



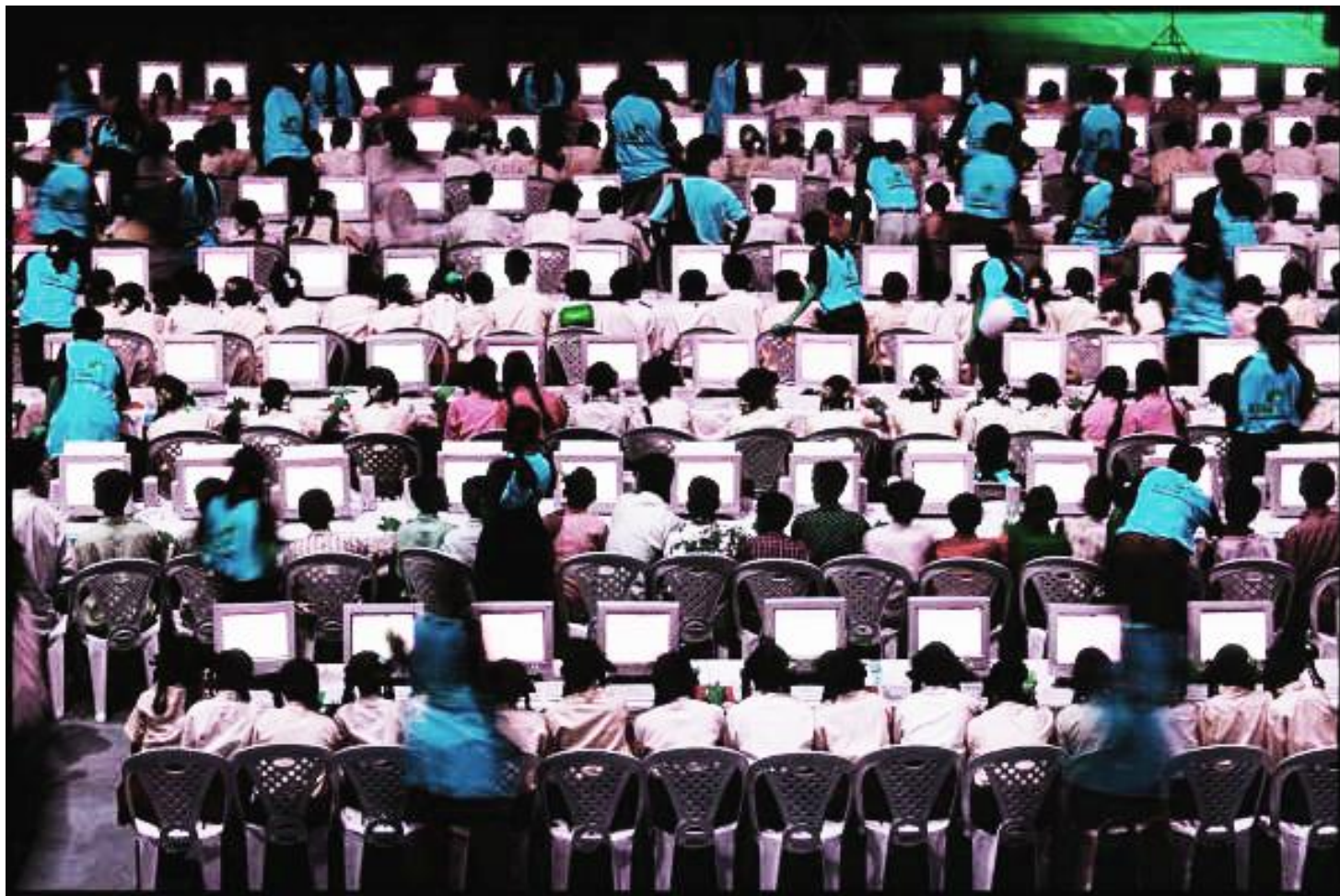
# ТЕХНОСФЕРА

## Введение в информационный поиск

Сергукова Юлия,  
программист отдела инфраструктуры проекта  
Поиск@Mail.Ru

# Что такое Поиск?

# Что такое Поиск?





# Что такое Поиск?



# Поиск@Mail.Ru

@поиск

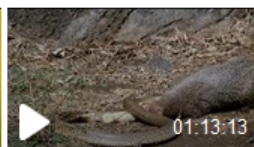
Интернет   Картинки   Видео   Новости   Ответы

