться! 2 ночи в Three ...
4 часа назад

Хотите Барселону? Туристический сезон в ...
6 часов назад

Горящие туры
@vktrip

Италия, Римини, Rainbow 3*. Вылет: 18.01, 8 дней, завтраки. Стоимость: от 11 600 рублей С ...
44 минуты назад

Доминикана, Бока-Чика, Отели 4*. Вылет: 25.01, на 8 дней, все включено. Стоимость: от ...
1 час назад

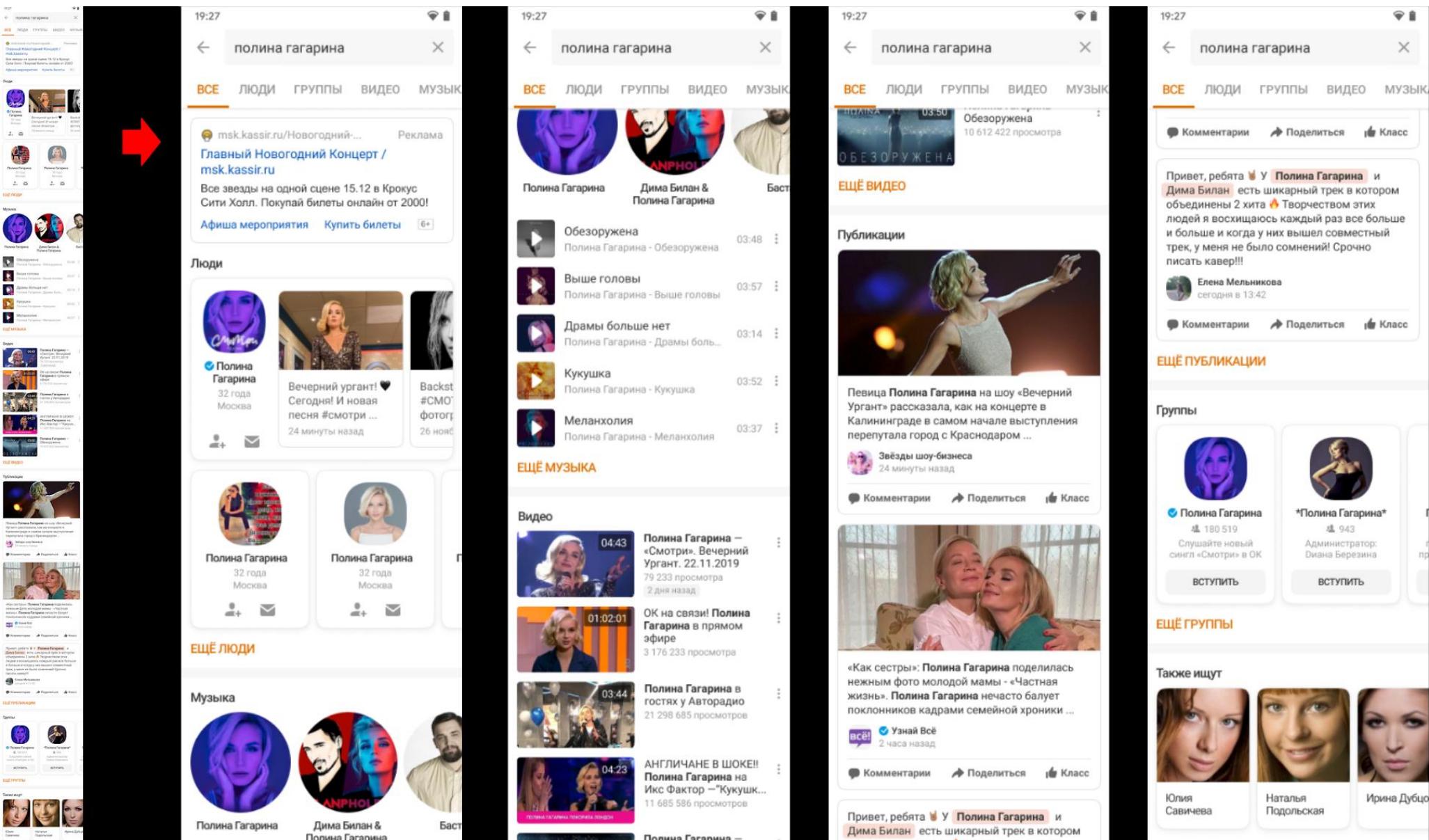
Горячие туры и горячие путевки из ...
@club.hot.tours

ВНИМАНИЕ! ТУРЦИЯ НА НОВЫЙ ГОД СУПЕР ЦЕНА! Отели 4-5 звезд 30.12. и 31.12.2019 из Москвы на 7 ночей ...
3 часа назад

ВНИМАНИЕ! ИНДИЯ НА НОВЫЙ ГОД СУПЕР ЦЕНА! Отели 2-3 звезд Прямой перелет!!! 24.12.2019 из Москвы на ...
Вчера



Поиск в социальных сетях



Поиск в e-сом приложениях

[Регистрация](#)[Войти](#)[Все категории](#)[Товары за баллы](#)[Быстрая доставка](#)[Женская одежда](#)[Дом и интерьер](#)[Автотовары](#)[Электроника](#)

Главная / Поиск

Результаты поиска: iphone

[Телефоны и планшеты](#)[Электроника](#)[Автотовары](#)[Цвет](#)[Цена](#)

Заказов: 696 ★★★★★
Apple iPhone XR, смартфон (китайская версия)
47271 ₽



Заказов: 625 ★★★★★
Apple iPhone XS, смартфон
57119 ₽



Заказов: 965 ★★★★★
Apple iPhone 6S, восстановленный смартфон
9651 ₽



Заказов: 844 ★★★★★
Apple iPhone 7 32 Гб / 128 Гб / 256 Гб, восстановленный смартфон
15232 ₽



Заказов: 23 ★★★★★
Apple iPhone 6S Plus, восстановленный смартфон
16846 ₽



Заказов: 66 ★★★★★
Apple iPhone 11, китайская версия, смартфон
56463 ₽



Заказов: 239 ★★★★★
Apple iPhone 6 64 Гб, восстановленный смартфон
15695 ₽



Заказов: 167 ★★★★★
Apple iPhone 6 64 Гб, восстановленный смартфон
12327 ₽



Заказов: 202 ★★★★★
Apple iPhone 7 Plus 32 Гб / 128 Гб / 256 Гб...
24292 ₽



Заказов: 78 ★★★★★
Apple iPhone 5S 16 Гб / 32 Гб / 64 Гб, смартфон...
9652 ₽



Заказов: 1 ★★★★★
iPhone XR, восстановленный смартфон
51720 ₽

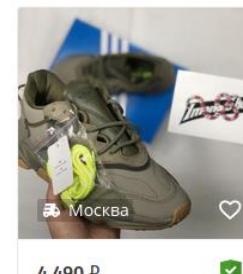


Заказов: 1 ★★★★★
iPhone 8 2/64/256 Гб, восстановленный смартфон
52131 ₽

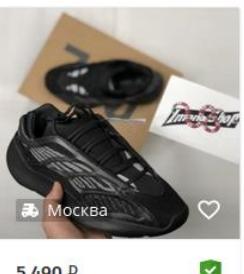
[Категории](#)[Найти](#)

Главная / Мужской гардероб / Обувь в Москве

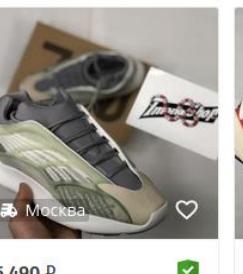
Обувь

[По умолчанию](#)

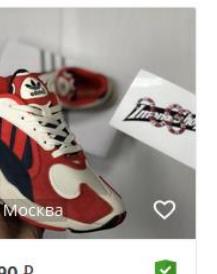
4 490 ₽
Новые Кроссовки Adid...



5 490 ₽
Новые Кроссовки Adid...



5 490 ₽
Новые Кроссовки Adid...



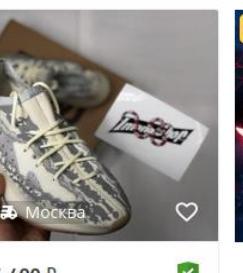
3 990 ₽
Новые Кроссовки Adid...



4 490 ₽
Новые Кроссовки Adid...



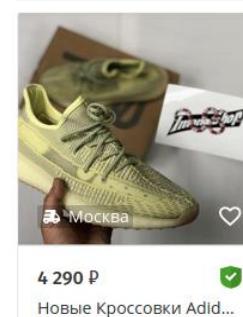
5 100 ₽
Кроссовки Adidas Temp...



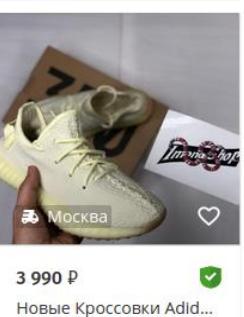
4 490 ₽
Новые Кроссовки Adid...



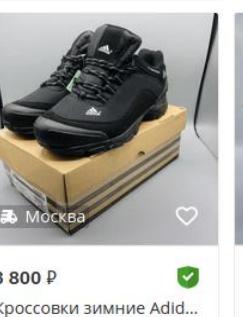
РЕКЛАМА
«Звёздные Войны:
Восход» (16+)



4 290 ₽
Новые Кроссовки Adid...



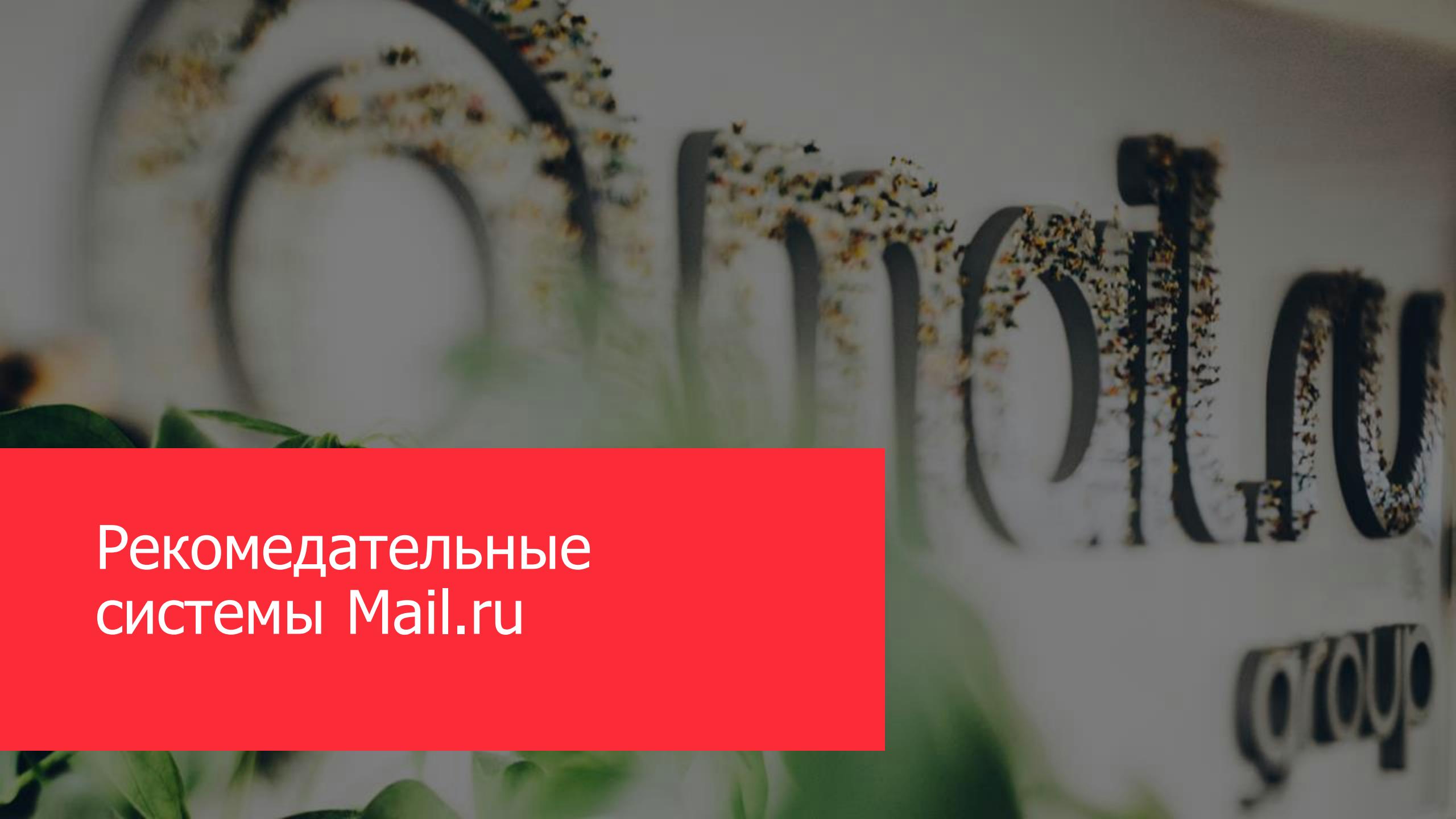
3 990 ₽
Новые Кроссовки Adid...



3 800 ₽
Кроссовки зимние Adid...



5 099 ₽
Осенние Кроссовки Adi...

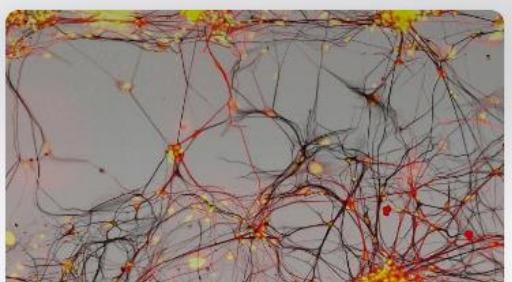


Рекомендательные
системы Mail.ru

Продукты

Создаем персонализированные продукты

- Лента рекомендаций - Пульс
- Персонализированное телевидение
- Рекомендации фильмов
- Рекомендации музыкальных треков
- Ленты товарных рекомендаций
- ...



Трансплантация нейронов при болезни Паркинсона...

Как известно, симптомы...

VK ЭТО РАБОТАЕТ



Правительство РФ разработало законопроект о банковских...

Принимать банковские...

INC-NEWS.RU



Сооснователь Nginx назвал действия Rambler...

Неправомерные действия...

RBC.RU



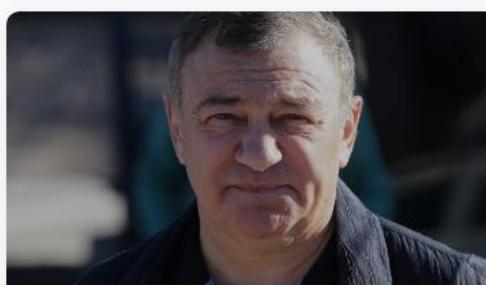
Выходец из семьи ирландского эмигранта и швейцарской эмигрантки Фрэнк Ширан (Роберт Де Ниро...)

EXLER.RU



Вкусная пицца прямо из печи. Привезем за 30 минут или пицца...

Реклама



Конец эпохи: как изменилась бизнес-империя Ротенбергов к концу...

FORBES.RU



В «Газпроме» рассказали, как выручат Сербию при



Водитель подорвал свой...

Взрыв, повредивший автомобиль и



Как меняется организм...

Обрезание крайней плоти мужского

Пульс

[Сериалы](#) [Шоу](#) [Фильмы](#) [Музыка](#) [Новости](#) [Мультфильмы](#)

рекомендованное

Подборка видео на основе ваших вкусов



далее

Идёт

Музыка

Gorillaz - Feel Good Inc.



Noize MC - Вселенная бесконечна?



Кипелов - Я свободен (Official video)



ещё

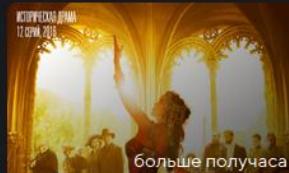
21:37 Шоу

60 минут. Зеленский высказался по ситуации в



Канал Mail.ru

Вы смотрели



Burito - Пока город спит

Звонкий, Рем Дигга - Из окон

Three Days Grace - So What

Мата Хари. Сезон 1. Серия 1

Метод Фрейда. Сезон 1.

Imagi
It Take

E-commerce



***** MTS RUS 16:30

Смартфоны на Android
Покупай по безопасной сделке

Юла
Прод.

2,9 км

4 900 ₽

Куртка зимняя му...

3,2 км

900 ₽

Книга Гарри поте...

2,5 км

25 000 ₽

Mercedes Brabus

39,1 км

6 850 000 ₽

Квартира, 3 комнаты,...

***** MTS RUS 16:28

PANDAO

Поиск товаров в Китае

Лучшее Товары за баллы Красота и здс

-50%

Наушники с
экранированным пр...

30 205 заказов ★★★★★

Подставкой для
мобильных телефон...

30 146 заказов ★★★★★

49 ₽ 16 ₽

-51%

-70%

Q

Heart

Bag

Folder

User

***** MTS RUS 16:29

Кооперативная улица, Зк5

Сортировка и фильтры

IL Патио

★★★★★
Заказ от 300 ₽ • Итальянская
Delivery Express

45 мин

45 мин

BB&Burgers

★★★★★

DC Hot

IR Tasks



Поиск

Яндекс

котики



Найти

Поиск Картинки Видео Карты Маркет Новости Переводчик Музыка Еще

Котики — ВКонтакте

[vk.com/v.kote](#)

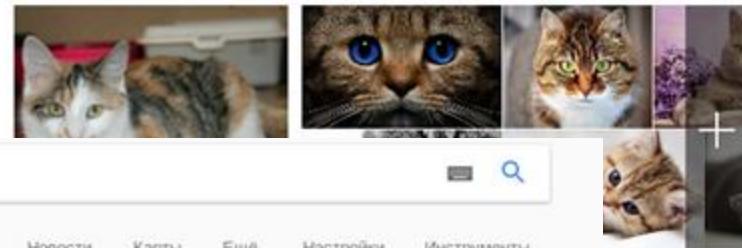
Дата рождения: 19 янв 2011

О себе: Комиксы: #котмиксы@v.kote Анимашки: #котогифки@v.k...

[Читать ещё >](#)



Кошка



Google

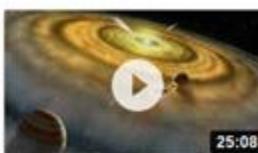
космос вселенная



Все Картинки Видео Новости Карты Еще Настройки Инструменты

Результатов: примерно 5 580 000 (0,42 сек.)

Видео



Мстители
The Avengers, США, 2012, 12+

★ Кино Mail.Ru 7.9 ★ КиноПоиск 7.9 ★ IMDB 8.1

Фильм Marvel «Мстители» является продолжением эпических борьб, начавшихся в фильмах «Железный человек», «Невероятный Железный человек 2», «Тор» и «Первый мститель». В прокат супергерои объединяются в уникальную команду. Когда члены

[Читать дальше >](#)

Жанр: фэнтези, боевик, приключения, фантастика

Режиссёр: Джосс Уидон

Продюсеры: Стэн Ли, Джон Фавро, Кевин Файги, Аллан Файн, Луис Д'Эспозито, +

Длительность: 2 ч. 22 мин.

Актёры

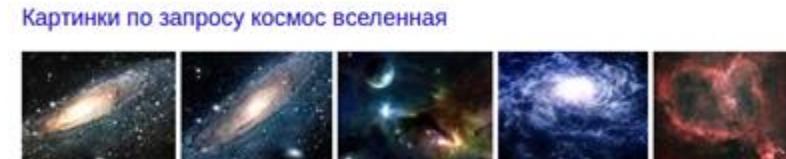


Роберт Дауни мл.

Крис Эванс

Марк Руффало

Крис Хемсворт



→ Другие картинки по запросу "космос вселенная"

Пожаловаться на картинки

Почему космос такой темный, если во Вселенной миллиарды ...

<https://hi-news.ru/Tемы/Космос>

1 июн. 2016 г. - Если Вселенная безгранична и полна звезд и галактик, почему мы не видим их повсюду, куда бы ни посмотрели? Иногда самые ...

мстители

- Мстители (2012) смотреть онлайн бесплатно
kinogo.net/110-mstiteliess-2012...

«Мстители» – фантастический боевик о новых приключениях супергероев. Сюжет картины посвящен извечному противостоянию добра и зла.

- Мстители 3: Война бесконечности смотреть онлайн фильм бесплатно в...

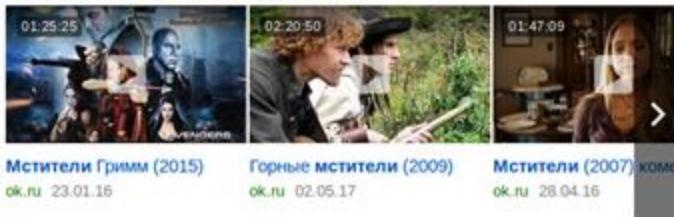
kino-1080hd.ru › Боевик
Пока правительство Соединенных Штатов Америки пытается восстановить разрушения после вторжения пришельцев, а Мстители устраивают разборки из-за...

- Смотреть Первый мститель: Противостояние в хорошем качестве hd 720

kinohabr.net › Фантастика

Все это заканчивается тем, что между членами Мстителей происходит серьезное противостояние, имеющее массу проблемных последствий в будущем.