## Видео по запросу «рики-тики-тави»

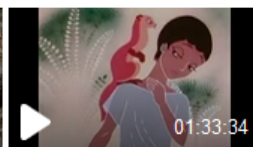
3 тыс. результатов



**Рикки-Тикки-Тави**  
youtube  
28 мая 2012



**РИКИ. ТИКИ. ТАВИ.** 1975 г. СССР  
ok.ru  
7 сентября 2015



Сборник Мультфильмов  
- **Рики-Тики-Тави**  
mail.ru  
6 октября 2015



**Рики-Тики-Тави**  
Совет  
youtube  
31 мар

## Рикки-Тикки-Тави — Youtube.com

[youtube.com/watch?v=ar-lv03EIXI](https://www.youtube.com/watch?v=ar-lv03EIXI)

Мультки студии Союзмультфильм, 3 526 966 просмотров

Мультки : **Рикки тикки тави**. Советский мультфильм для детей по рассказу Р. Киплинга.

Нажми лайк, если любишь советские мультфильмы. Добрые...

## Сказка Рикки Тикки Тави читать

[deti-online.com/skazki/skazki-...](https://deti-online.com/skazki/skazki-...)

**Рикки Тикки Тави** - сказка Киплинга. **Рикки Тикки Тави** читать онлайн или скачать в формате doc или pdf.

## Рикки-Тикки-Тави



«Рикки-Тикки-Тави» — рассказ из «Книги джунглей» Редьярда Киплинга, а также имя главного героя этого рассказа. [Википедия](#)

Год первого издания: 1893

Язык оригинала: английский язык

Авторы: Редьярд Киплинг

[Сообщить об ошибке](#)   [Википедия](#)   [Wikidata](#)

# Поиск@Mail.Ru

- Текстовый поиск
- Картинки
- Видео
- Новости
- Ответы
- Карточка «быстрого ответа»

а еще:

- Приложения
- Обсуждения
- и многое другое





# Что такое информационный поиск

Информацио́нный по́иск — процесс поиска неструктурированной документальной информации, удовлетворяющей информационные потребности пользователя



# Информационный поиск

## Наука

- математические модели
- вычислительная лингвистика
- статистика и теорвер
- машинное обучение

## Практика

- огромный объем данных
- KISS («keep it simple, stupid»)
- хаки практической реализации

# Цель

**«удовлетворение информационной потребности  
пользователя»**

# Цель

**«удовлетворение информационной потребности пользователя»**

У пользователя есть потребность в получении определенной информации

Цель – найти **релевантную** его запросу информацию

# Входные данные

- **корпус документов** – набор документов (текстов, изображений, новостных блоков и т.д.)
- **ссылочная информация** – PageRank
- **обратная связь** – запросы пользователей и клики в выдачу