Больше видео по запросу: Мстители



- Мстители (2012) смотреть онлайн бесплатно в хорошем hd 720 качестве

kinogid.me › Фильмы 2012

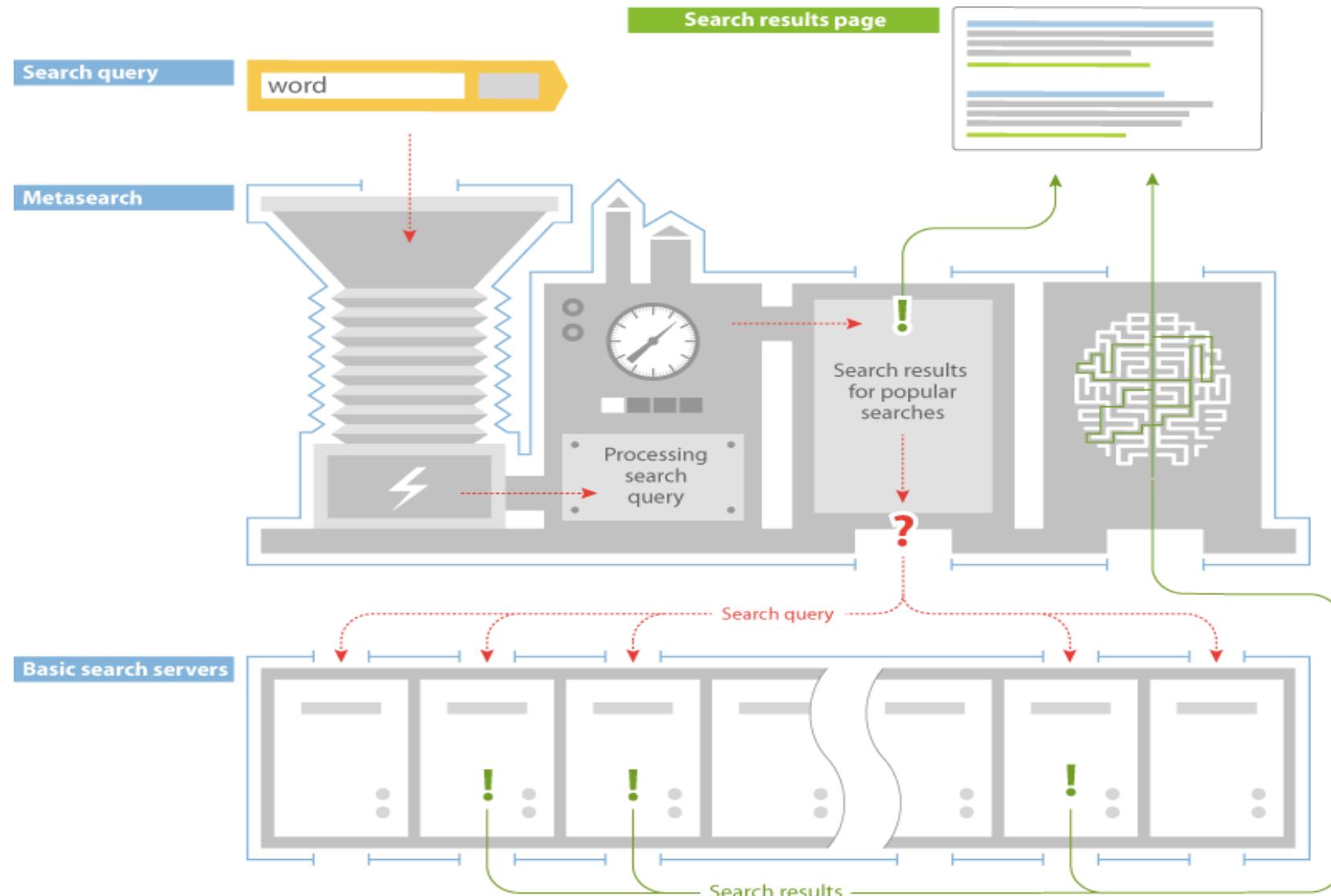
Жанр: Боевики, Приключения, Фантастика, Фэнтези. Мстители 2: Эра Альтрана. Безумный Макс 3: Дорога ярості. Тор 2: Царство тьмы

- Мстители — трейлеры, даты премьер — КиноПоиск

kinopoisk.ru/film/11175

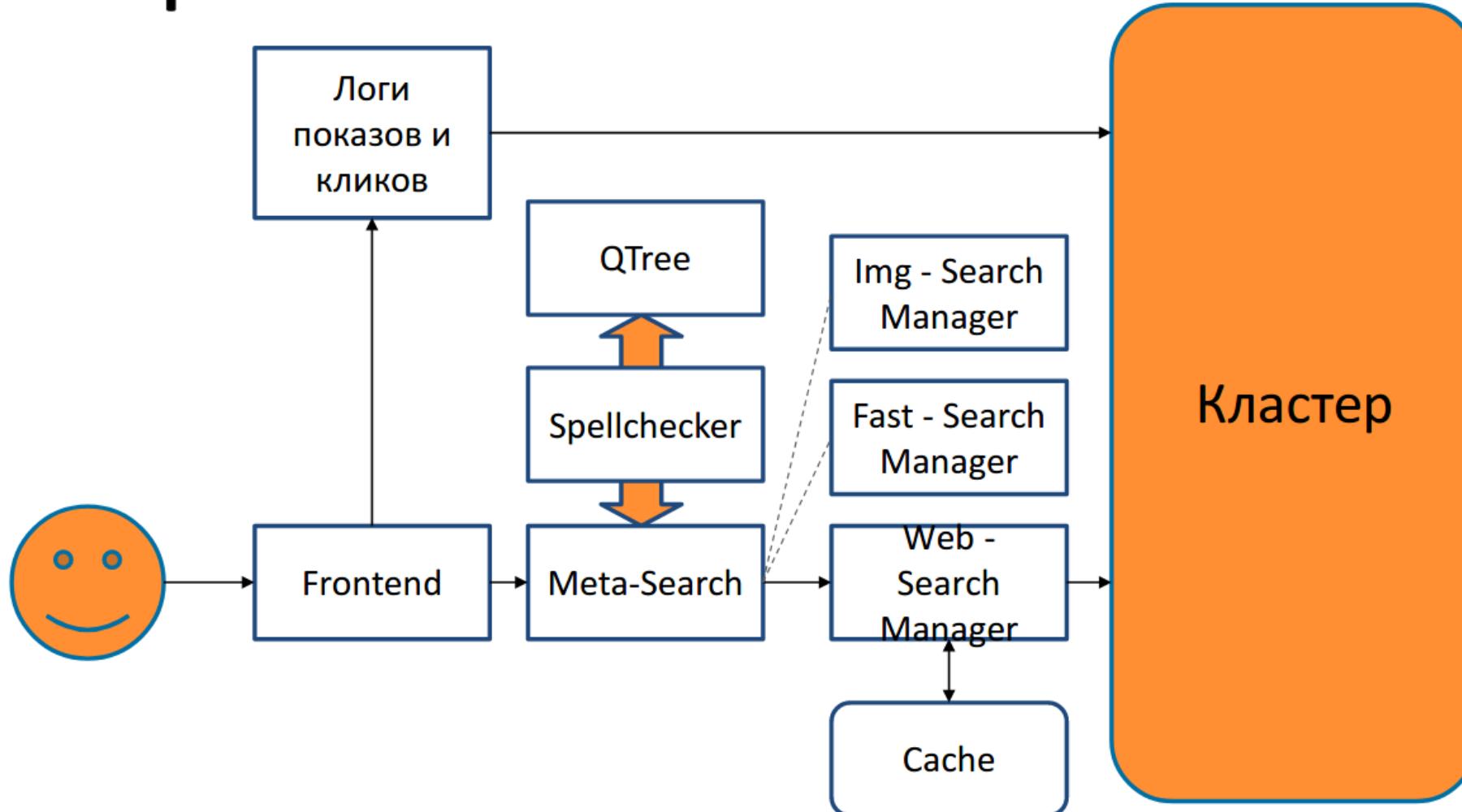
Для того чтобы добавить рецензию на фильм Мстители, необходимо: Звёздные войны

Архитектура

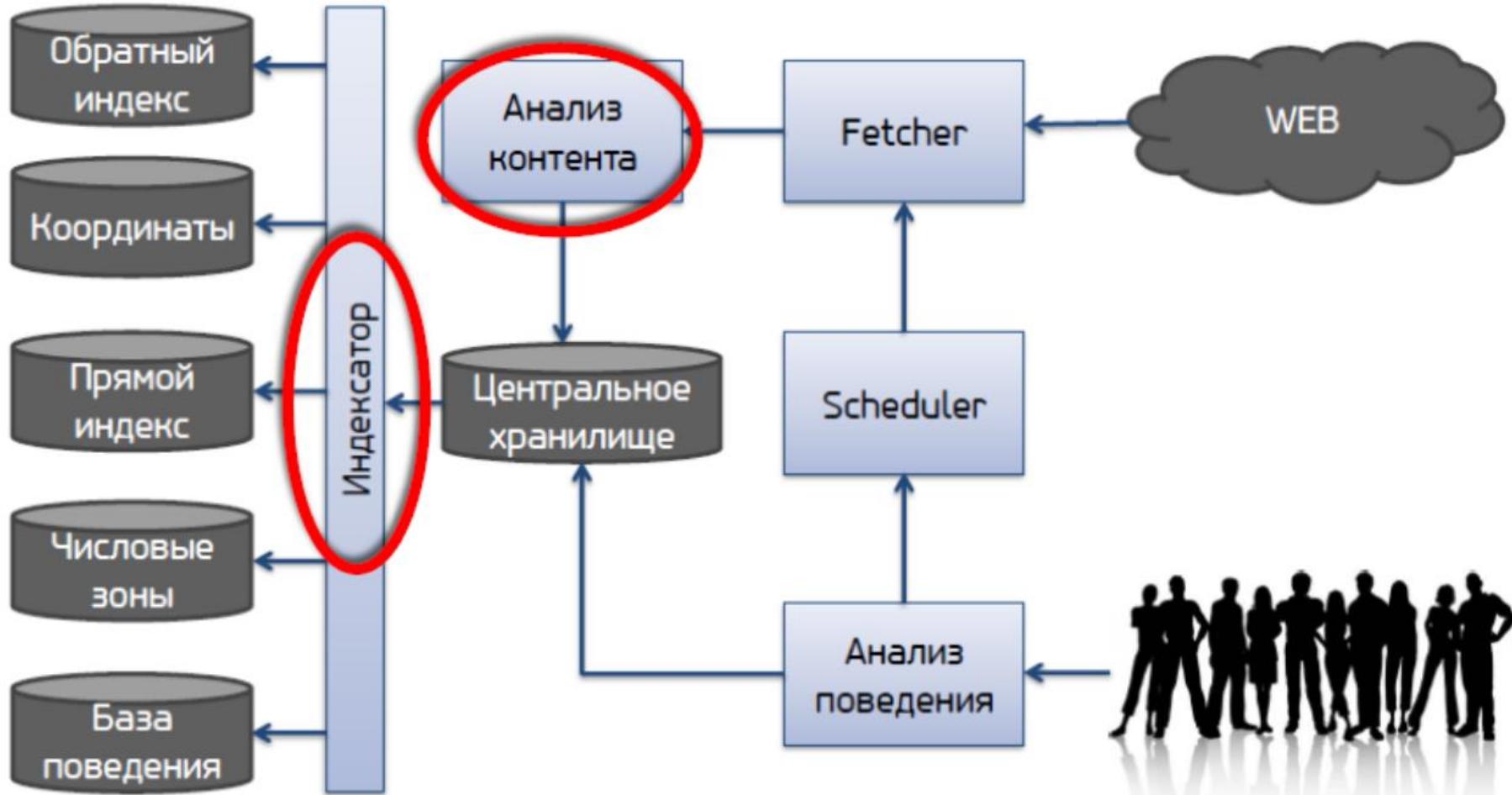


Упрощенная схема

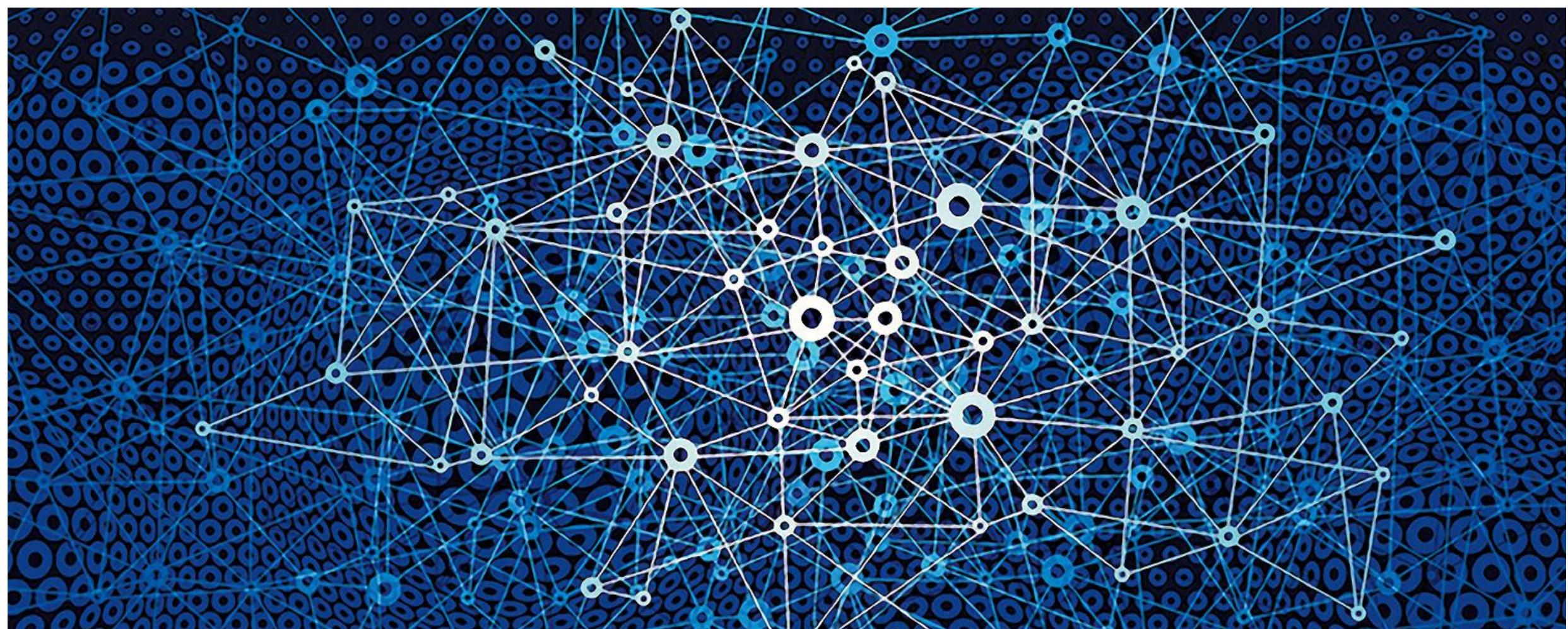
Как работает поиск



Оффлайн кусок



Фокусировка краулера



Фильтрация спама



Фильтрация порнографии



Исправление опечаток



поиск адронный коллайдер Найти

Интернет Соцсети beta Картинки Видео Новости Ответы

Исправлена опечатка: андронный коллайдер Отменить

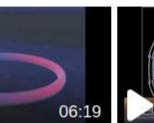
w **Большой адронный коллайдер — Википедия**
ru.wikipedia.org/wiki/Большой_...
Большой адронный коллайдер, сокращённо БАК — ускоритель заряженных частиц на встречных пучках, предназначенный для разгона протонов и тяжёлых ионов (ионов свинца) и изучения продуктов их соударений.
Коллайдер построен в ЦЕРНе (Европейский совет ядерных исследований), находящемся около Женевы...
Поставленные задачи Распределённые вычисления
Технические характеристики Научные результаты
Строительство и эксплуатация Финансирование проекта

Картинки по запросу «андронный кол...» 12 тыс. результатов



fb **Адронный коллайдер** зачем нужен? Для чего нужен большой адронный...
fb.ru/article/189775/adronnyiy-...
Открывшиеся возможности человека, управляющего гравитацией, многократно превышают все видимые перспективы развития технического прогресса. Тем более что передовых учёных больше интересует не само...

Вideo по запросу «андронный коллайдер» 9 тыс. результатов



Большой адронный коллайдер National youtube 47:30
5 января 2014

Как работает Большой Адронный youtube 06:19
14 сентября 2008

За гранью. Коллайдер Большой youtube 26:04 HD
20 августа 2017

колла youtube 7 февр

Саджесты



пресс

x

Найти

пресс конференция путина

прессбол

пресс конференция путина 2019

свободная пресса

престанс

престириум

преступление и наказание

игра престолов

преступная россия

игра престолов онлайн

Ранжирование

Этапы ранжирования



Построение сниппетов



@ ПОИСК сколько миллиграмм в грамме Найти ::

Интернет Соцсети beta Картинки Видео Новости Ответы

В 1 г — 1000 мг
[Сколько в грамме миллиграмм?](#)
skolkoru.ru/skolko-milligramm-v-odnom-gramme/

FB Сколько в грамме миллиграмм - простой ответ

fb.ru/article/34584/skolko-v-...

Заканчивая обучение, мы часто забываем многое, что проходили по программе. Например, не все помнят, **сколько в грамме миллиграмм**. Однако эти знания бывают порой просто необходимы и в повседневной...

поиск дмитрий медведев Найти

Интернет Соцсети **beta** Картинки Видео Новости Ответы

Медведев, Дмитрий Анатольевич — Википедия
ru.wikipedia.org/wiki/Медведев,_Дмитрий_Анатольевич

Дмитрий Анатольевич Медведев — российский государственный и политический деятель. Председатель Правительства Российской Федерации с 8 мая 2012 года. Председатель партии «Единая Россия» с 26 мая 2012 года. Президент Российской Федерации с 7 мая 2008 по 7 мая 2012. В 2000—2001, 2002—2008 гг.

Биография Председатель Правительства...
Участие в выборах... Бизнес
Президентство В сфере информационных...

Картинки по запросу «дмитрий медве...» 12 тыс. результатов

Председатель Правительства России Д.А.Медведев
da-medvedev.ru

Встреча Дмитрия Медведева с руководством украинской политической партии «Оппозиционная платформа – За жизнь»
 Видеоблог Дмитрия Медведева Тема: Здравоохранение Кстати, о часовых поясах

Дмитрий Медведев (@damedvedev) • Фото и видео в Instagram
instagram.com/damedvedev

Дмитрий Медведев 2,7 млн подписчиков 1 подписка

Видео по запросу «дмитрий медве...» 11 тыс. результатов

Почему преподаватели мало зарабатывают? 03:36
[youtube](https://www.youtube.com/watch?v=3JyfjwvXWUo)
 3 августа 2016

Дмитрий Медведев на закрытии осенней youtube 09:23
[youtube](https://www.youtube.com/watch?v=3JyfjwvXWUo)
 2 дня назад

Дмитрий Медведев подводит итоги работы mail.ru 01:29:35 FHD
[mail.ru](https://www.youtube.com/watch?v=3JyfjwvXWUo)
 8 декабря 2018

Дмитрий Медведев решает youtube 09:23
[youtube](https://www.youtube.com/watch?v=3JyfjwvXWUo)
 9 дека...

Дмитрий Медведев
 Российский государственный и политический деятель

Дмитрий Анатольевич Медведев — российский государственный и политический деятель. Председатель Правительства Российской Федерации с 8 мая 2012 года. Председатель партии «Единая Россия» с 26 мая 2012 года. Википедия

Родился: 14 сентября 1965 г. (54 года), Санкт-Петербург, СССР
 Рост: 163 см
 В браке с: Светлана Медведева (с 1993 г.)

Фильмы и сериалы

Он вам не Димон 2017
 Ёлки 2010
 Путин, Россия и Запад 2011
 Уроки русского 2010
 Показательный процесс: Ист... 2013

Смотрите также

Владимир Путин
 Светлана Медведева
 Владимир Жириновский
 Борис Ельцин
 Виктор Черномырдин

Blender

А ещё...

- Определение языка
- Выделение именованных сущностей
- Поиск похожих запросов
- Тематическое моделирование
- Поиск дубликатов
- Active Learning
- Semisupervised Learning
- Reinforcement Learning
- ...



Search quality
estimation



Что такое хороший поиск?



Релевантность - сложное комплексное понятие, учитывающее множество факторов и зачастую крайне субъективное



Что такое хороший поиск?



В информационном поиске релевантность рассматривается с нескольких сторон

- тематическая релевантность
- пользовательская релевантность
- текстовая релевантность
- ...

Классическое ранжирование



Как оценивать качество ранжирования?

- Возьмем запросы, которые задают пользователи
- Для каждого запроса рассмотрим документы, которые демонстрирует эта поисковая система
- Про каждый документ скажем, насколько он хорош

@ поиск конституция РФ Найти

Интернет Соцсети beta Картинки Видео Новости Ответы

w Конституция Российской Федерации — Википедия ru.wikipedia.org/wiki/... Конституция Российской Федерации — высший нормативный правовой акт Российской Федерации. Принята народом России 12 декабря 1993 года, вступила в силу 25 декабря 1993 года. Конституция обладает высшей юридической силой, закрепляющей основы конституционного строя России, государственное устройство...

RG Конституция Российской Федерации от 25 декабря 1993 года, С... rg.ru/2009/01/21/konstitucia-dok... Конституция Российской Федерации от 25 декабря 1993 года, с изменениями от 30 декабря 2008 года

constitution.kremlin.ru Конституция Российской Федерации Разъяснения отдельных положений Конституции РФ 1993 года (федеральные конституционные законы; федеральные законы и постановления Конституционного Суда...)

vladileitor.ru/constitution Конституция РФ 2019 последняя редакция Конституция РФ 2019 скачать бесплатно Конституцию Российской Федерации все главы и статьи Конституции России последняя редакция

kodeks.systecs.ru/constitution Конституция Российской Федерации: Все главы и статьи Судебная власть и прокуратура. Конституционные поправки и пересмотр Конституции. Заключительные и переходные положения

constrf.ru Конституция РФ с Комментариями. Последняя действующая редакция КРФ... Актуальная редакция Конституции Российской Федерации с изменениями на 2019 год, а также Комментарии ко всем статьям КРФ.

zakonrf.info/konstitucia Конституция Российской Федерации (РФ) Конституция Российской Федерации (РФ) последняя действующая редакция

constitutionrf.ru Конституция РФ 2019 с Комментариями. Последняя редакция КРФ с... Новая редакция Конституции Российской Федерации с последними изменениями и поправками на 2019 год. Комментарий к статьям КРФ.

Классическое ранжирование

Как оценивать качество ранжирования?

- Возьмем запросы, которые задают пользователи
- Для каждого запроса рассмотрим документы, которые демонстрирует эта поисковая система
- Про каждый документ скажем, насколько он хорош

@ поиск конституция РФ Найти

Интернет Соцсети beta Картинки Видео Новости Ответы

w Конституция Российской Федерации — Википедия ru.wikipedia.org/wiki/... Конституция Российской Федерации — высший нормативный правовой акт Российской Федерации. Принята народом России 12 декабря 1993 года, вступила в силу 25 декабря 1993 года. Конституция обладает высшей юридической силой, закрепляющей основы конституционного строя России, государственное устройство...

RG Конституция Российской Федерации от 25 декабря 1993 года, С... rg.ru/2009/01/21/konstitucia-dok... Конституция Российской Федерации от 25 декабря 1993 года, с изменениями от 30 декабря 2008 года

constitution.kremlin.ru Конституция Российской Федерации constitution.kremlin.ru Разъяснения отдельных положений Конституции РФ 1993 года (федеральные конституционные законы; федеральные законы и постановления Конституционного Суда...)

vladrevisor.ru/constitution Конституция РФ 2019 последняя редакция vladrevisor.ru/constitution Конституция РФ 2019 скачать бесплатно Конституцию Российской Федерации все главы и статьи Конституции России последняя редакция

kodeks.systecs.ru/constitution Конституция Российской Федерации: Все главы и статьи kodeks.systecs.ru/constitution Судебная власть и прокуратура. Конституционные поправки и пересмотр Конституции. Заключительные и переходные положения

constrf.ru Конституция РФ с Комментариями. Последняя действующая редакция KRF... constrf.ru Актуальная редакция Конституции Российской Федерации с изменениями на 2019 год, а также Комментарии ко всем статьям КРФ.

zakonrf.info/konstitucia Конституция Российской Федерации (РФ) zakonrf.info/konstitucia Конституция Российской Федерации (РФ) последняя действующая редакция

constitutionrf.ru Конституция РФ 2019 с Комментариями. Последняя редакция KRF с... constitutionrf.ru Новая редакция Конституции Российской Федерации с последними изменениями и поправками на 2019 год. Комментарий к статьям КРФ.

Классическое ранжирование



Как оценивать качество ранжирования?

- Придумаем из этого метрику
- Вариант: доля релевантных документов среди показанных
- Проблема: не учитывается позиция релевантных
- Решение: метрика DCG

Поиск: конституция РФ

Интернет Соцсети beta Картинки Видео Новости Ответы

W Конституция Российской Федерации — Википедия ru.wikipedia.org/wiki/... Конституция Российской Федерации — высший нормативный правовой акт Российской Федерации. Принята народом России 12 декабря 1993 года, вступила в силу 25 декабря 1993 года. Конституция обладает высшей юридической силой, закрепляющей основы конституционного строя России, государственное устройство...

RG Конституция Российской Федерации от 25 декабря 1993 года, c... rg.ru/2009/01/21/konstitucia-dok... Конституция Российской Федерации от 25 декабря 1993 года, с изменениями от 30 декабря 2008 года

RU Конституция Российской Федерации constitution.kremlin.ru Разъяснения отдельных положений Конституции РФ 1993 года (федеральные конституционные законы; федеральные законы и постановления Конституционного Суда...)

RU Конституция РФ 2019 последняя редакция vladileitor.ru/constitution Конституция РФ 2019 скачать бесплатно Конституцию Российской Федерации все главы и статьи Конституции России последняя редакция

RU Конституция Российской Федерации: Все главы и статьи kodeks.systecs.ru/constitution Судебная власть и прокуратура. Конституционные поправки и пересмотр Конституции. Заключительные и переходные положения

RU Конституция РФ с Комментариями. Последняя действующая редакция KRF... constrf.ru Актуальная редакция Конституции Российской Федерации с изменениями на 2019 год, а также Комментарии ко всем статьям КРФ...

Z Конституция Российской Федерации (РФ) zakonrf.info/konstitucia Конституция Российской Федерации (РФ) последняя действующая редакция

RU Конституция РФ 2019 с Комментариями. Последняя редакция KRF с... constitutionrf.ru Новая редакция Конституции Российской Федерации с последними изменениями и поправками на 2019 год. Комментарий к статьям КРФ...

Классическое ранжирование

DCG

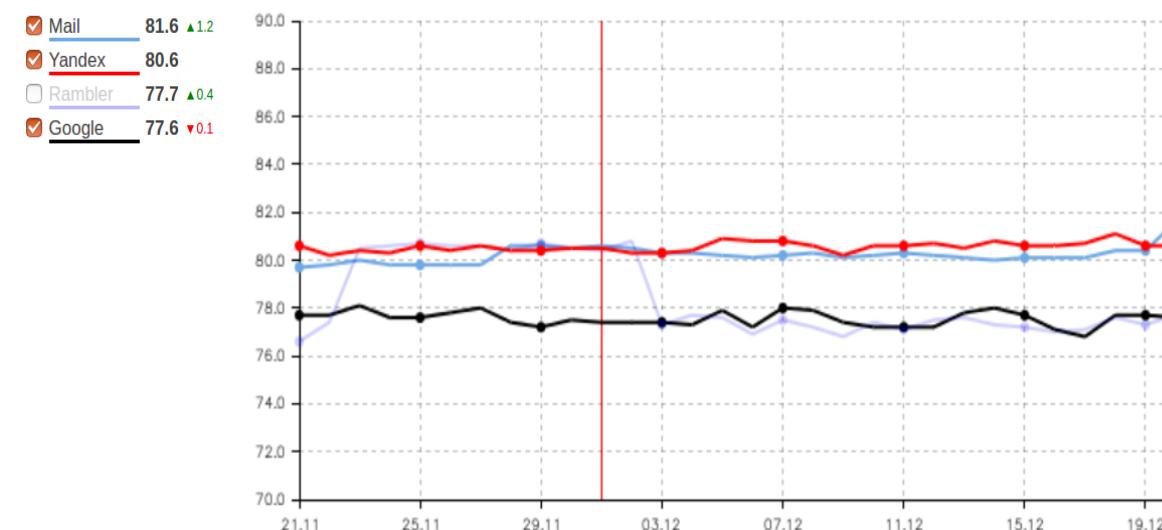
- › Предположим, что релевантности – это действительные числа из диапазона [0,1]
- › Пусть документы расположены в некотором порядке: d_1, d_2, \dots, d_n
- › Пусть им соответствуют релевантности r_1, r_2, \dots, r_n
- › Хотим, чтобы ответ считался хорошим, если более релевантные документы стоят выше
- › DCG – discounted cumulative gain

$$DCG = \sum_{i=1}^n \frac{r_i}{\ln(i+1)}$$

Оценка качества поиска



Ассесорские оценки



Общее качество

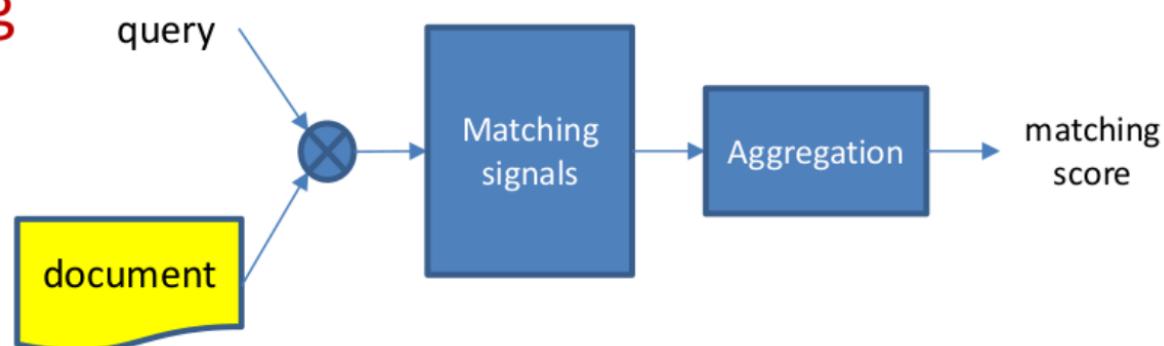
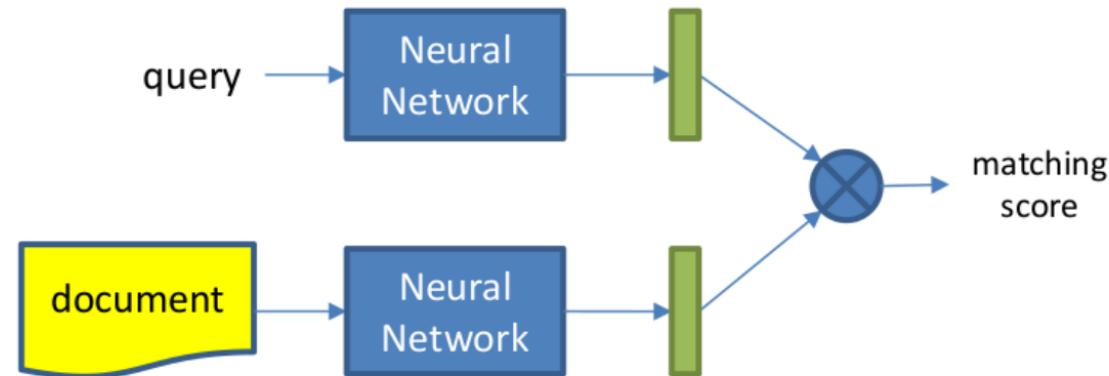


A black and white photograph showing a close-up view of a server or networking equipment. The foreground is dominated by a red rectangular overlay containing the text. Behind the red box, several metal brackets with circular holes are visible, along with numerous small, rectangular components and a row of five numbered slots (1, 2, 3, 4, 5) at the top. The background is blurred, showing more of the server's internal structure.

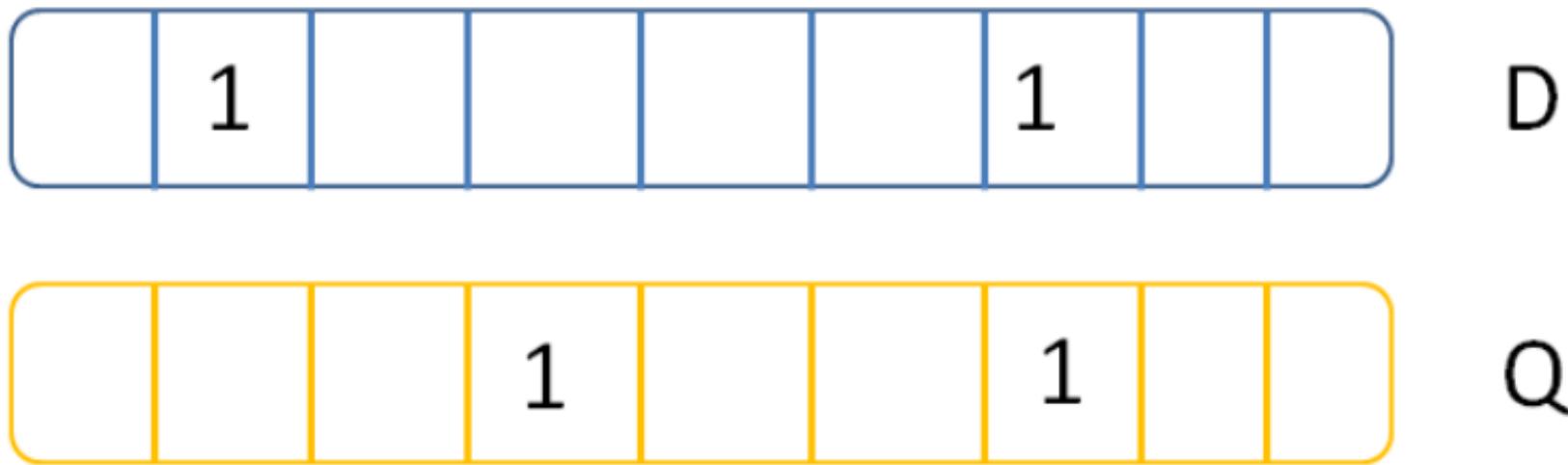
Learning to match

Learning to match

- Representation learning:
representing queries and document in semantic space
- Matching function learning:
discovering and aggregating the query-document matching patterns



Vector space ranking model



TfIdf

$$TfIdf = \sum_{w \in q} tf_w * idf_w$$

BM25

BM25

$$BM25 = \sum_{w \in q} \frac{(k_1 + 1)tf_w}{k_1((1 - b) + b\frac{dl}{avdl}) + tf_w} Idf_w$$

BM25F

$$BM25F = \sum_{w \in q} \frac{\sum_E rank(E)}{\sum_E rank(E) + k} Idf_w$$

Классические модели

Недостатки

- ✖ Соответствие происходит по полному совпадению слов из запроса в документе
- ✖ Не понятно как определить важность слов
- ✖ Не понятно как учитывает порядок слов и контекст, в котором они использованы
- ✖ Приходится думать на лингвистикой (морфология, синонимы, транслитры и т.д)

Векторная модель



$$\text{Sim}(Q, D) = Q^T D = \sum_{w \in q} tfd_w \mathbf{imp}_w$$

Обучаем важность слов на кликах

Векторная модель

deep learning

w Deep learning — Wikipedia

en.wikipedia.org/wiki/Deep_learning

Deep learning (also known as **deep structured learning** or **hierarchical learning**) is part of a broader family of machine learning methods based on **learning** data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep learning models...

в Глубокое обучение — Википедия

ru.wikipedia.org/wiki/Глубокое_...

Глубокое обучение — совокупность методов машинного обучения (с учителем, с частичным привлечением учителя, без учителя, с подкреплением), основанных на обучении представлениям, а не на специализированных алгоритмах, разработанных для конкретных задач. Многие методы глубокого обучения были...

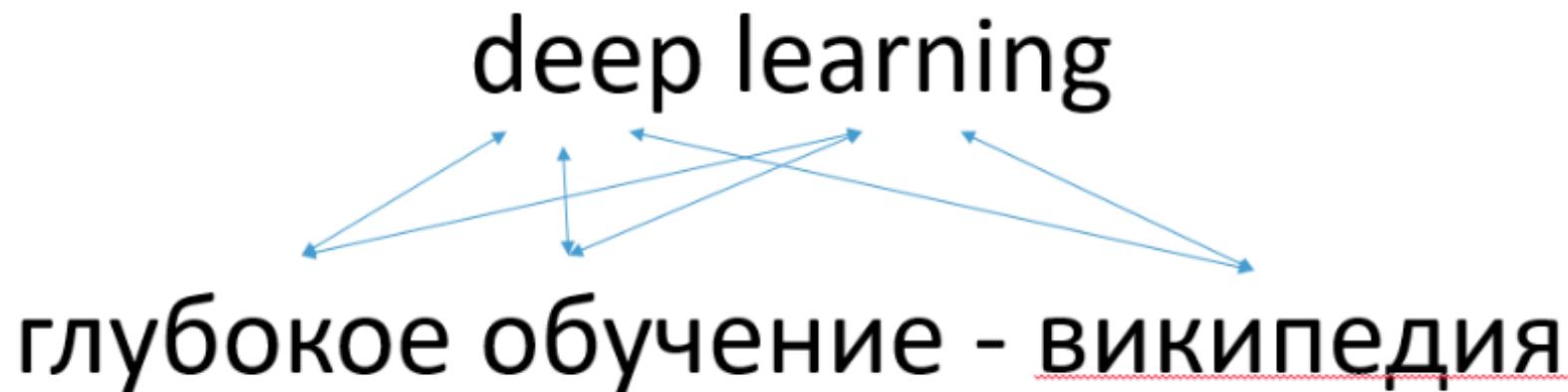
[История](#) [Определения](#) [Содержание](#)

$$\text{Sim}(Q, D) = Q^T D = \sum_{w \in q} tfd_w idf_w$$

Надо подправить модель так, чтобы смочь выучить такие зависимости из данных

Векторная модель

Идея: Будем рассматривать не только пересечение слов запроса и документа, но все пары слов из запроса и документа



$$\text{Sim}(Q, D) = \sum_{q \in Q} \sum_{w \in D} tfd_{qw} idf_{qw}$$

Lexical and Semantic matching

Query: united states president

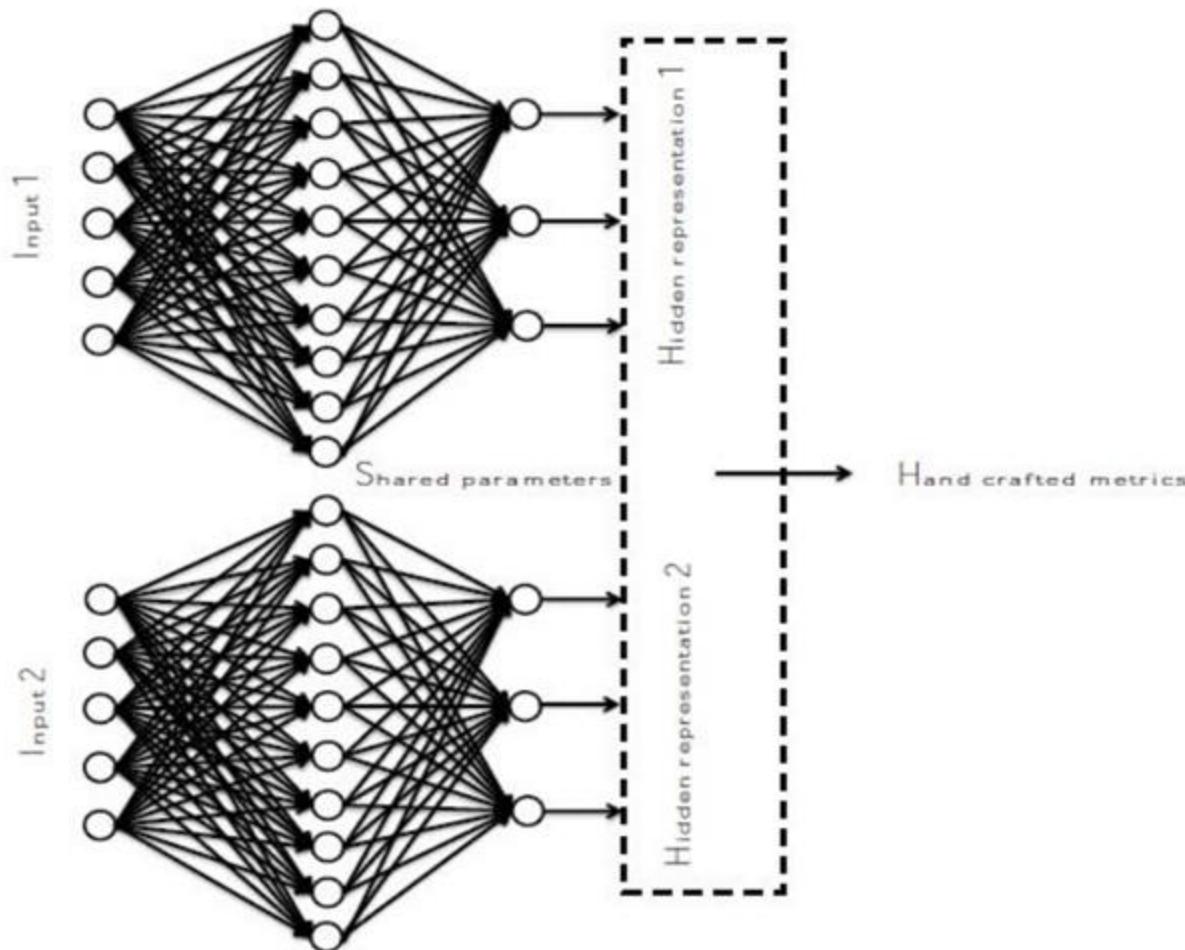
The President of the United States of America (POTUS) is the elected head of state and head of government of the United States. The president leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States.

The President of the United States of America (POTUS) is the elected head of state and head of government of the United States. The president leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States.

- › Традиционные модели инфопоиска работают исключительно на лексическом матчинге
- › Representation модели могут учитывать как все термы документа соотносятся с запросом
- › И те и другие могут быть смоделированы с помощью нейросетей

Инструмент

Siamese Neural Network



Word embeddings and convolution neural networks for NLP

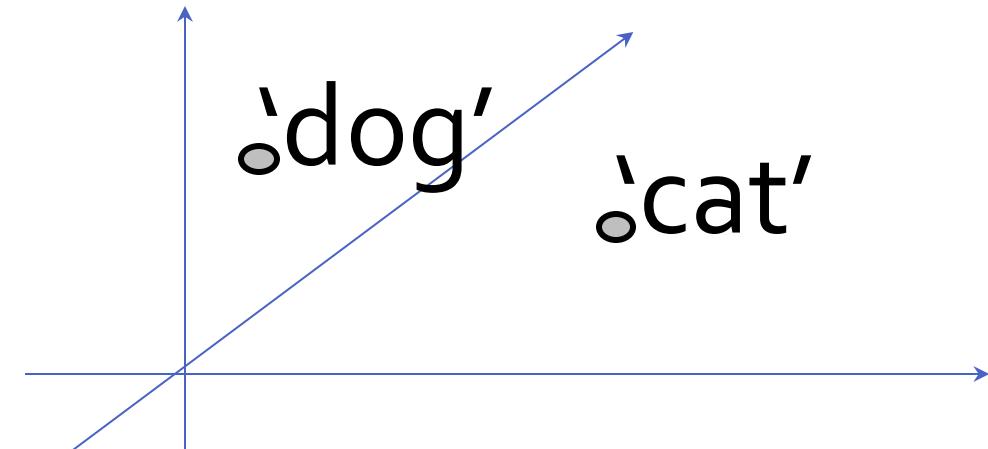
Semantic Level NLP

- Поиск синонимов
- Ответы на вопросы
- Чат боты
- Машинный перевод
- Построение аннотаций
-

Word-based NLP

Классический подход:

- Перенумеровываем слова (“один”, “два”)
- Используем *one-hot encoding*



$$x_{\text{мальчик}}: [0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$$

$$x_{\text{парень}}: [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$$

$$x_{\text{олень}}: [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Проблема: $\|x_{\text{мальчик}} - x_{\text{парень}}\| = \|x_{\text{парень}} - x_{\text{олень}}\|$

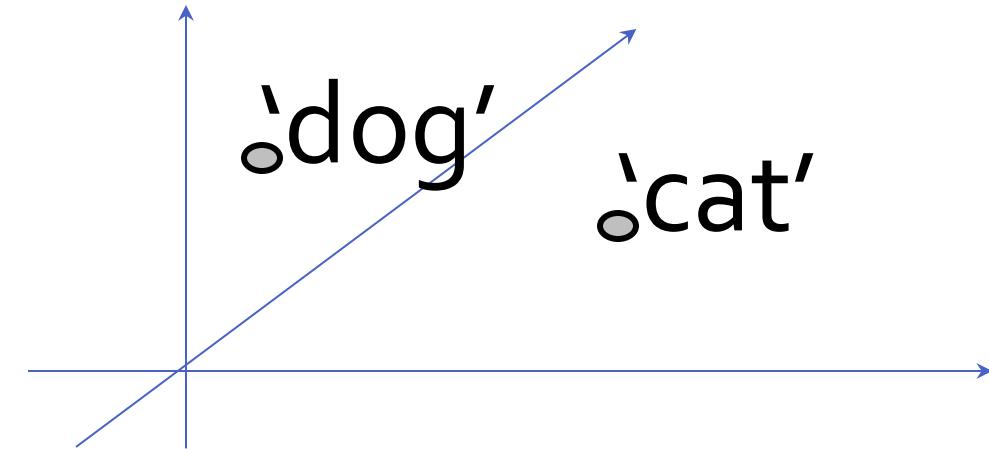
Решение: компактные представления

Word-based NLP

$x_{\text{мальчик}}: [0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$

$x_{\text{парень}}: [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$

$x_{\text{олень}}: [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$



Проблема: Очень высокая размерность пространства

Костыль 1: Лемматизация (“Парни -> парень”, “Шёл” ->“Идти”)

Side-effect 1: Лемматизация приводит к потере информации

Костыль 2: Заменить редкие слова заглушкой

Side-effect 2: Во многих задачах редкие слова крайне информативны

Дистрибутивная гипотеза

Гипотеза:

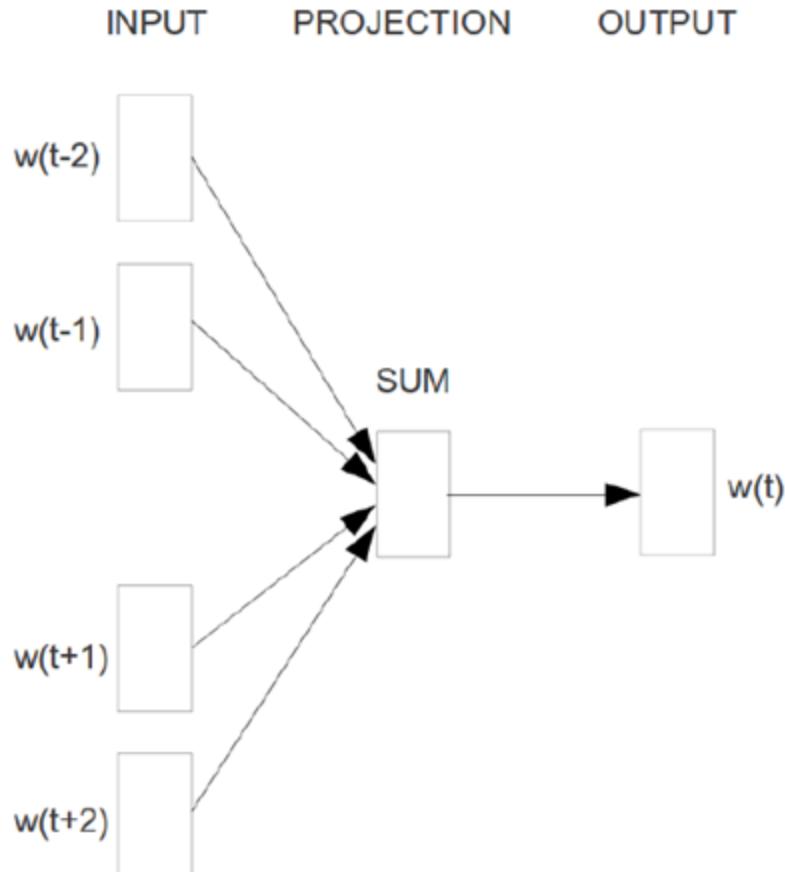
Лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения.