# Пример

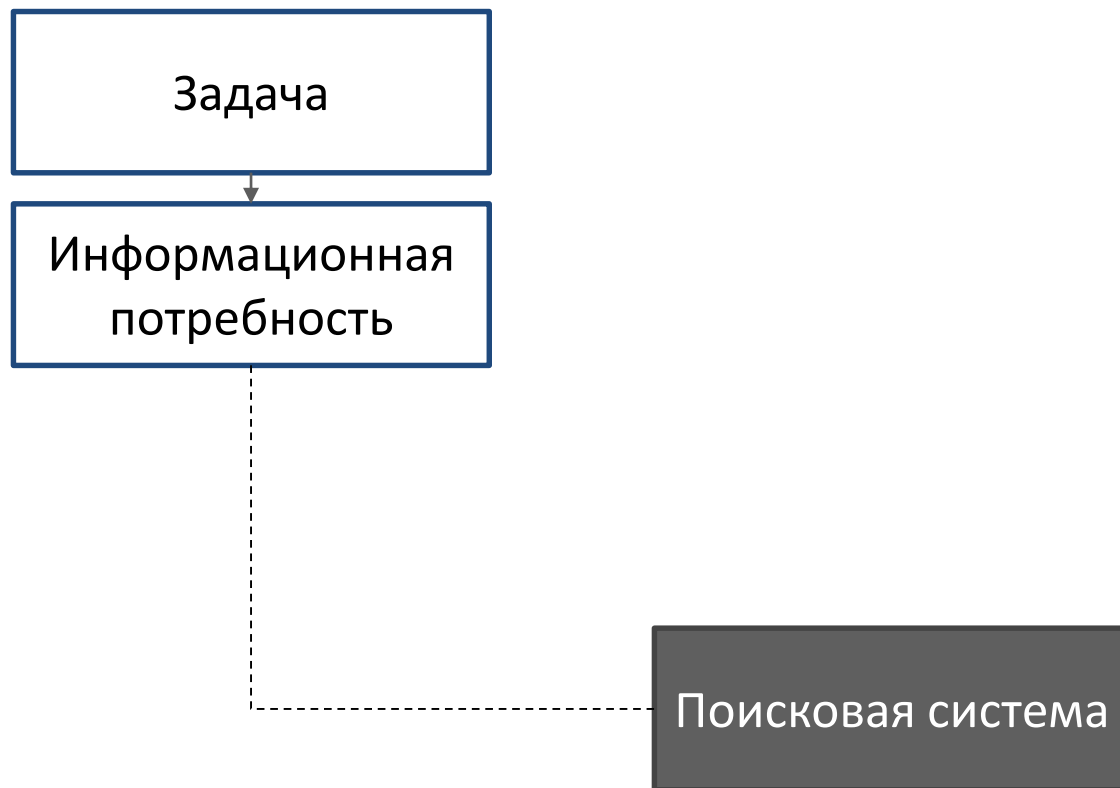
Задача

Поисковая система

Смерть («Плоский мир» Т.Пратчетта)



# Пример



Смерть («Плоский  
мир» Т.Пратчетта)