Вывод:

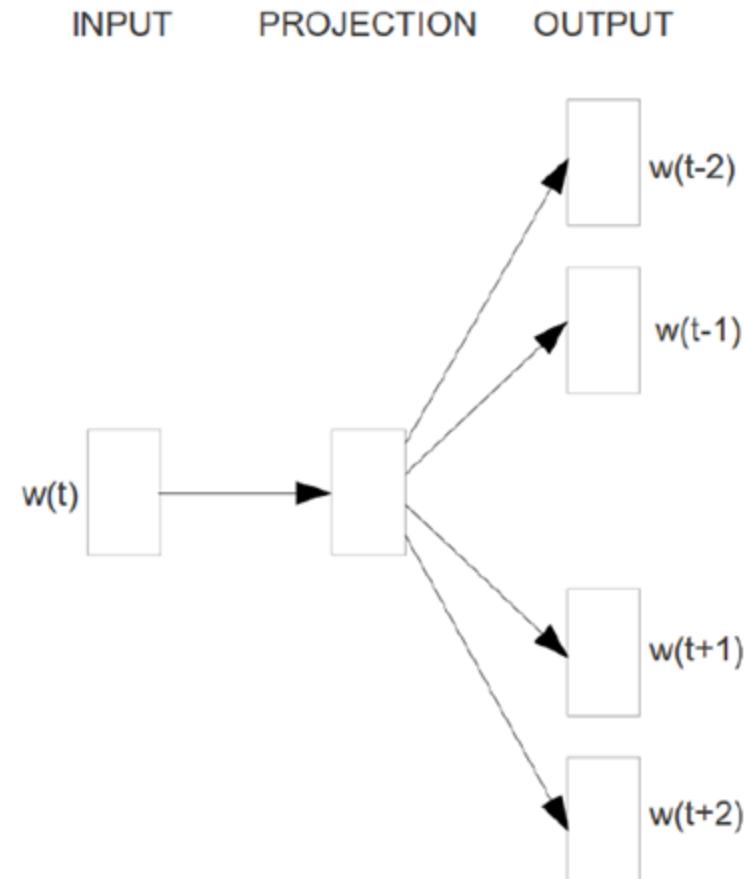
Значит, представления слов, можно построить с помощью контекстов, в которых эти слова употребляются.

Word2Vec (2013)

Архитектуры CBOW и Skip-gram

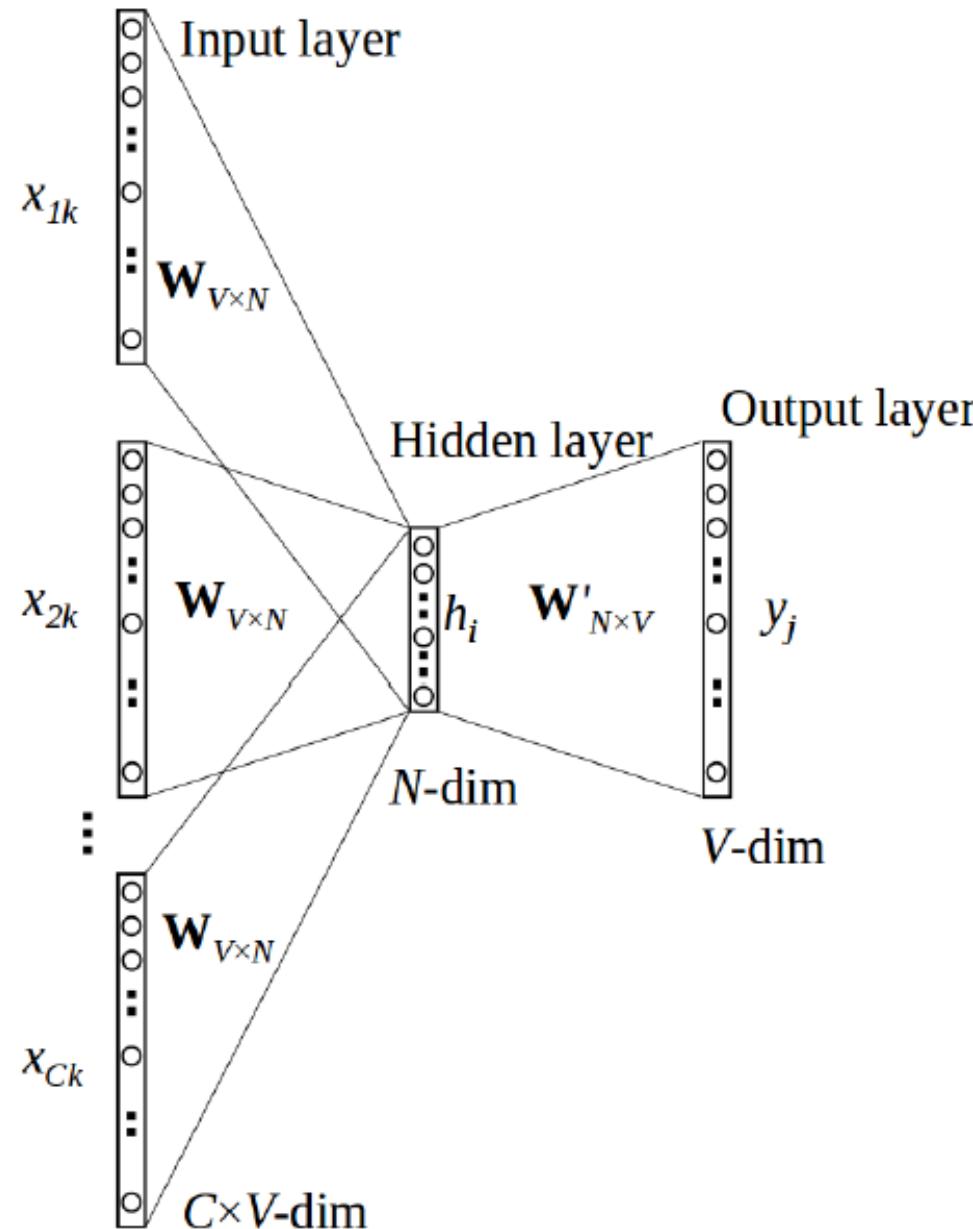


CBOW

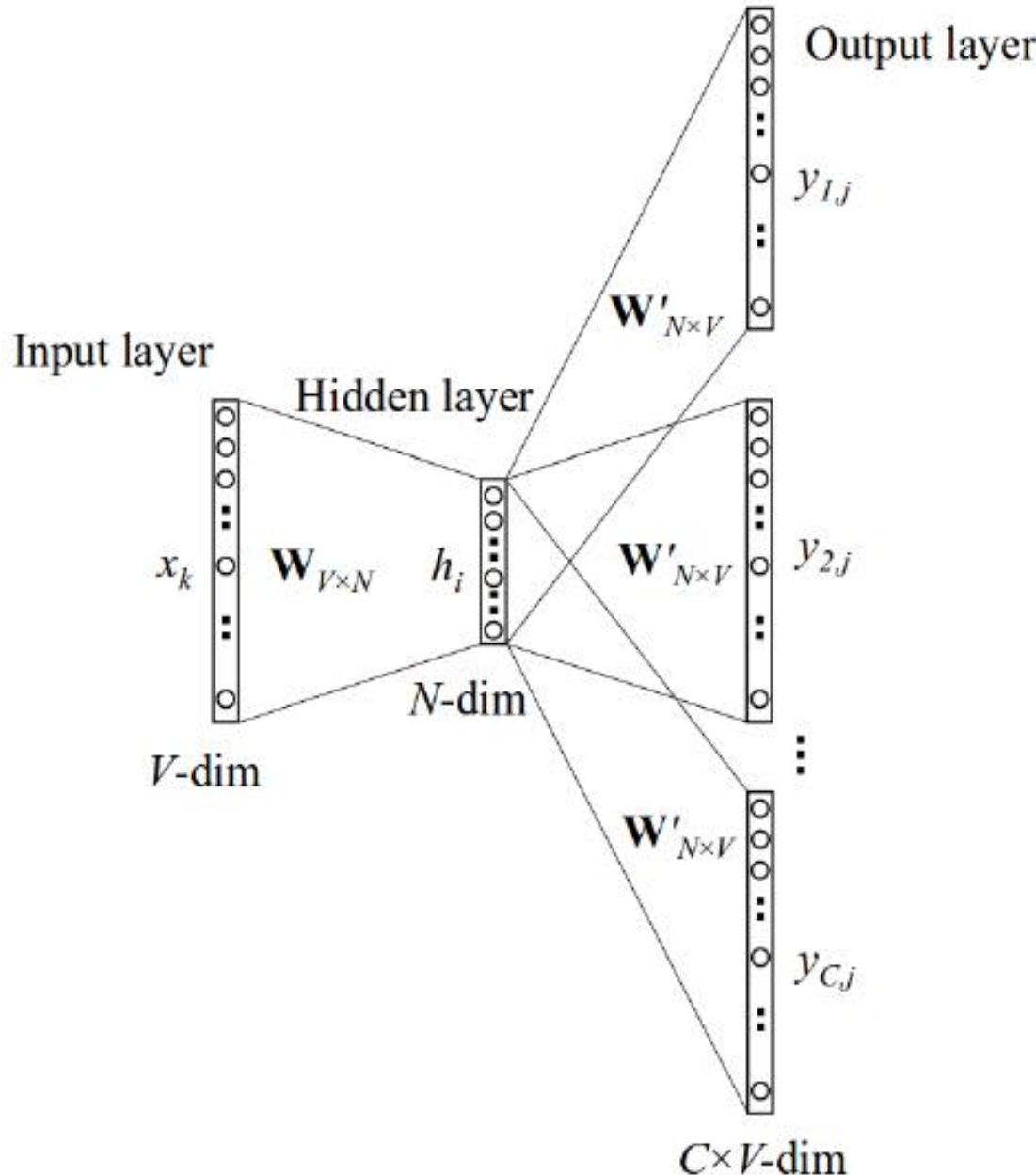


Skip-gram

CBOW (Continuous Bag of Words)



Skip-gram



Skip-gram

Оптимизируем

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$

Skip-gram

Вычисление вероятностей выходных слов

Используется softmax

$$p(w_O|w_I) = \frac{\exp \langle \mathbf{v}'_{w_O}, \mathbf{v}_{w_I} \rangle}{\sum_{w=1}^{|V|} \exp(\langle \mathbf{v}'_w, \mathbf{v}_{w_I} \rangle)}$$

На практике эту формулу применять сложно, так как вычисление градиента пропорционально $|V|$.

На практике применяют разные аппроксимации: иерархический softmax или negative sampling.

Negative sampling

Идея:

Не будем рассматривать все слова из словаря, а учтем только рассматриваемое слово + подмешаем еще k отрицательных примеров.

Заменим $\log p(w_O|w_I)$ на

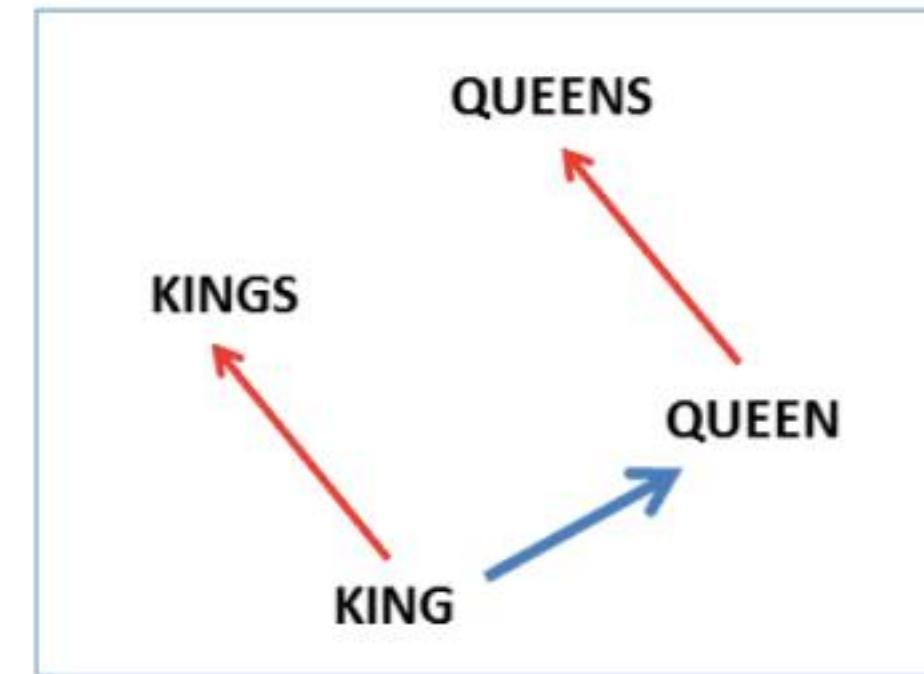
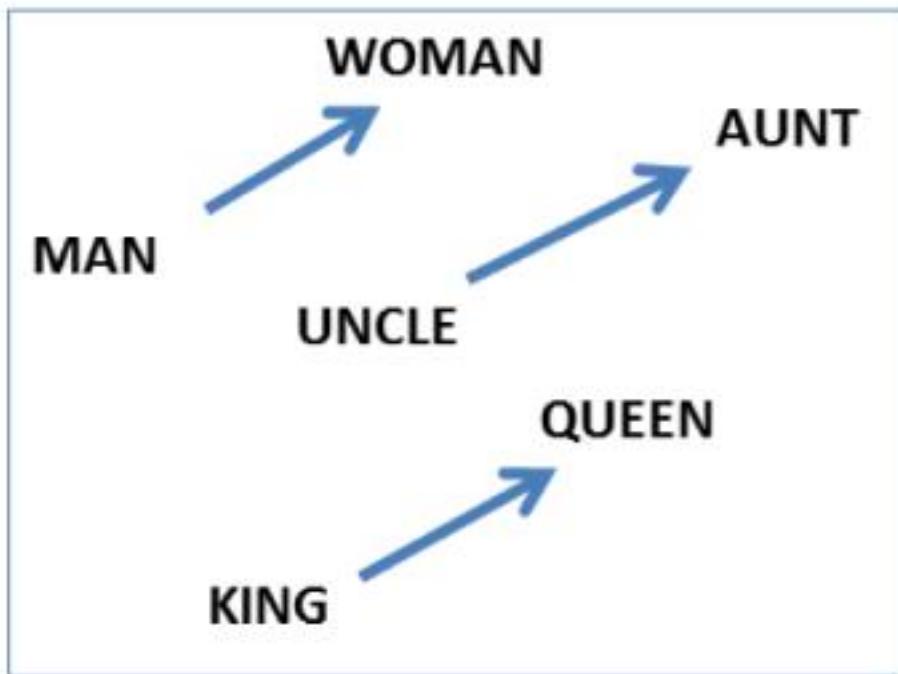
$$\log \sigma(\langle \mathbf{v}'_{w_O}, \mathbf{v}_{w_I} \rangle) + \sum_{i=1}^k \mathbb{E}_{w_i \sim P_n(w)} [\log \sigma(-\langle \mathbf{v}'_{w_i}, \mathbf{v}_{w_I} \rangle)]$$

- ▶ $k \approx 5 - 20$ для небольших выборок
- ▶ $k \approx 2 - 5$ для больших данных

$$\frac{1}{Z} P_{\text{empirical}}^{3/4}(w)$$

Свойства выученных представлений

$$\text{vec}(\text{"man"}) - \text{vec}(\text{"king"}) + \text{vec}(\text{"woman"}) = \text{vec}(\text{"queen"})$$



Пример модели

Ближайшие по контекстам слова для *дефицит* (с косинусной близостью)

1. [недостаток](#) 0.65354729
2. [профицит](#) 0.62004328
3. [госдолг](#) 0.61431211
4. [несбалансированность](#) 0.59282905
5. [дефицитность](#) 0.58463109
6. [дисбаланс](#) 0.57497036
7. [недопоступление](#) 0.57044709

Ближайшие по контекстам слова для *сингл* (с косинусной близостью)

1. [альбом](#) 0.91364485
2. [лонгплей](#) 0.88309568
3. [пластинка](#) 0.85516018
4. [кавер](#) 0.84436756
5. [студийник](#) 0.83440500

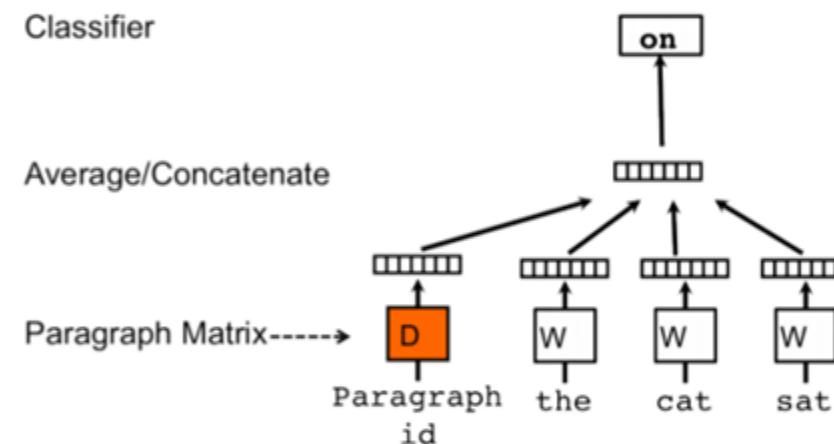
Doc2vec (Mikolov, 2014)

- Обобщение word2vec на целые документы (фразы, предложения и т.д.)
- Преобразует текст произвольной длины в вектор фиксированного размера
- Distributed Memory (DM)
- Distributed Bag of Words (DBOW)

Doc2vec (Mikolov, 2014)

Distributed Memory (DM)

- ▶ Назначим и рандомно проинициализируем paragraph vector
- ▶ Будем предсказывать слово из текста используя контекст и paragraph vector
- ▶ Идем скользящим окном по всему документу, сохраняя при этом paragraph vector фиксированным (поэтому Distributed Memory)
- ▶ Обновление происходит при помощи SGD и backprop

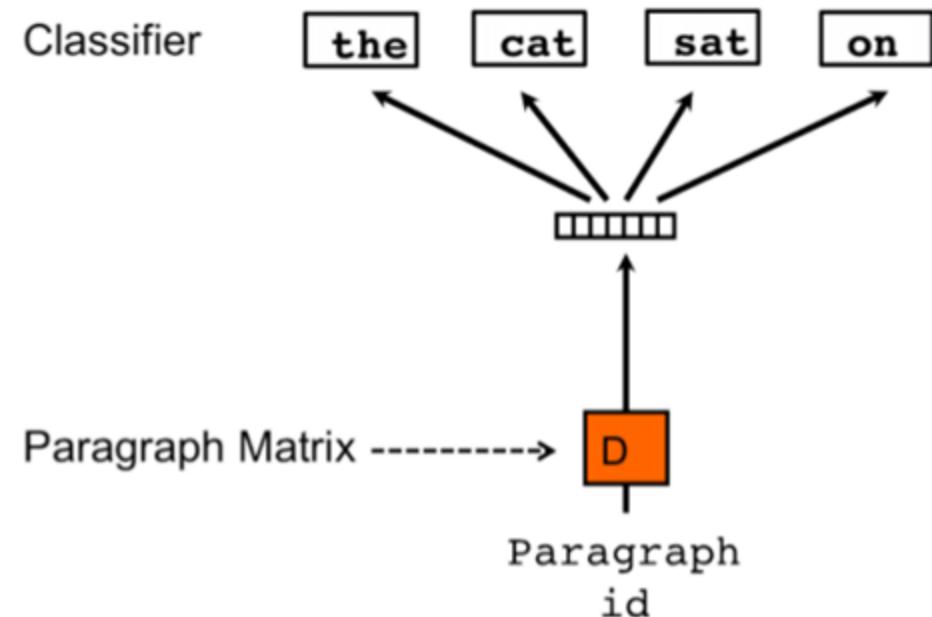


Doc2vec (Mikolov, 2014)

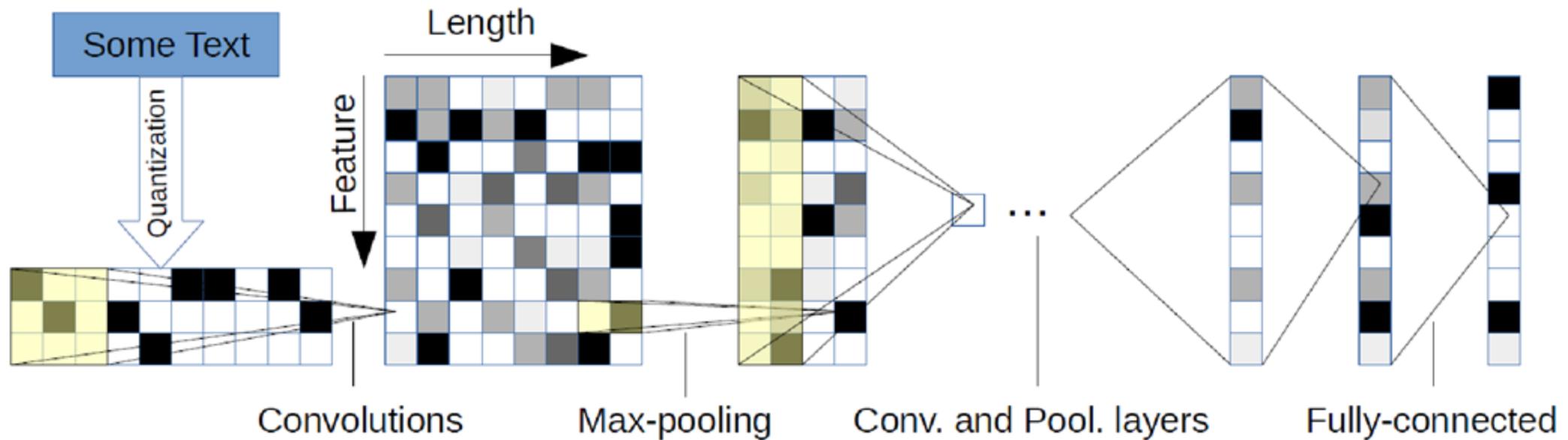
Distributed Bag of Words (DBOW)

- ▶ Используем только paragraph vector (вектора слов не используем)
- ▶ Берем окно из слов в параграфе и случайно семплируем какое из слов предсказать используя paragraph vector (игнорируем порядок слов)
- ▶ Очень просто и требует меньше ресурсов

Classifier

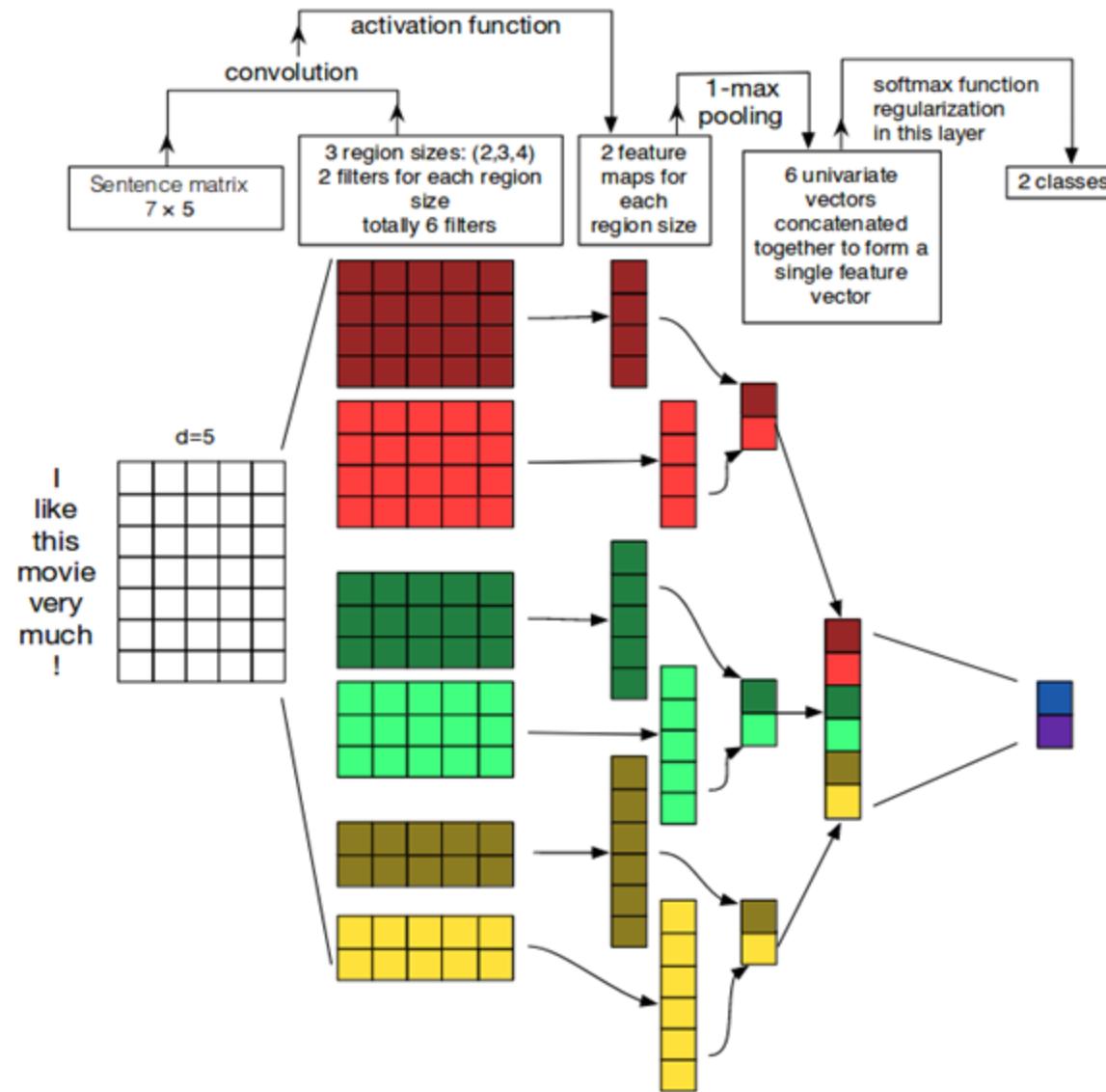


Character-level ConvNet

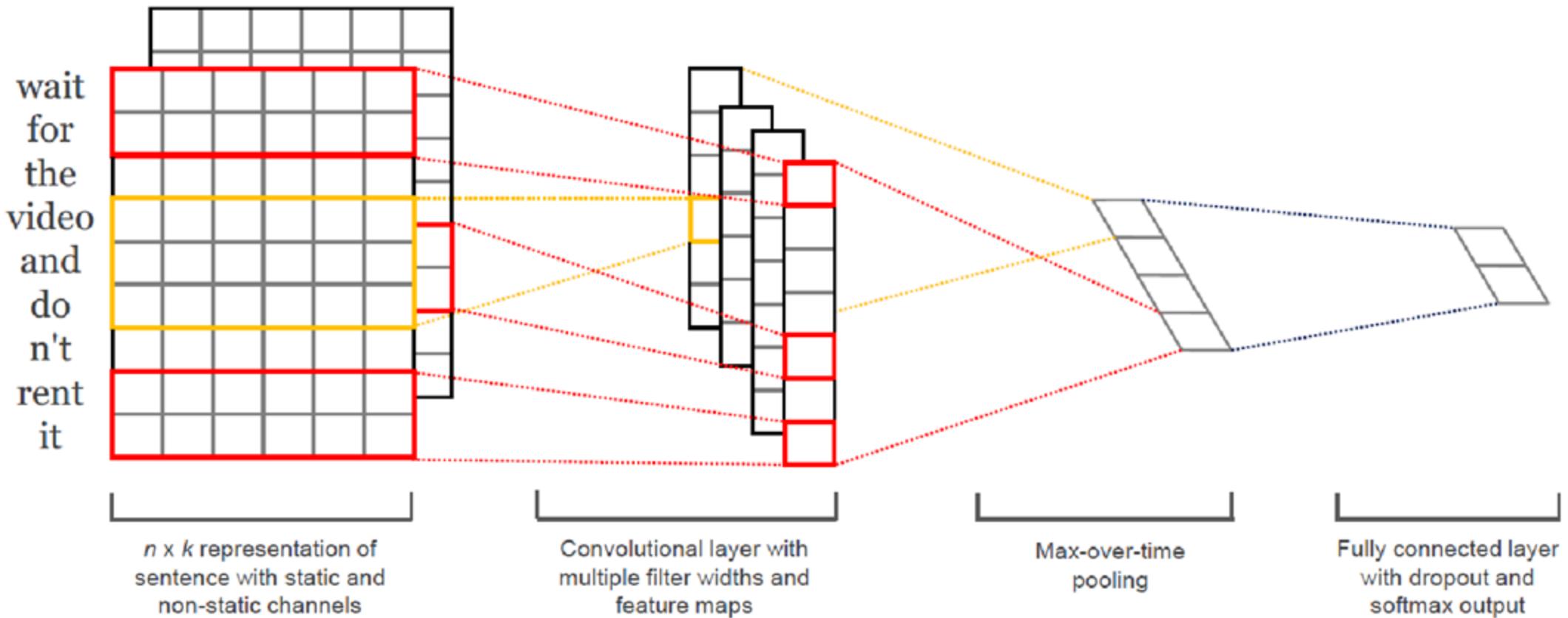


Trained for text classification

Векторное представление текста как изображения



Convolutional Neural Networks for Sentence Classification



Обычная сверточная сеть, проинициализированная word2vec представлениями слов

Deep Structured Semantic Model (DSSM) (2013)

Posterior probability
computed by softmax

Relevance measured
by cosine similarity

Semantic feature

Multi-layer non-linear projection

Word Hashing

Term Vector

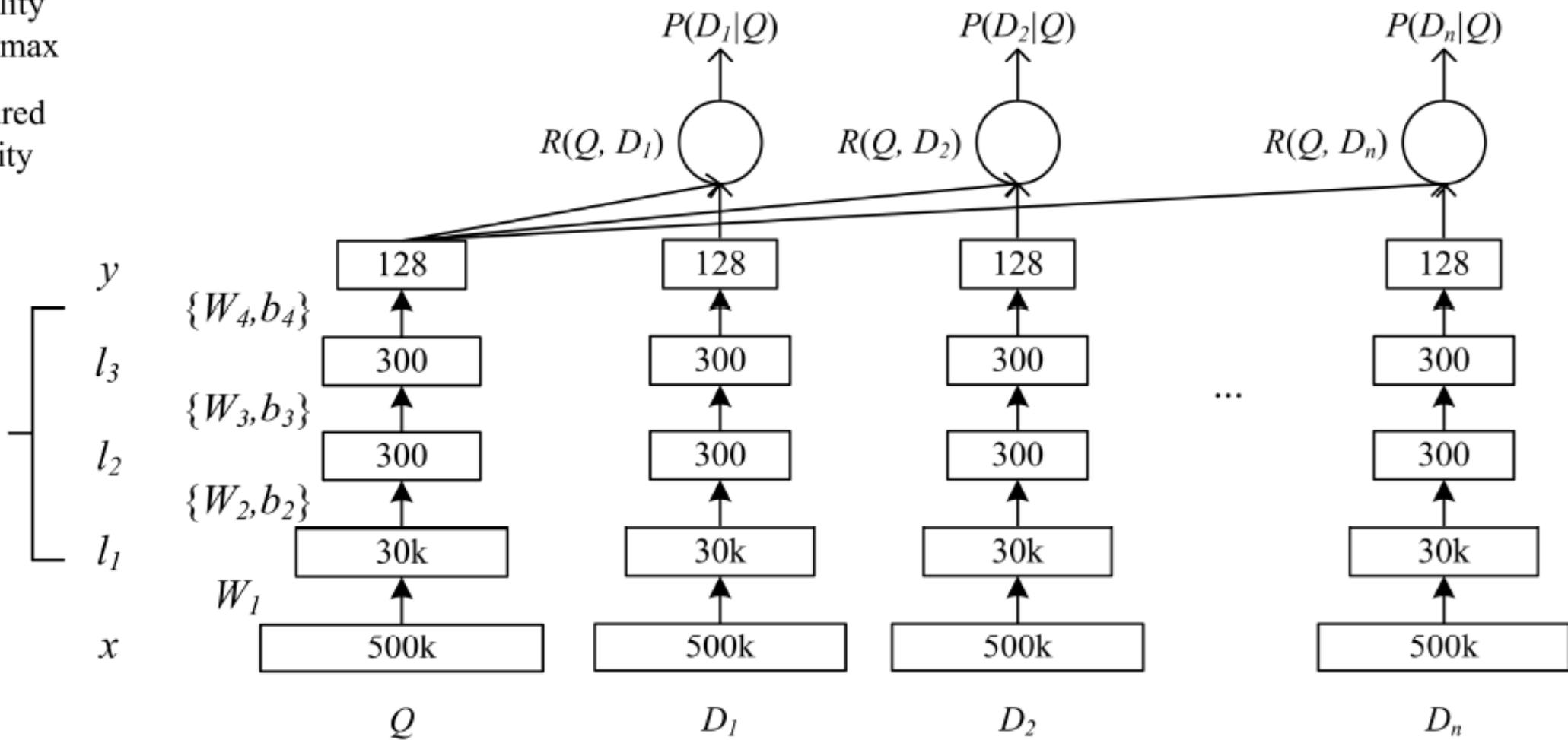


Figure 1: Illustration of the DSSM. It uses a DNN to map high-dimensional sparse text features into low-dimensional dense features in a semantic space. The first hidden layer, with 30k units, accomplishes word hashing. The word-hashed features are then projected through multiple layers of non-linear projections. The final layer's neural activities in this DNN form the feature in the semantic space.

Deep Structured Semantic Model (DSSM) (2013)

Collisions

Word Size	Letter-Bigram		Letter-Trigram	
	Token Size	Collision	Token Size	Collision
40k	1107	18	10306	2
500k	1607	1192	30621	22

Table 1: Word hashing token size and collision numbers as a function of the vocabulary size and the type of letter ngrams.

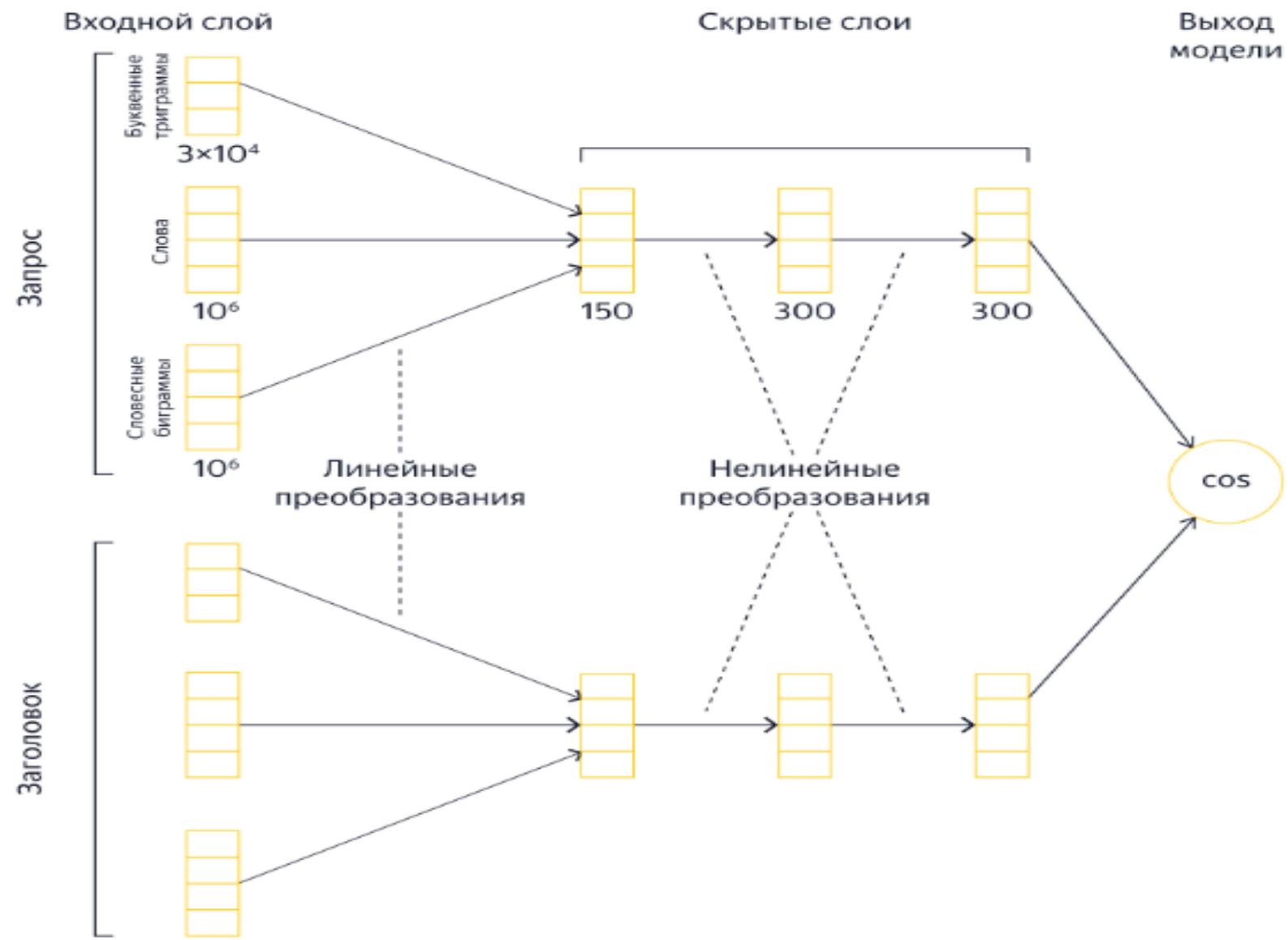
Learning

$$R(Q, D) = \frac{y_Q^T y_D}{\|y_Q\| \|y_D\|}$$

$$P(D|Q) = \frac{\exp(R(Q, D))}{\sum_{D' \in \mathcal{D}} \exp(R(Q, D'))}$$

$$L = -\log \prod_{Q, D^+} P(D^+|Q) \rightarrow \min$$

DSSM в Яндексе



DSSM в яндексе

Запрос: [келлская книга]

Заголовок страницы	BM25	Нейронная модель
келлская книга википедия	0.91	0.92
ученые исследуют келлскую книгу вокруг света	0.88	0.85
book of kells wikipedia	0	0.81
ирландские иллюстрированные евангелия vii viii вв	0	0.58
икеа гипермаркеты товаров для дома и офиса ikea	0	0.09

DSSM в яндексе

Запрос: [евангелие из келлса]

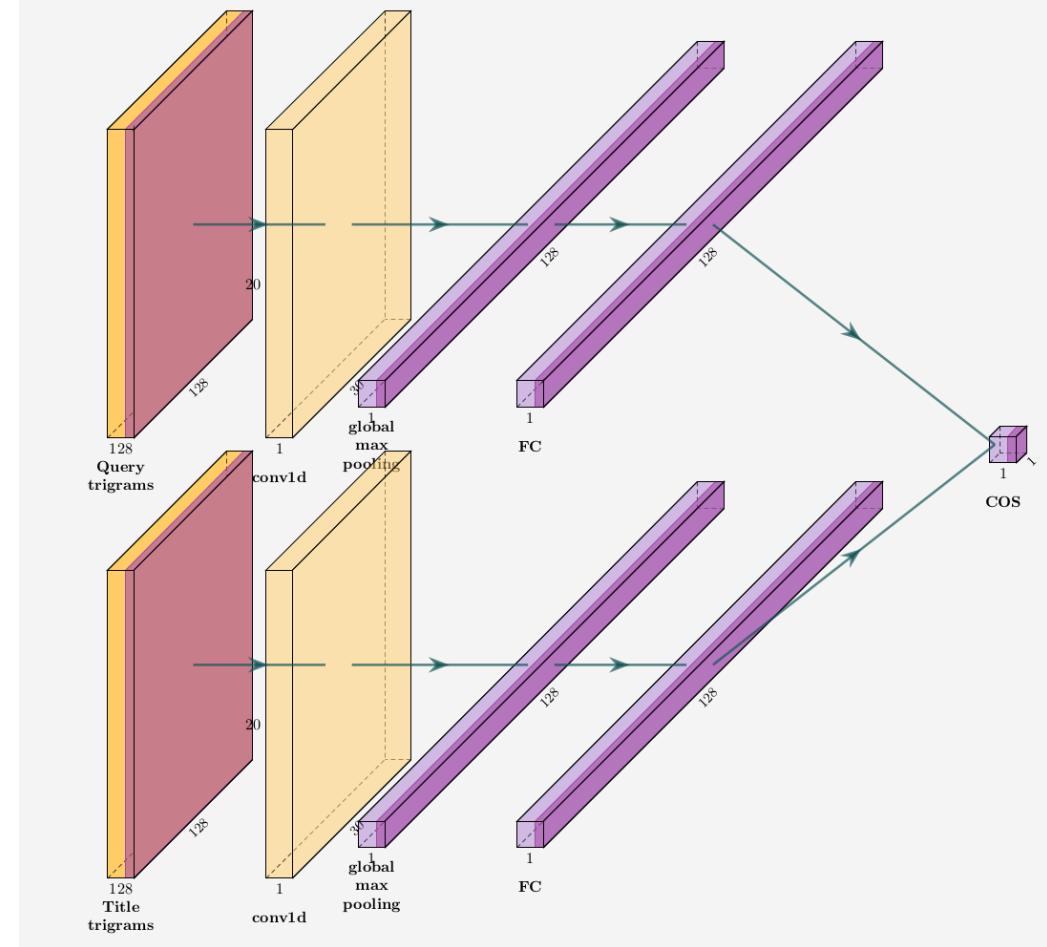
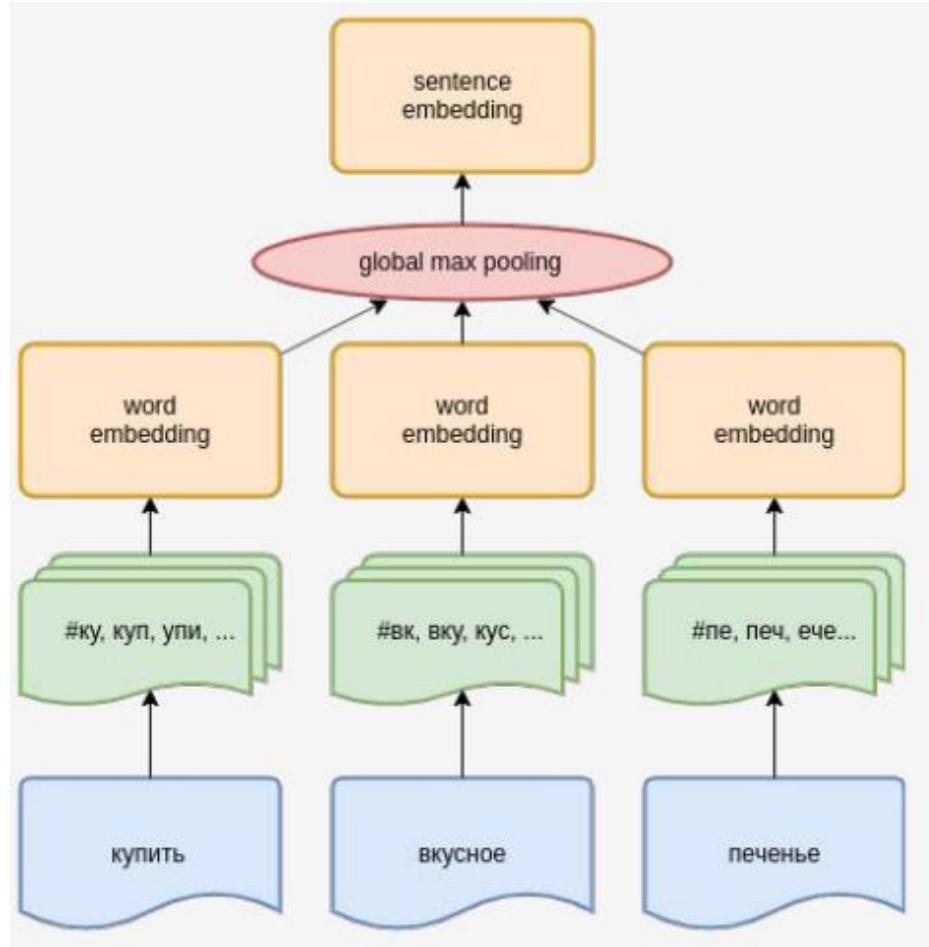
Заголовок страницы	BM25	Нейронная модель
келлская книга википедия	0	0.85
ученые исследуют келлскую книгу вокруг света	0	0.78
book of kells wikipedia	0	0.71
ирландские иллюстрированные евангелия vii viii вв	0.33	0.84
икеа гипермаркеты товаров для дома и офиса ikea	0	0.10