# Пример



# Пример

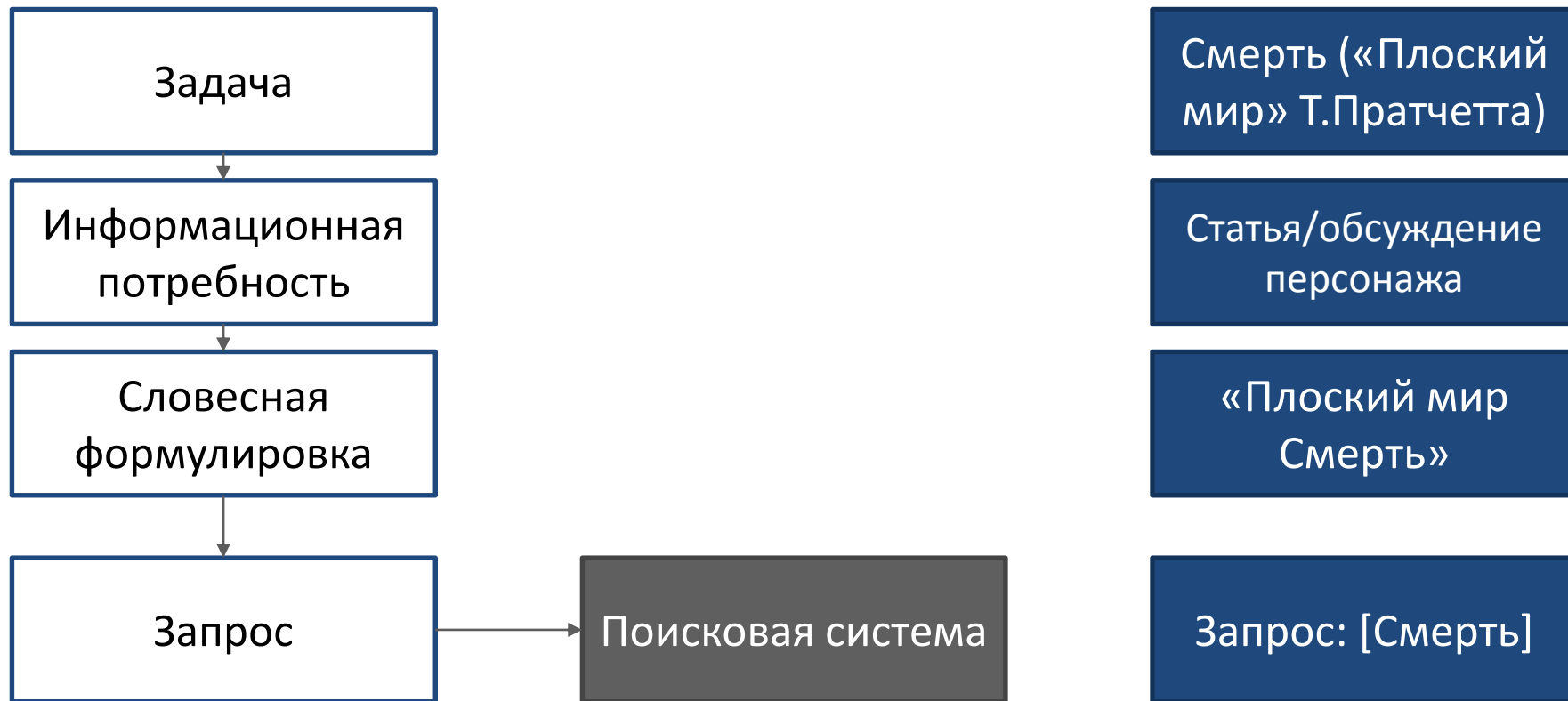




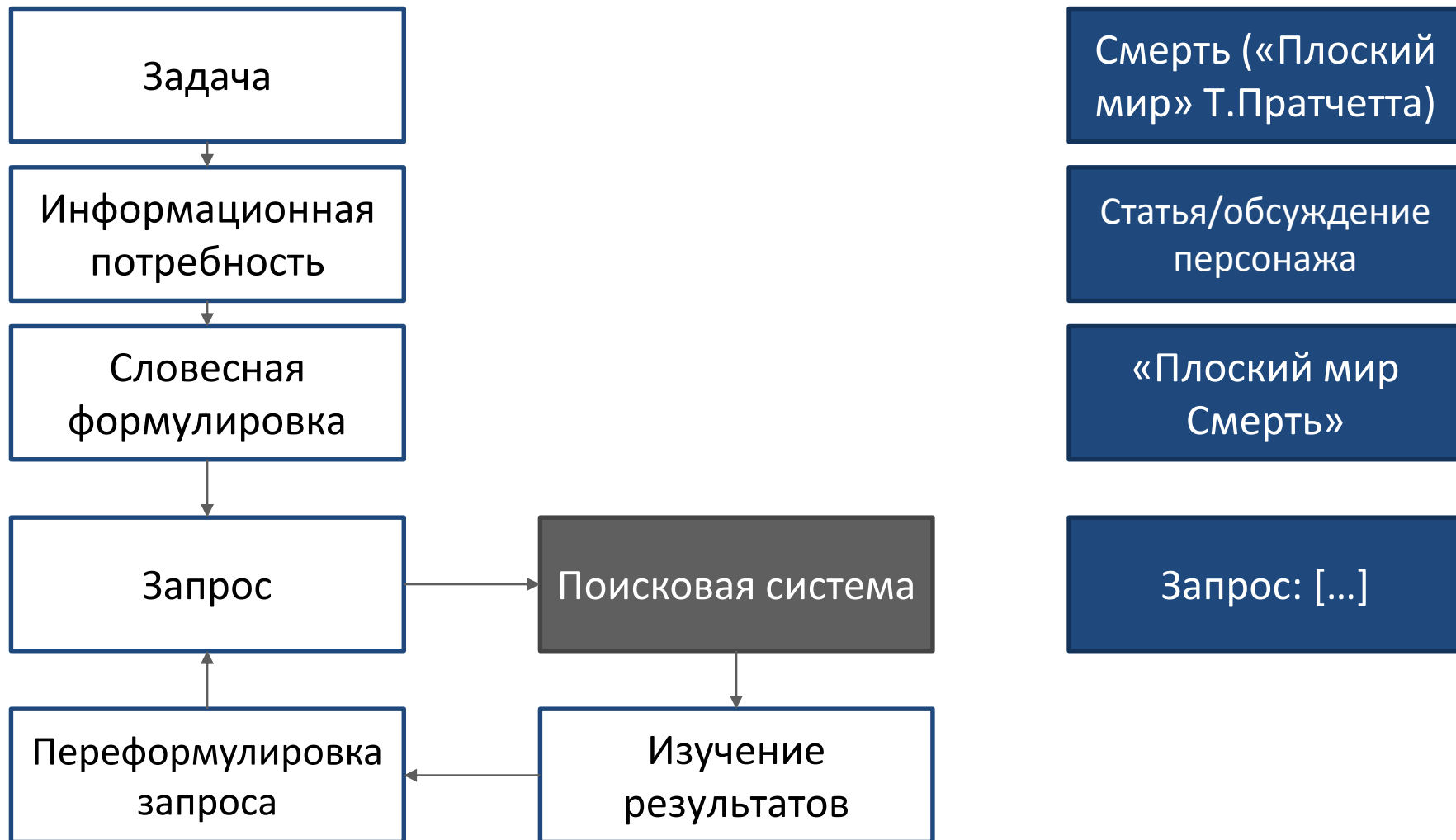
# Пример



# Пример



# Пример



# Что может пойти не так?



# Что может пойти не так?

- Не поняли пользователя
  - Синонимы
  - Опечатки
  - Несодержательный запрос
- Нашли не то
  - Омонимы
  - Нет знаем нужный документ (непопулярные запросы)
  - Пользователь не согласен с результатами



# Сложности обработки запросов от людей

- лингвистика: что написано в тексте
- запросы: что хотел пользователь
- статистика
- машинное обучение
- огромные объемы данных
- большие нагрузки
- пользовательский интерфейс
- форматы данных

# Связанные задачи

- рекомендательные системы
- машинный перевод
- извлечение мнений
- распознавание и синтез речи
- семантический анализ текста
- диалог с пользователем
- и т.д.