DSSM в яндексе

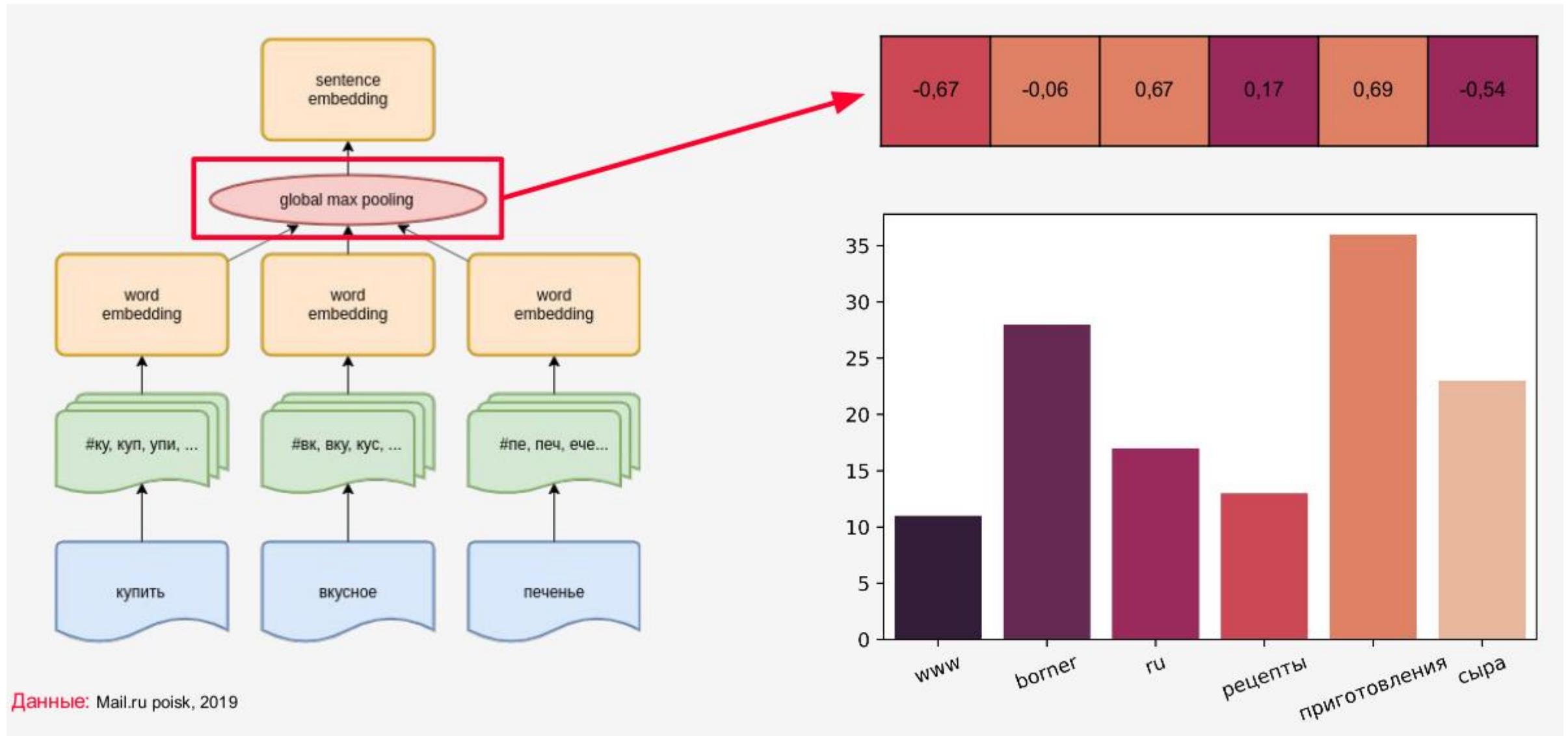
Запрос: [рассказ в котором раздавили бабочку]

Заголовок страницы	BM25	Нейронная модель
фильм в котором раздавили бабочку	0.79	0.82
и грянул гром википедия	0	0.43
брэдбери рэй википедия	0	0.27
машина времени роман википедия	0	0.24
домашнее малиновое варенье рецепт заготовки на зиму	0	0.06

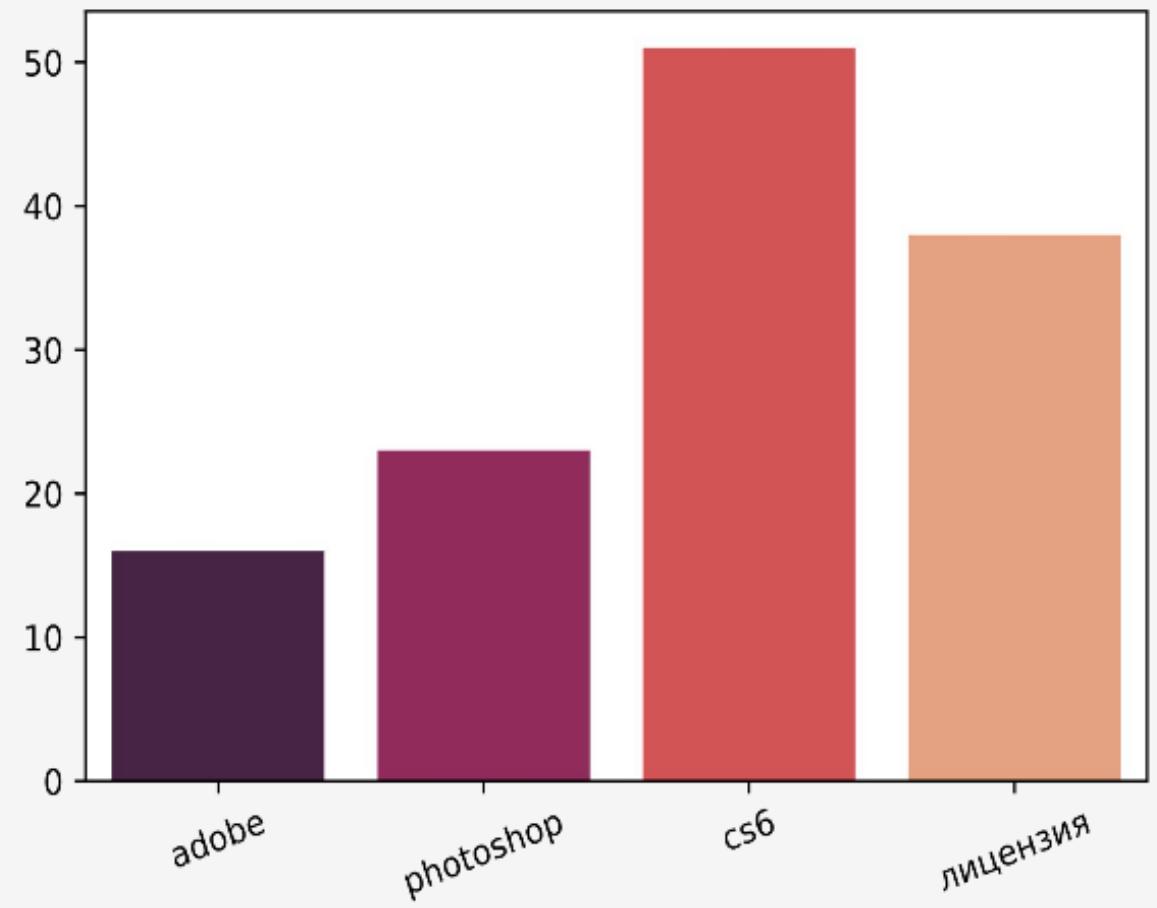
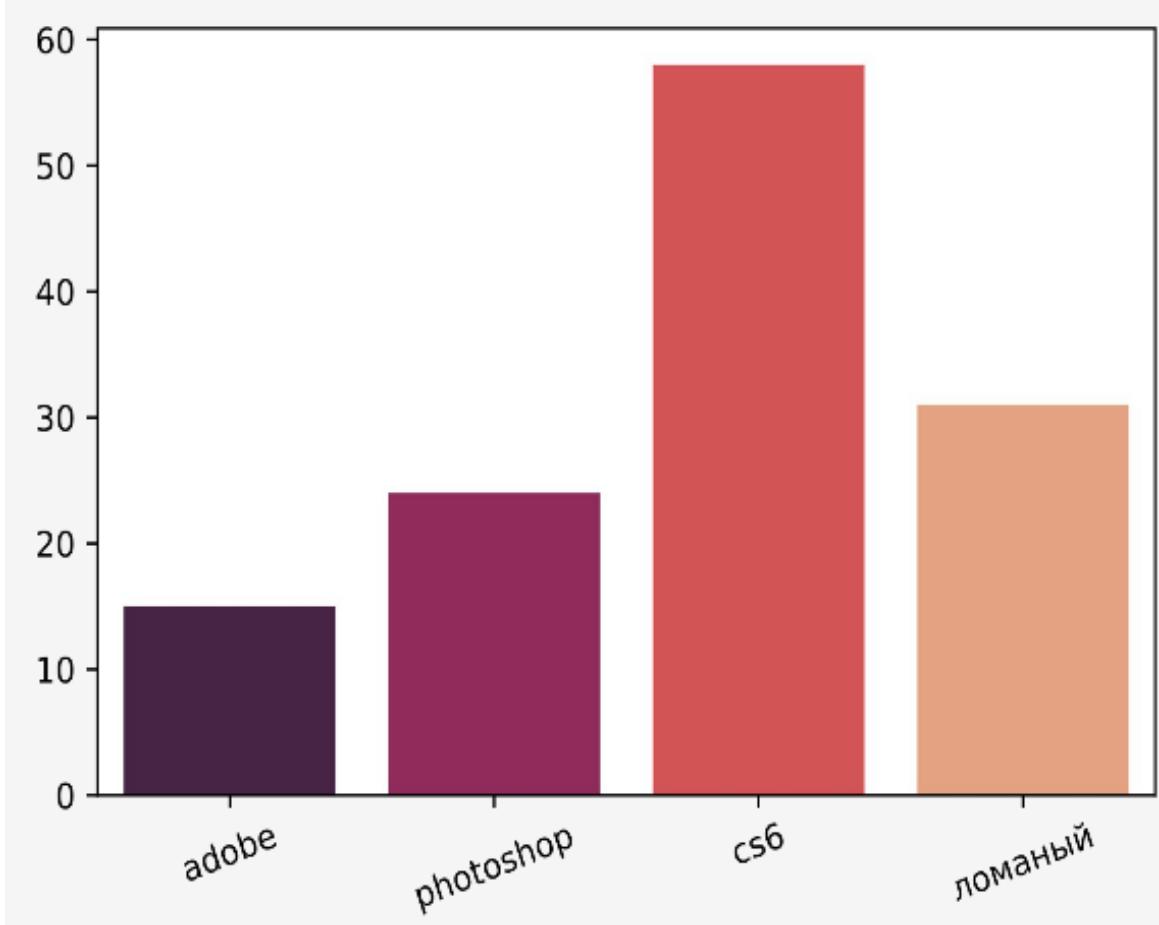
DSSM в Mail.ru



Важность слов в запросе



Важность слов в запросе



DSSM в Mail.ru

Запрос	Документ	BM25	Нейроранк
астана тойота	Официальный сайт Toyota в Казахстане	0	0.839
поэзия	Стихи.ру	0	0.823
ворлд дрим паттайя	Бангкок Dream World	0	0.818
спб картошка	Купить картофель в Санкт-Петербурге	0	0.811

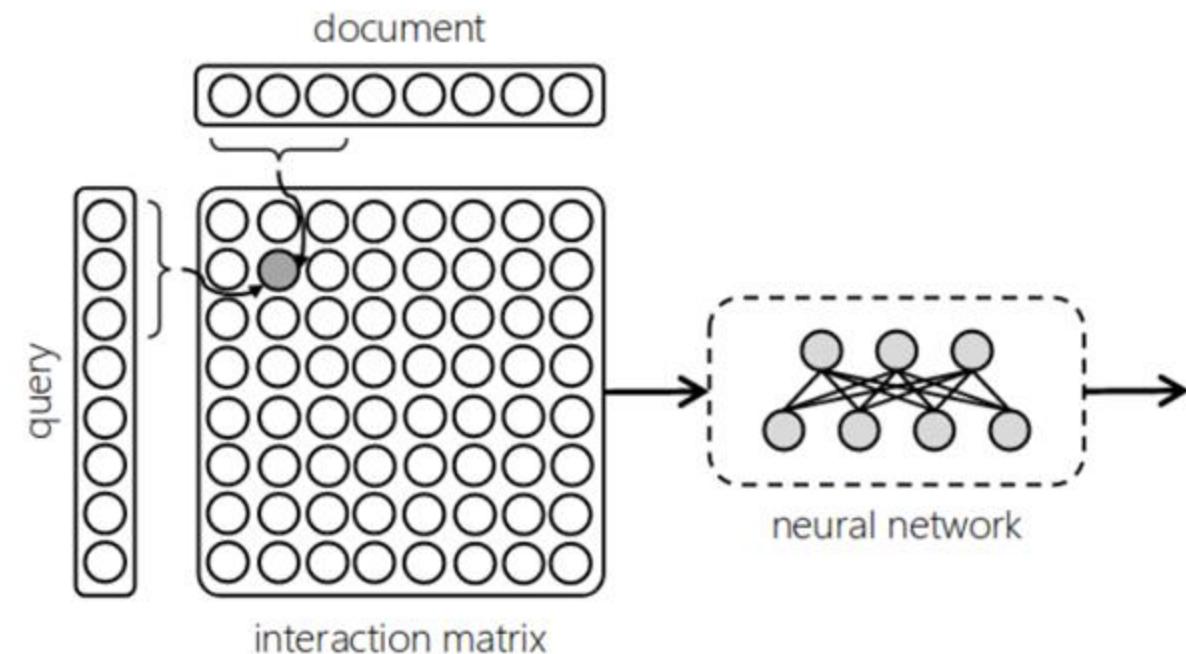
Interaction matrix based approaches

Alternative to Siamese networks

Interaction matrix X , where $x_{i,j}$ is obtained by comparing the the i^{th} word in source sentence with j^{th} word in target sentence

Comparisons can be both lexical or semantic

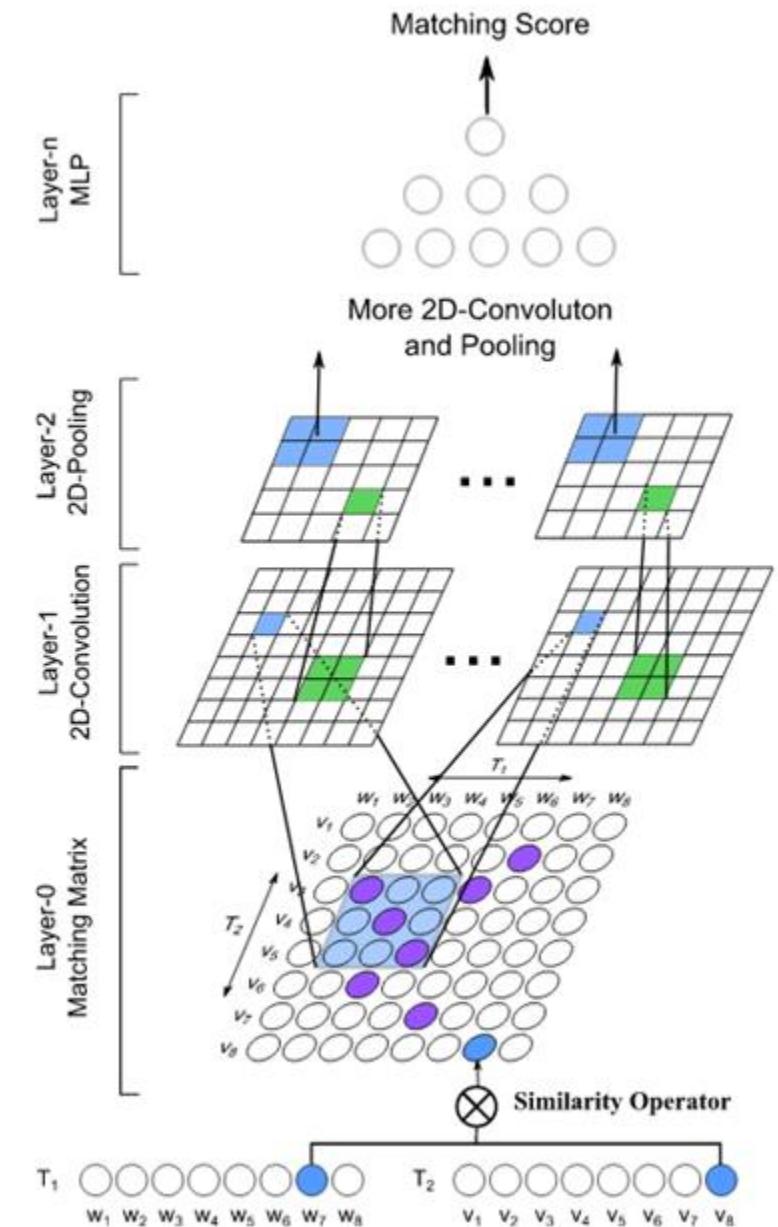
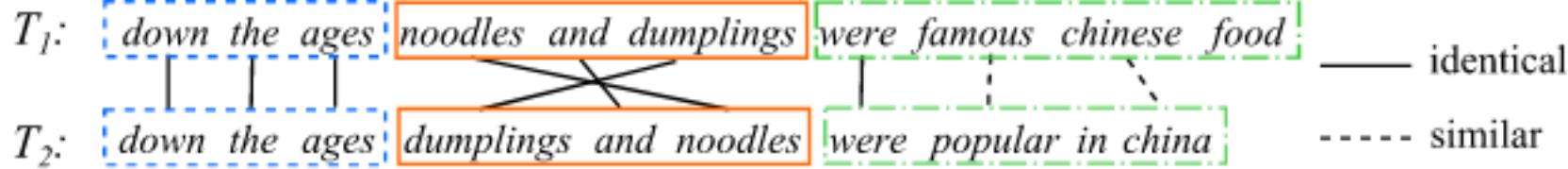
Вопрос:



- ▶ Как такое использовать с линейными моделями?

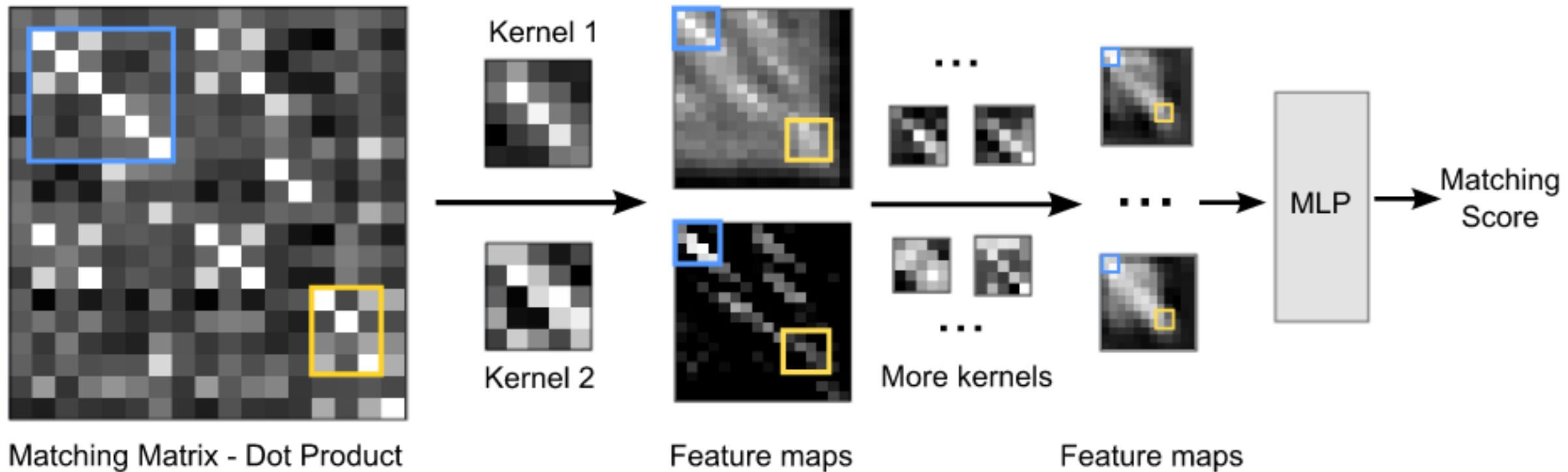
MatchPyramid

- Сверточные слои поверх interaction matrix между запросом и документом
- После конволюций полно связные слои



MatchPyramid

T₁: PCCW's chief operating officer, Mike Butcher, and Alex Arena, the chief financial officer, will report directly to Mr So.
T₂: Current Chief Operating Officer Mike Butcher and Group Chief Financial Officer Alex Arena will report to So.



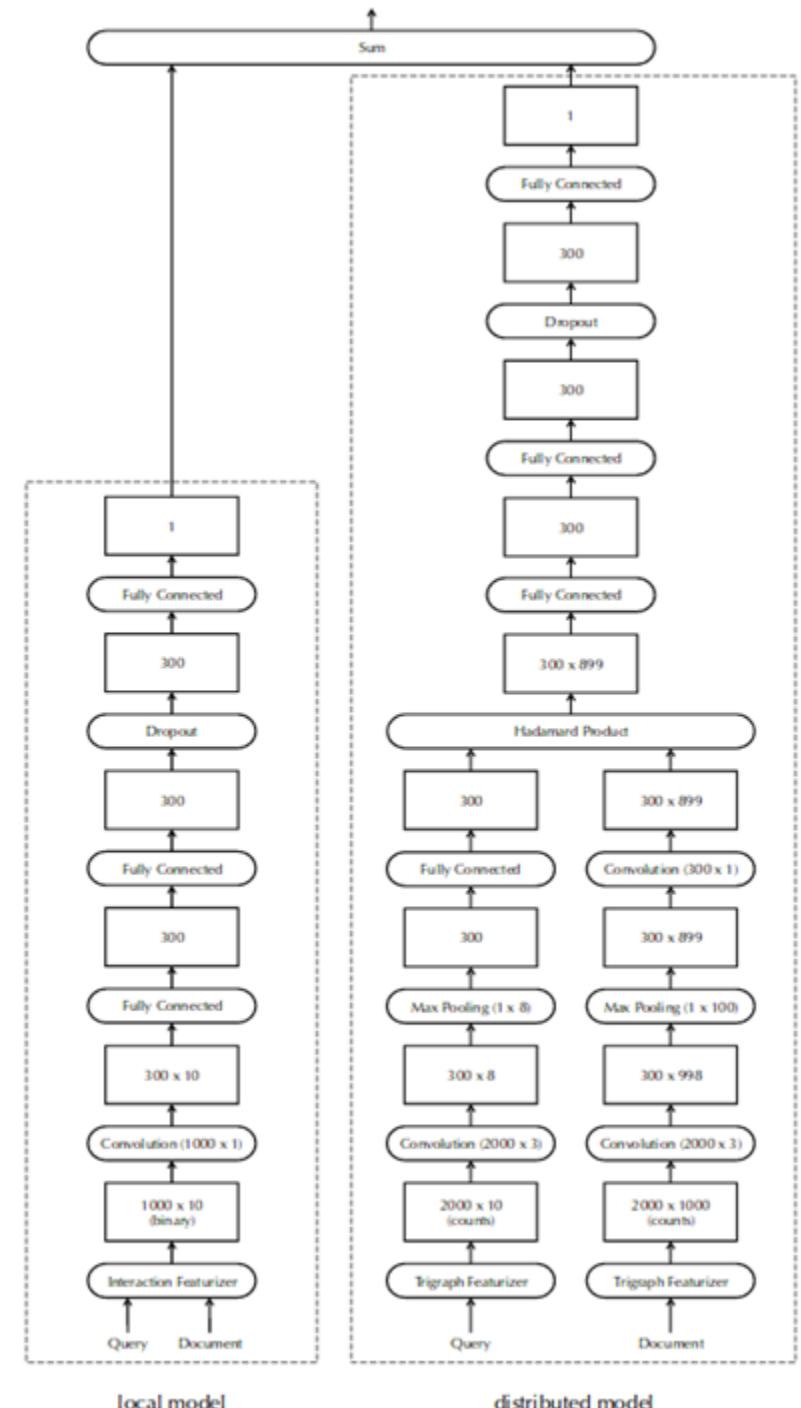
Duet Model

- Совместно тренируем 2 подсетки, сфокусированные на лексическом и семантическом матчинге

Training sample:

$q, d^+, d_1, d_2, d_3, d_4$

$$p(d^+|q) = \frac{e^{nrdm(q, d^+)}}{\sum_{d \in D^-} e^{nrdm(q, d)}}$$



Duet Model

Term importance

Local model

Only query terms have an impact

Earlier occurrences have bigger impact

Query: united states president

The President of the United States of America (POTUS) is the elected head of state and head of government of the United States. The president leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States.

Visualizing impact of dropping terms on model score

Duet Model

Term importance

Distributed model

Non-query terms (e.g., *Obama* and *federal*) has positive impact on score

Common words like 'the' and 'of' probably good indicators of well-formedness of content

Query: united states president

The President of the United States of America (POTUS) is the elected head of state and head of government of the United States. The president leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States.

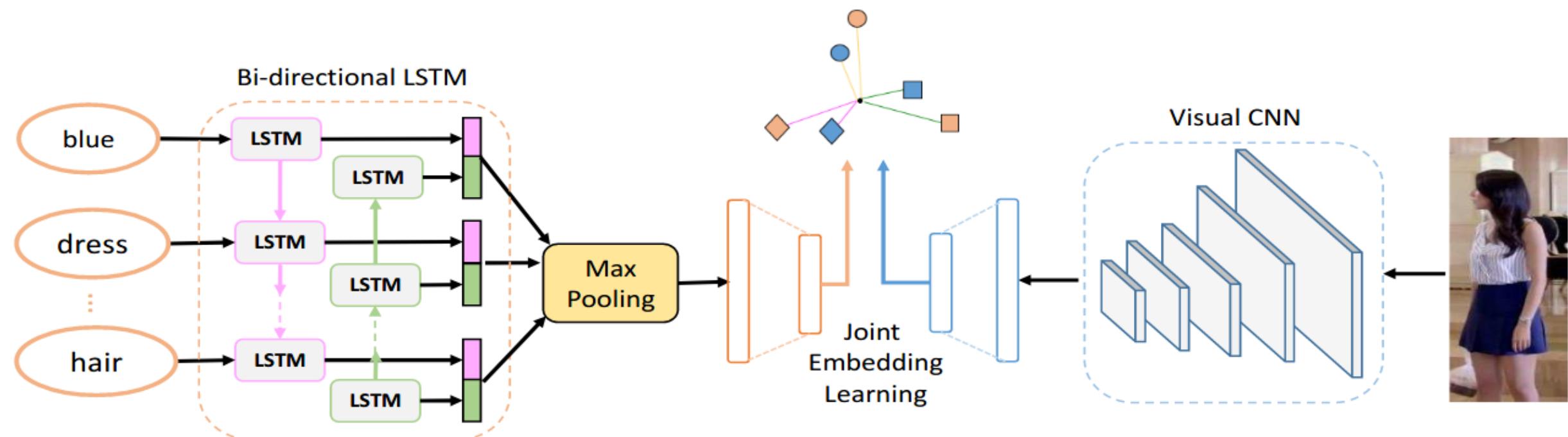
Visualizing impact of dropping terms on model score

Duet Model (Результаты)

	NDCG@1	NDCG@10
Non-neural baselines		
LSA	22.4	44.2
BM25	24.2	45.5
DM	24.7	46.2
QL	24.6	46.3
Neural baselines		
DRMM	24.3	45.2
DSSM	25.8	48.2
CDSSM	27.3	48.2
DESM	25.4	48.3
Our models		
Local model	24.6	45.1
Distributed model	28.6	50.5
Duet model	32.2	53.0

Матчинг картинок

Можно использовать аппарат симских нейросетей и для сопоставления текстовых запросов с изображениями в поиске по картинкам



Матчинг картинок

Query: птичка

mark: 0.3991589; id: 100033



mark: 0.45618516; id: 45853



mark: 0.46587723; id: 17447



mark: 0.47023547; id: 25005

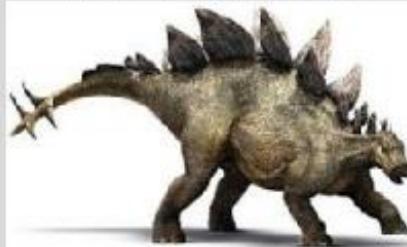


Query: тиранозавр

mark: 0.5119386; id: 329417



mark: 0.5251976; id: 370003



mark: 0.5283124; id: 434927



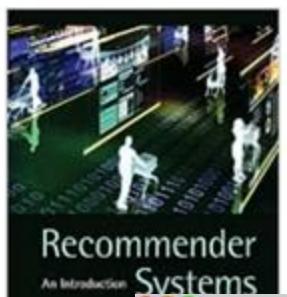
mark: 0.5353282; id: 30751



Смежные задачи информационного поиска



Рекомендательные системы



Recommender Systems: An Introduction
by Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich

AVERAGE CUSTOMER RATING:

★★★★★ (Be the first to review)



Rегистrieren, um sehen zu können, was
deinen Freunden gefällt.

FORMAT:
Hardcover

NEW FROM BN.COM

\$65.00 List Price

\$52.00 Online Price
(You Save 20%)

Add to Cart

Spotify Premium - E

Customers

Table of Contents

Customers

Search Pattern

The Sunshine Underground
Raise The Alarm

Wake Up
The Sunshine Underground

Recommender Systems

lost.fm

Most played albums in the last 3 months

- Strangers In The Night (Remastered) - UFO
- Grace/Wastelands - Peter Doherty
- Natural Mystic - Bob Marley
- Audioslave - Audioslave
- Kiss This - Sex Pistols
- She Wants Revenge - She Wants Revenge
- Los Angeles - X
- Forever Delayed - Music Street Preachers
- AMERICAN CHARGE - AMERICAN CHARGE
- MARJORIE MEYERS - MARJORIE MEYERS
- MORIARTY - MORIARTY



PANDORA™
created by the Music Genome Project™

Реклама



Русские геймеры дождались

Игра "Игра Престолов" рвет все шаблоны.
Тысячи кланов и 0 доната. Ежедневно 1 млн
игроков.



Смартфоны Носимая электроника Компьютеры Телевизоры Камеры Бытовая техника Прочее

20 часов назад

Итоги слепого тестирования фронтальных камер. Какой из 9 флагманов делает лучшие селфи?

Мы провели действительно грандиозный тест и сравнили 9 селфи-камер флагманов. Голосование закончилось — суммарно читатели отдали больше 25 тысяч голосов. Настало время объявлять победителей и раскрывать карты, под каким номером скрывался тот или иной флагман. Помимо рейтинга читательских симпатий мы составили свой — редакционный. И не во всем оценки читателей и экспертов совпали.

Подписаться

Директ Учебный центр МГУТУ на Таганке! ipkit.ru Кадровое делопроизводство+1 за 11900 р. Удостоверение гос. университета! 18+ Адрес и телефон Москва

Английский язык для подростков big-been.ru/Школа-Английского Уникальная акция до 25.04! Обучение от 4896р. в месяц! Набор в группы! Адрес и телефон Москва

iphone 7

Интернет Картинки Видео Приложения Новости Ответы

Реклама

iPhone 7 - Официальный сайт Apple. apple.com/ru Реклама Это 7. Подробнее. Купить. Купить Новый (PRODUCT)RED™ Airpods Сравнение моделей

iPhone 7 в наличии. Москва! - 100% Оригинал. istore-apple.ru/iPhone-7 Реклама Фирменная гарантия 1 год. Магазин в центре Москвы. Доставка! Только оригинал Гарантия APPLE Удобная доставка Курьер-консультант Москва м. Новослободская +7 (495) 374-58-53 пн-вс 9:00-20:00

Apple iPhone 7 32Gb Silver / moscow.shop.megafon.ru moscow.shop.megafon.ru Реклама Цена 48990 руб. Купите в МегаФоне с бесплатной доставкой внутри МКАД! Подобрать номер Скидки за покупки Тарифы Контакты ★★★★☆ Рейтинг магазина Москва 8 (800) 550-58-58 круглосуточно

Сотовый телефон Apple iPhone 7 — Товары Mail.Ru torg.mail.ru Apple iPhone 7 128GB от 42 890 руб. Apple iPhone 7 256GB от 46 990 руб. Apple iPhone 7 32GB от 37 790 руб.

iPhone 7: новости, дата выхода и характеристики | Фото Apple Айфон 7 apple-geek.ru

Голосовые ассистенты

13



Recommendation systems



Recommender systems - The task

@ mail.ru
group

- Нужно построить модель, предсказывающую насколько user нравятся items
- У типичной рекомендательной системы имеется:
 - Матрица users и items
 - Рейтинги, отражающие их прошлые предпочтения
 - Задача состоит в предсказании будущих предпочтений
- В общем случае, все **НЕ** ограничивается предсказанием рейтинга



Approaches to recommender systems

- › Collaborative filtering

Основан на анализе предпочтений пользователей таких как рейтинги фильмов, лайки и т.д.

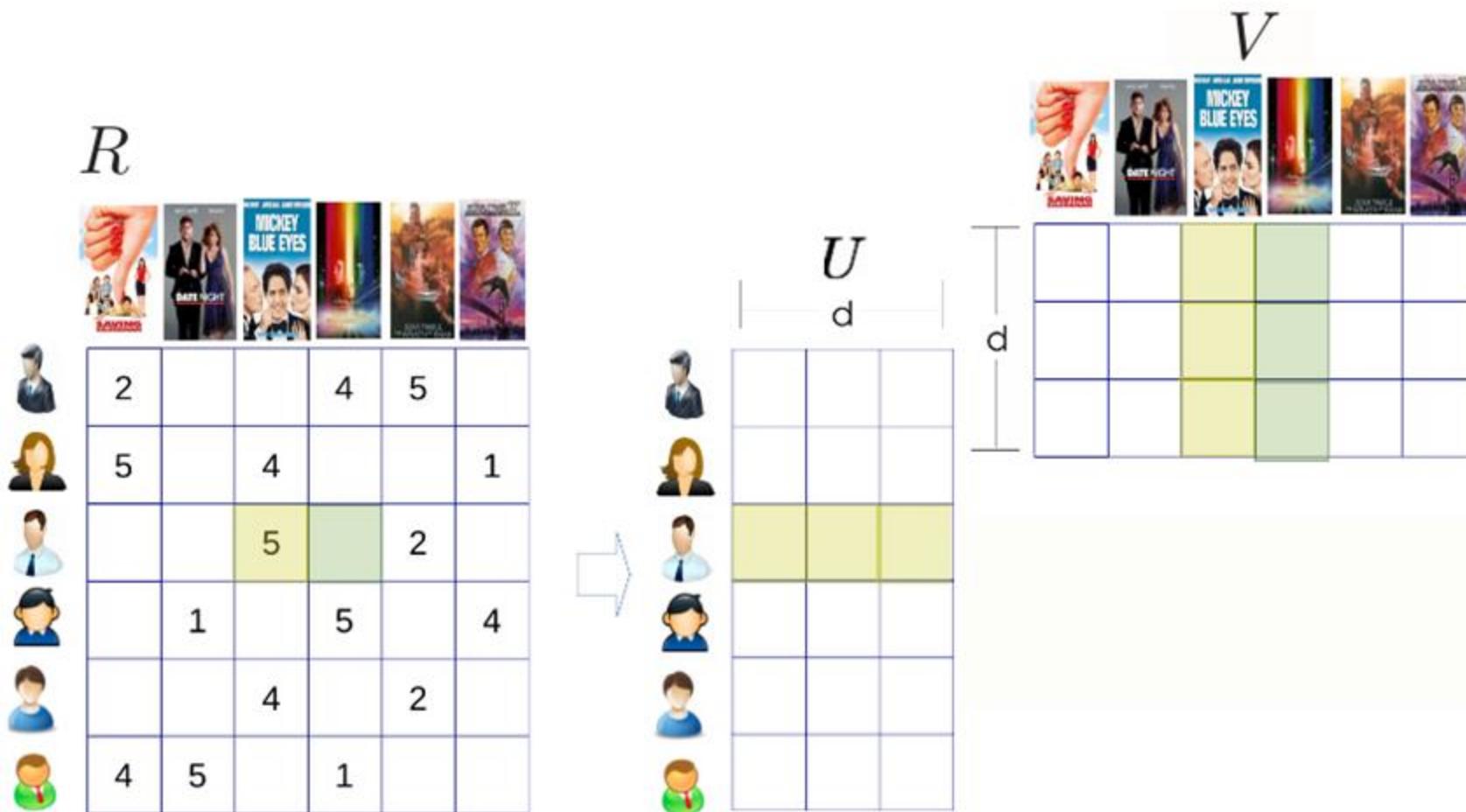
- › Content-based filtering

Основан на матчинге представлений items и users

Пользовательский профиль может представляться предыдущими рейтингами, посещенными страницами в интернете, поисковыми запросами и т.д.

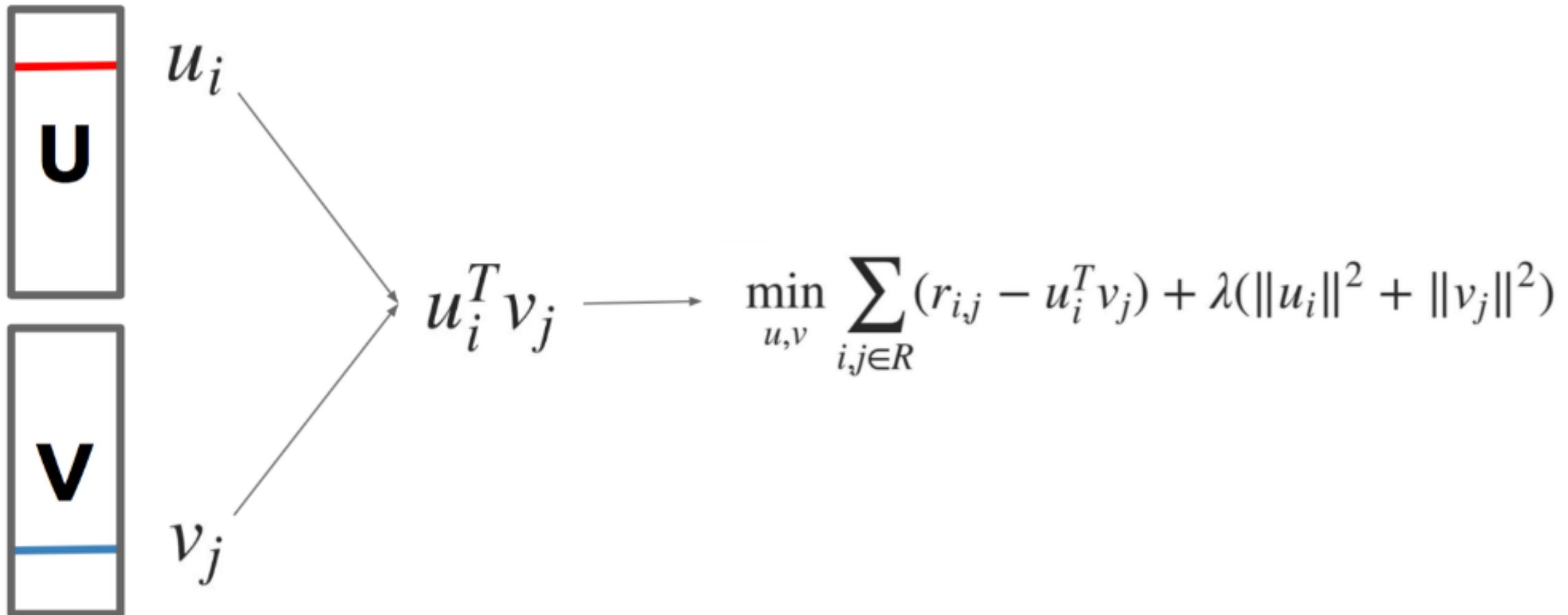
- › A hybrid approach

Matrix factorization



$$\min_{u,v} \sum_{i,j \in R} (r_{i,j} - u_i^T v_j) + \lambda(\|u_i\|^2 + \|v_j\|^2)$$

A feed-forward neural network view



Pytorch example

```
import torch
from torch.autograd import Variable

class MatrixFactorization(torch.nn.Module):
    def __init__(self, n_users, n_items, n_factors=20):
        super().__init__()
        # create user embeddings
        self.user_factors = torch.nn.Embedding(n_users, n_factors,
                                              sparse=True)
        # create item embeddings
        self.item_factors = torch.nn.Embedding(n_items, n_factors,
                                              sparse=True)

    def forward(self, user, item):
        # matrix multiplication
        return (self.user_factors(user)*self.item_factors(item)).sum(1)

    def predict(self, user, item):
        return self.forward(user, item)

model = MatrixFactorization(n_users, n_items, n_factors=20)
loss_fn = torch.nn.MSELoss()
optimizer = torch.optim.SGD(model.parameters(),
                            lr=1e-6)

for user, item in zip(users, items):
    # get user, item and rating data
    rating = Variable(torch.FloatTensor([ratings[user, item]]))
    user = Variable(torch.LongTensor([int(user)]))
    item = Variable(torch.LongTensor([int(item)]))

    # predict
    prediction = model(user, item)
    loss = loss_fn(prediction, rating)

    # backpropagate
    loss.backward()

    # update weights
    optimizer.step()
```

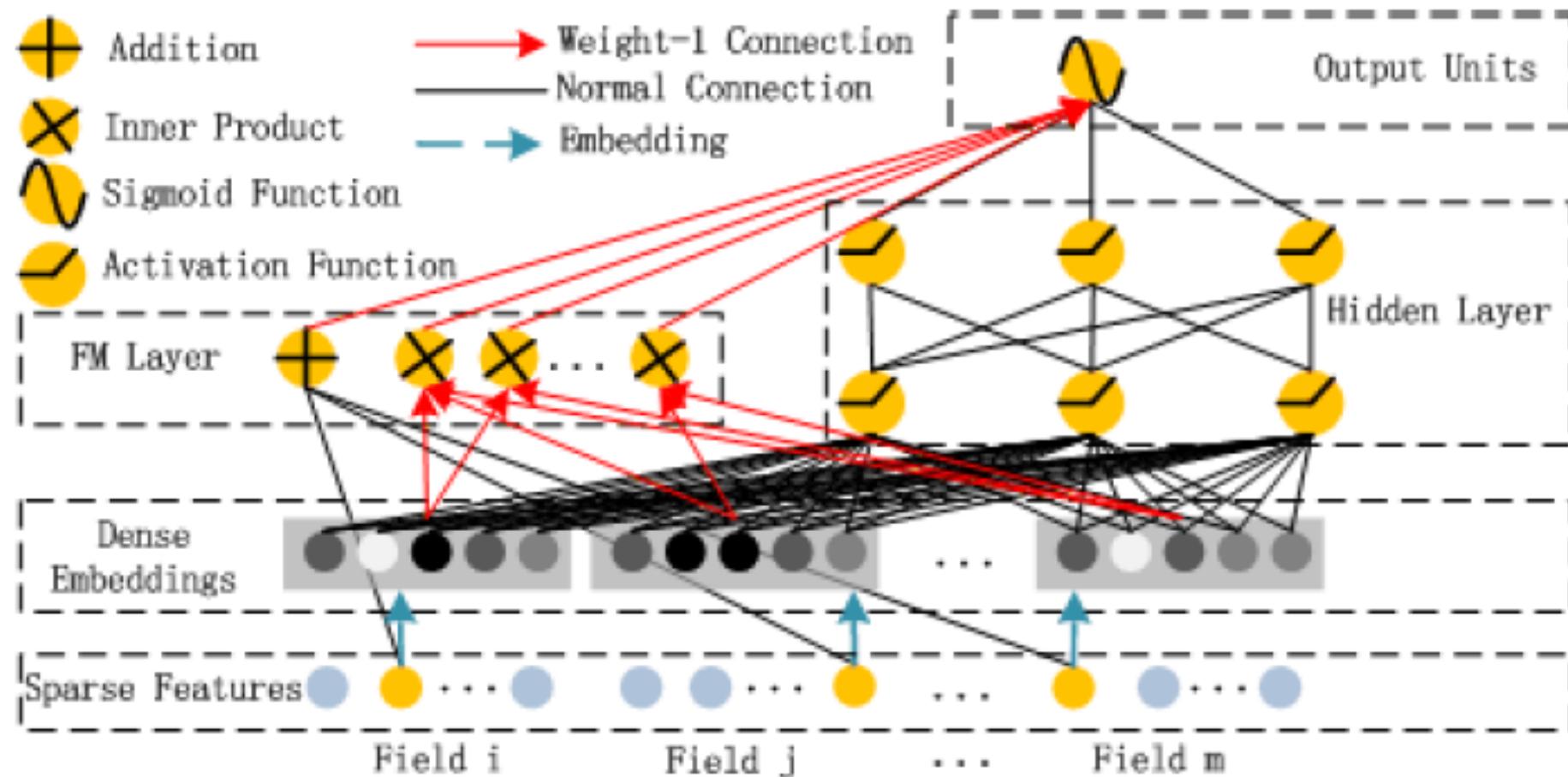
Factorization Machines

$$\hat{y}(x) = w_0 + \sum_{j=1}^p w_j x_j + \sum_{j=1}^p \sum_{j'=j+1}^p x_j x_{j'} \mathbf{v}_j^T \mathbf{v}_{j'}$$

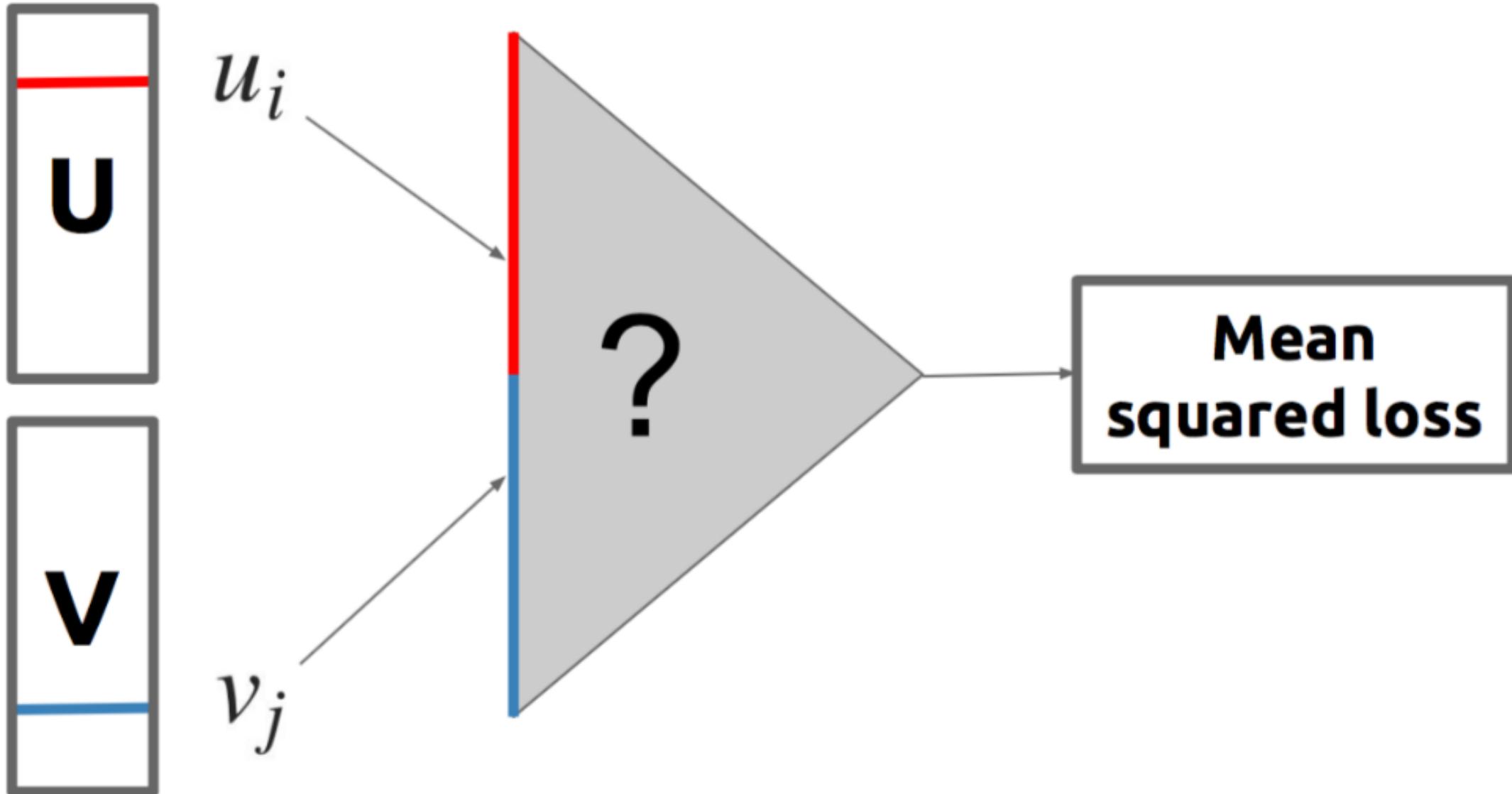
	Feature vector \mathbf{x}																				Target y	
\mathbf{x}_1	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	y_1
\mathbf{x}_2	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	y_2
\mathbf{x}_3	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	y_3
\mathbf{x}_4	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	y_4
\mathbf{x}_5	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	y_5
\mathbf{x}_6	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	y_6
\mathbf{x}_7	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	y_7
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...	Last Movie rated	
	User				Movie					Other Movies rated												

Factorization Machines as Neural Network

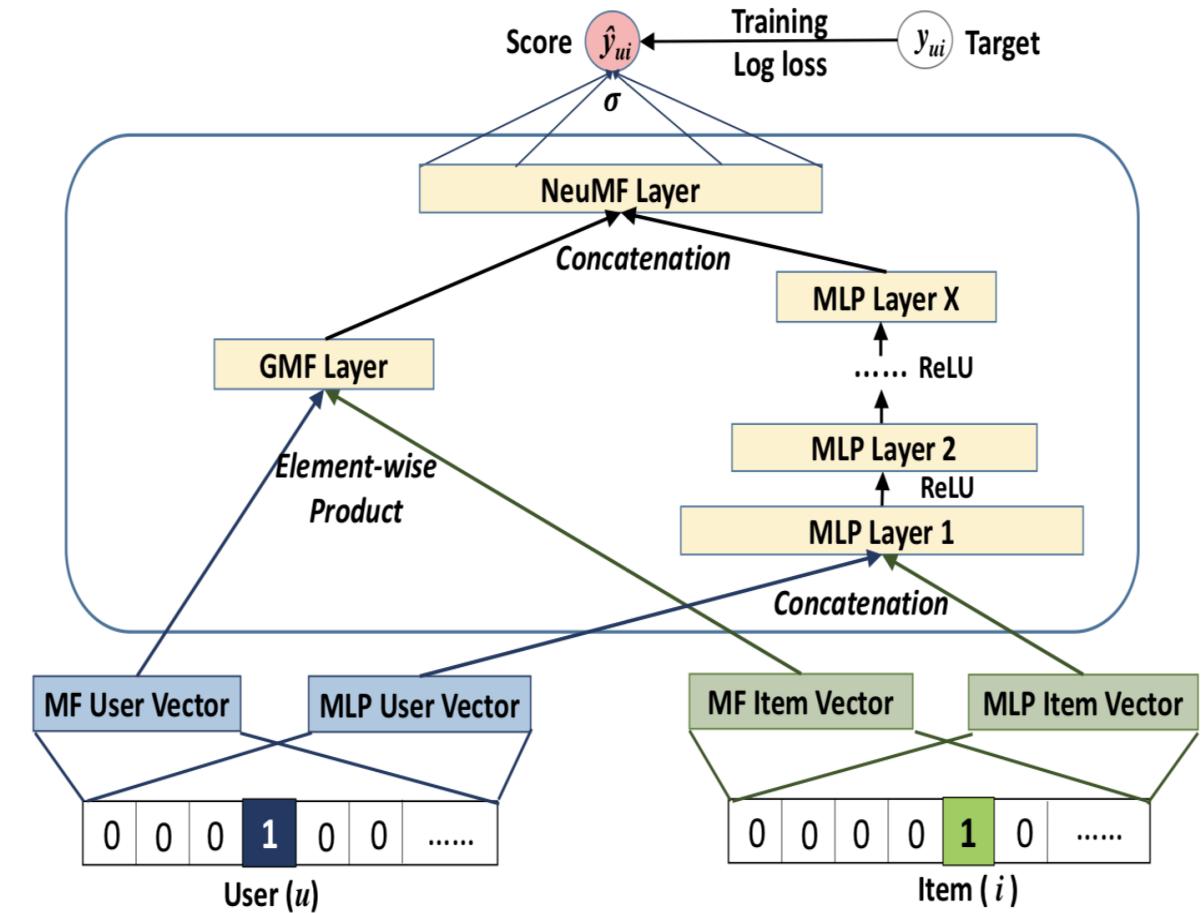
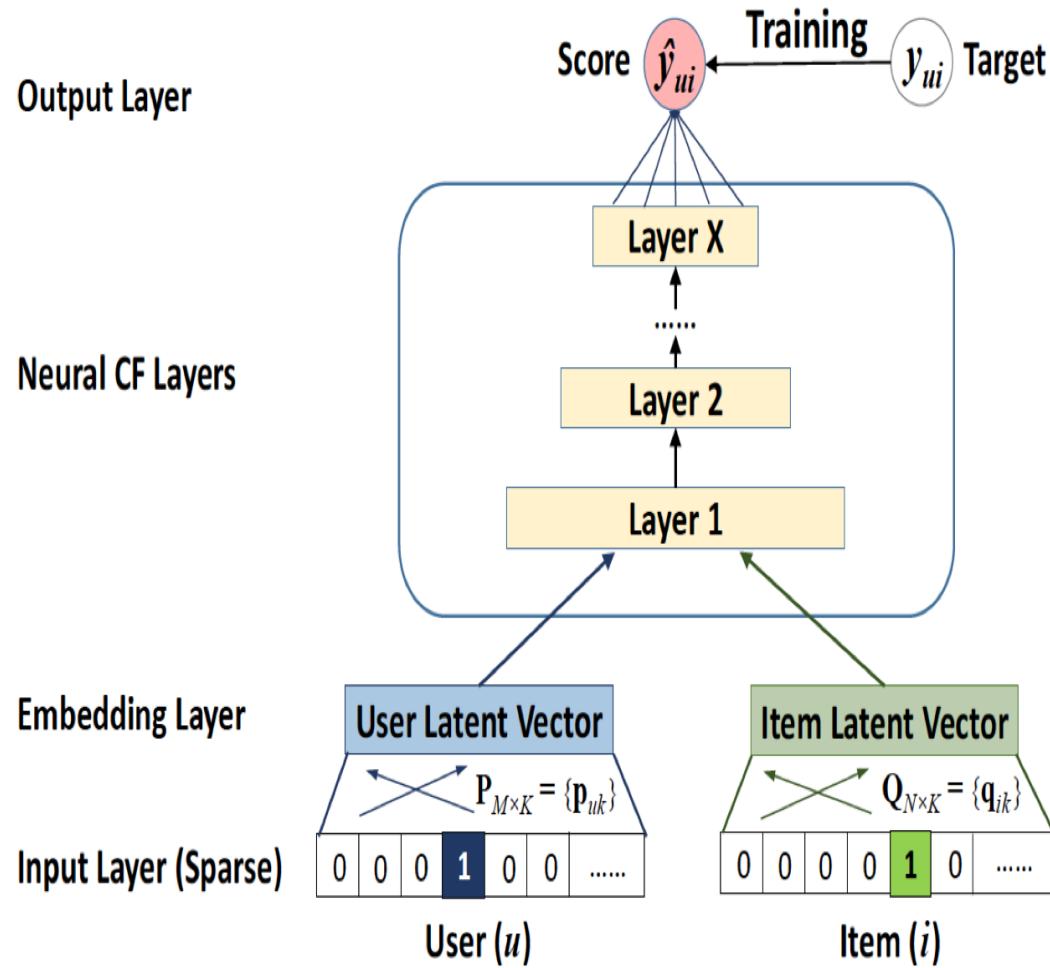
@ mail.ru
group



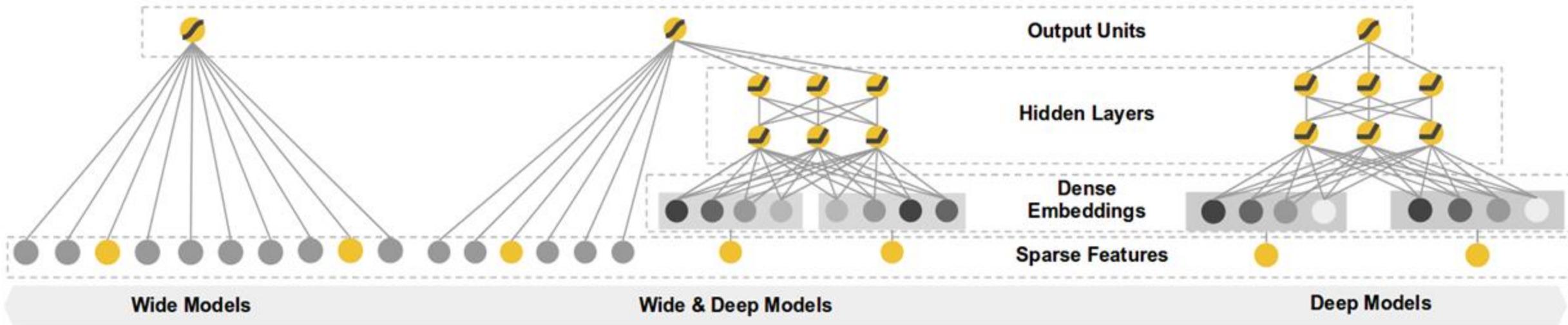
A deeper view



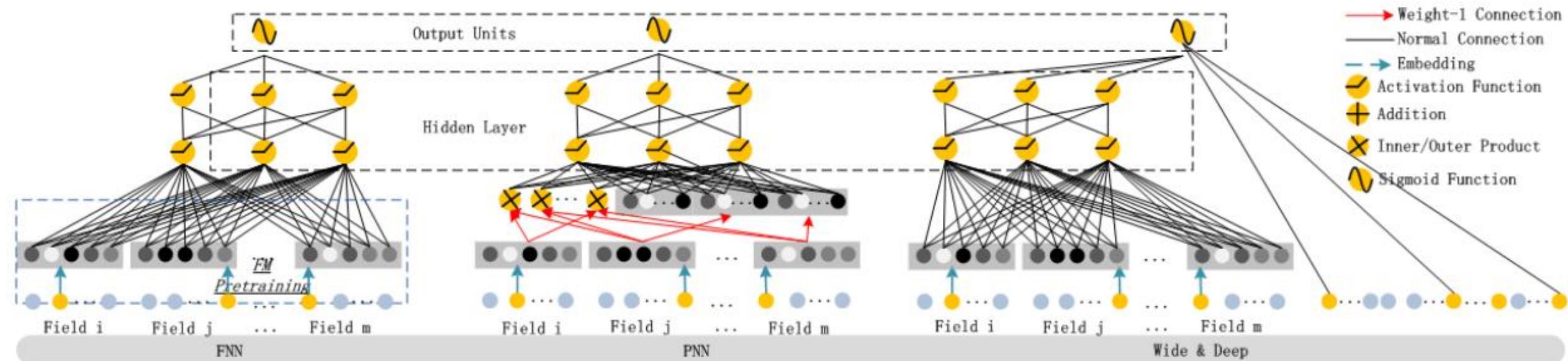
Neural collaborative filtering



Wide and Deep (Google 2016)



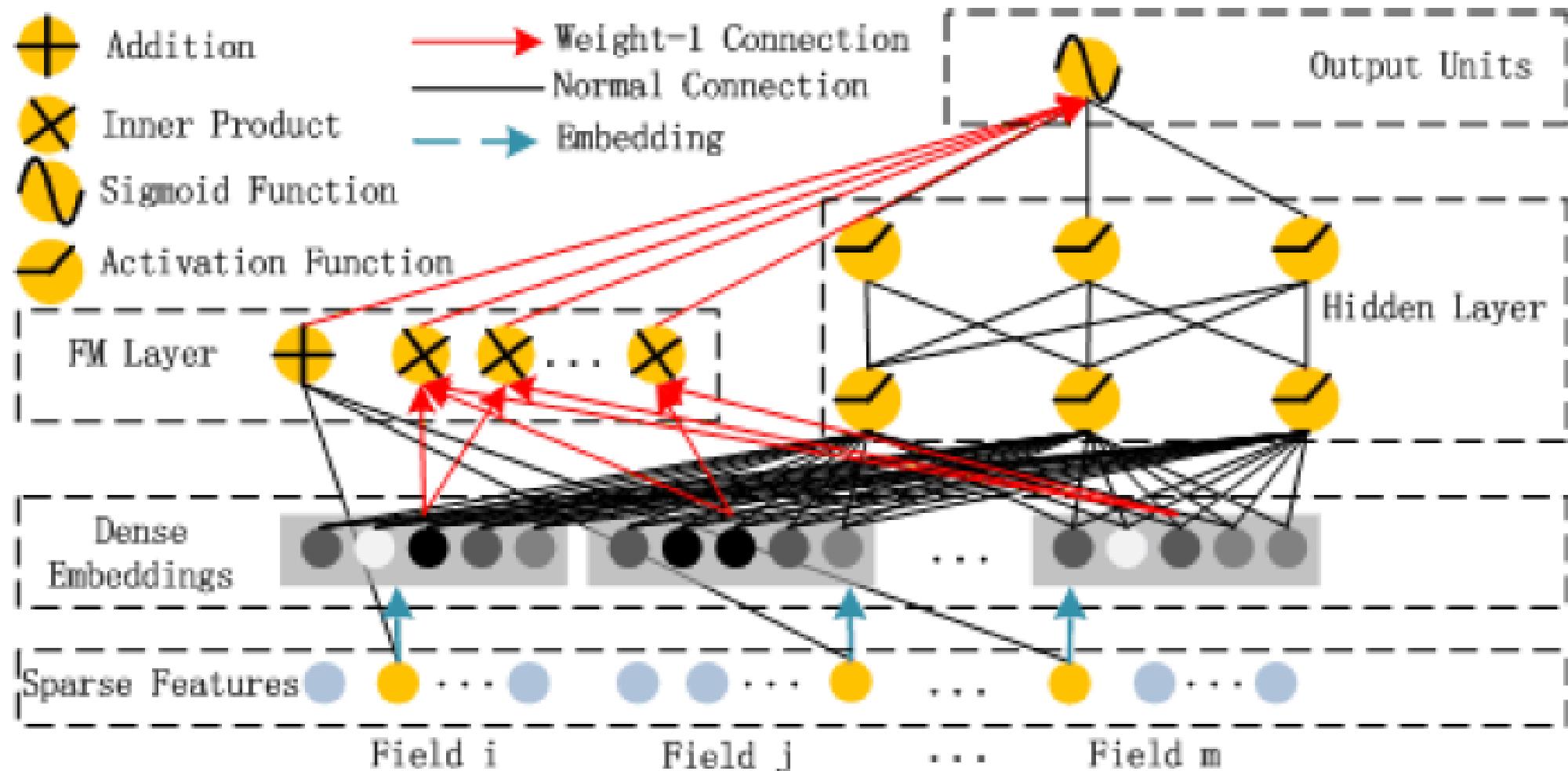
Нейросетевые архитектуры для CF



FNN - FM initialized feed forward neural network

PNN - Product based neural network

DeepFM (2017)



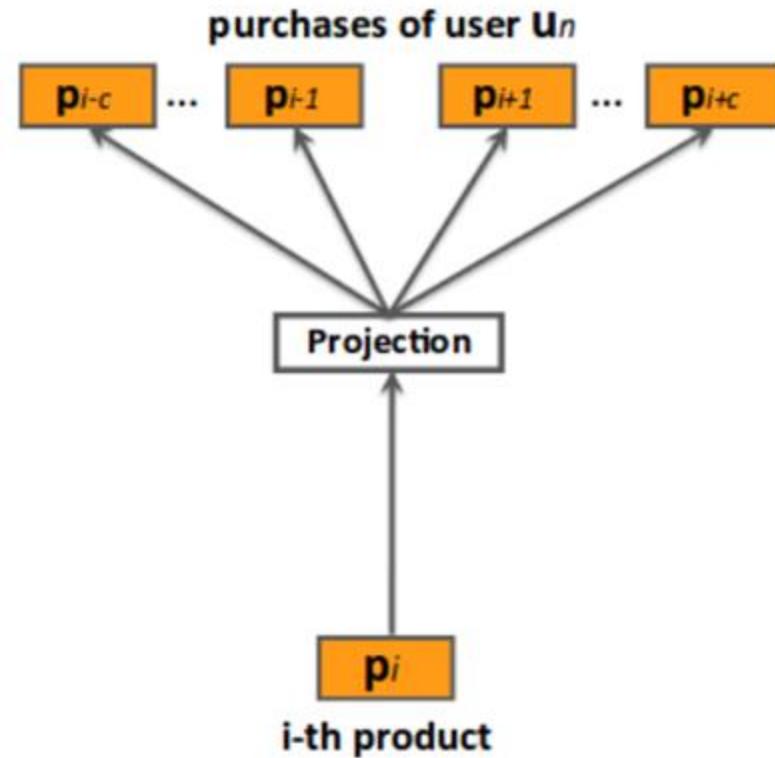
DeepFM (2017)

Результаты:

	Company*		Criteo	
	AUC	LogLoss	AUC	LogLoss
LR	0.8640	0.02648	0.7686	0.47762
FM	0.8678	0.02633	0.7892	0.46077
FNN	0.8683	0.02629	0.7963	0.45738
IPNN	0.8664	0.02637	0.7972	0.45323
OPNN	0.8658	0.02641	0.7982	0.45256
PNN*	0.8672	0.02636	0.7987	0.45214
LR & DNN	0.8673	0.02634	0.7981	0.46772
FM & DNN	0.8661	0.02640	0.7850	0.45382
DeepFM	0.8715	0.02618	0.8007	0.45083

Prod2Vec and Item2Vec

- › Prod2Vec & item2vec:
факторизация матрицы
совместной покупки
товаров
- › Аналог word2vec за
исключением
игнорирования
пространственной
составляющей



Prod2vec skip-gram model

$$\frac{1}{K} \sum_{i=1}^K \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{i+j} | w_i) \rightarrow \frac{1}{K} \sum_{i=1}^K \sum_{j \neq i}^K \log p(w_j | w_i)$$

Pandao2Vec

- › Вектора Pandao2Vec отражают семантику
- › Для них даже (иногда) работают аналогии:



Чехол
для
iPhone

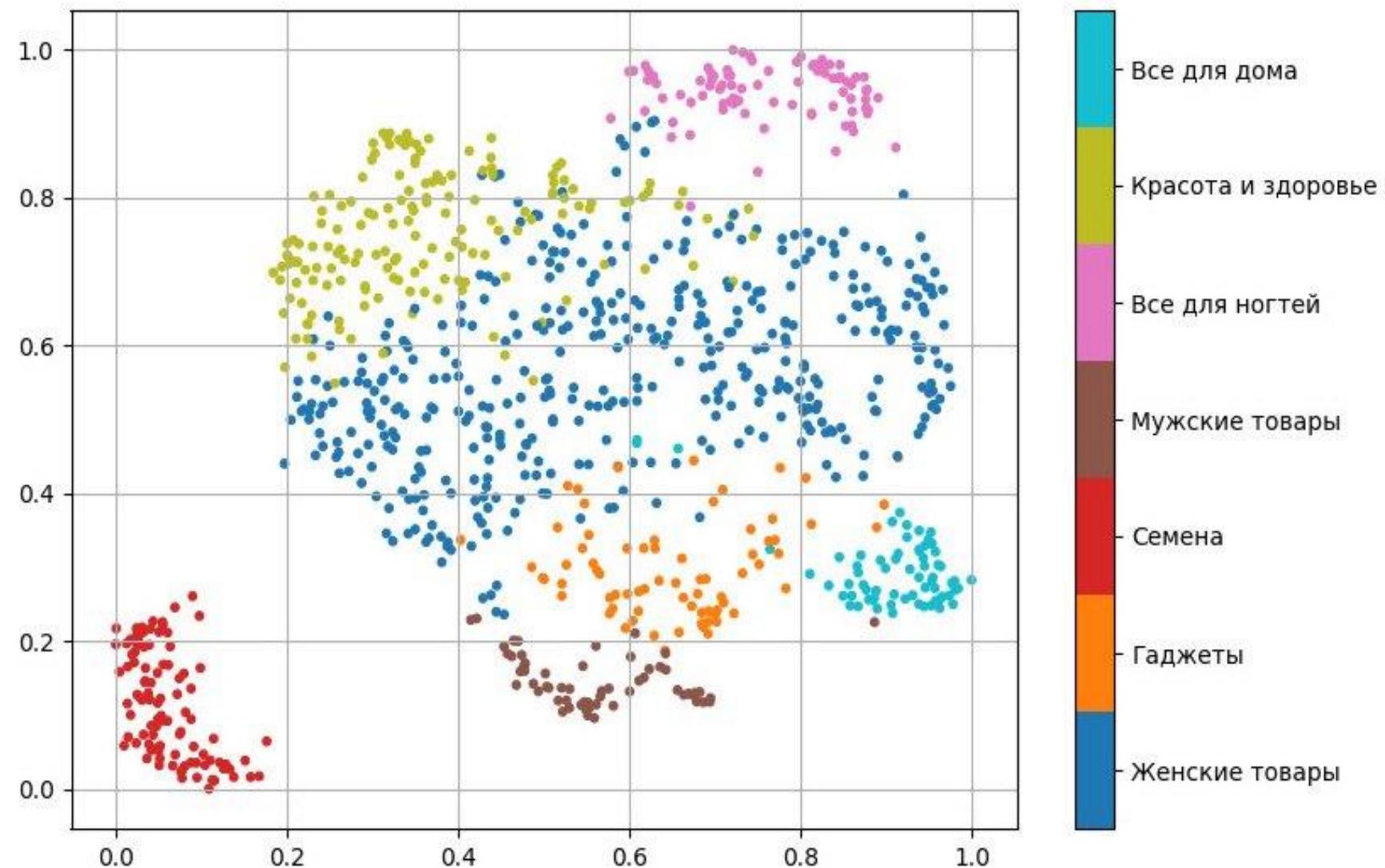
Стекло
для
iPhone

Стекло
для
Samsung

Чехол
для
Samsung

Pandao2Vec

Топ 1000 товаров



Выводы

Neighborhood methods

- Проблема холодного старта
- Матрица оценок используется непосредственно для получения предсказаний
- У больших компаний с большим количеством данных чаще всего реализуется item-to-item подход

Выводы

Model-based methods

- Основаны на оффлайн расчетах
- В продакшене используются уже обученные модели
- Модели необходимо обновлять и перестраивать

Можно применять комбинированный подход

Все это можно использовать в композициях

Открытые вопросы

- Как улучшить качество рекомендаций, а не RMSE?
- Как обосновать рекомендации?
- Как решать проблему холодного старта?

Открытые вопросы

- Как учитывать неявные предпочтения?
- Как учитывать контекст в рекомендациях?
- Как дополнительно учитывать признаки объектов?
- Как учитывать связи между пользователями?
- Как учитывать информацию из других источников (например из поиска)?



Технические вопросы



- Как быстро обновлять рекомендации?
- Как масштабировать расчет рекомендаций?
- Как быстро выбирать топ документов для предсказаний?
- Как измерять качество контекстных рекомендаций?