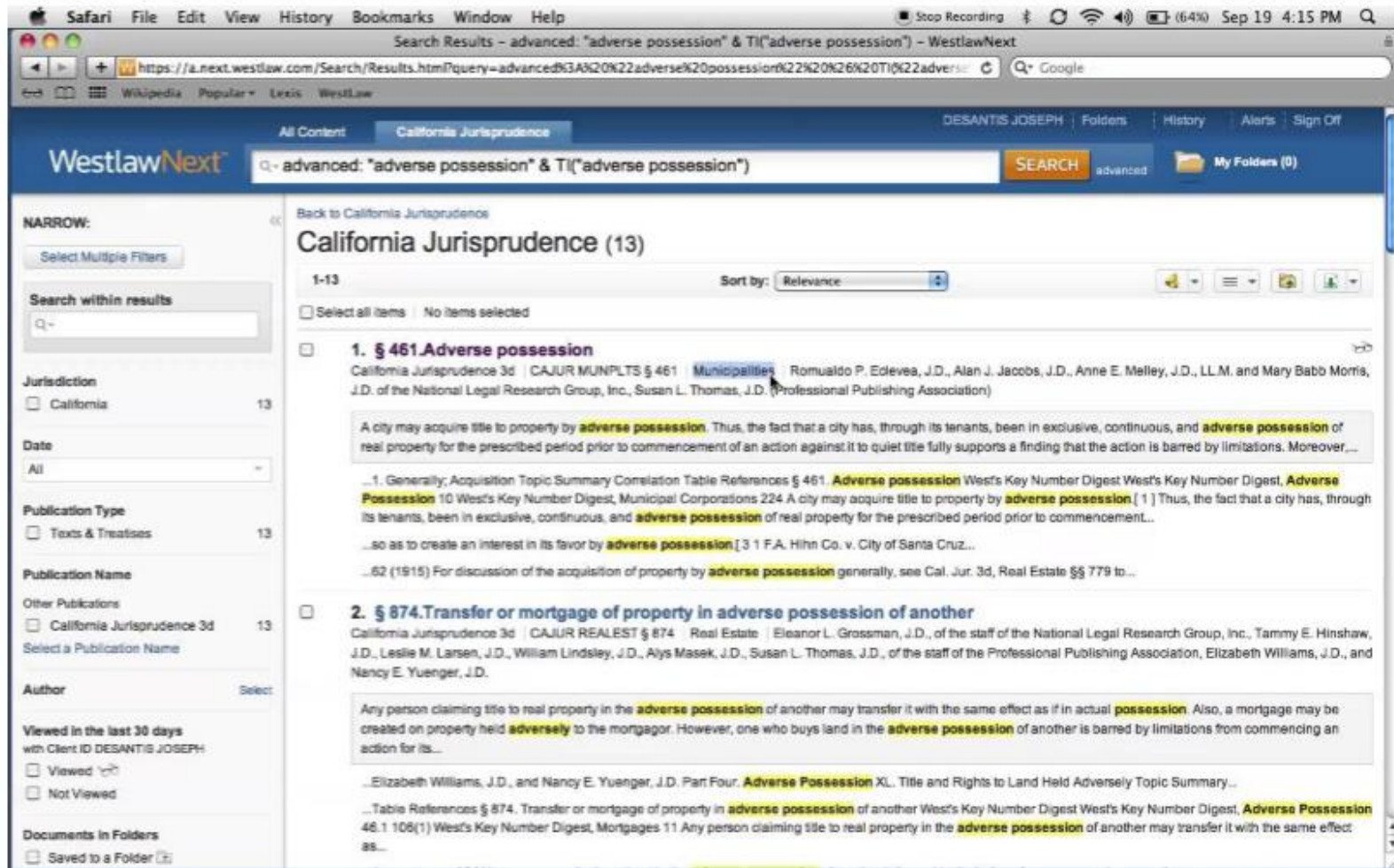
# Примеры поисковых систем



# Южная Корея - Naver



# Westlaw



The screenshot shows a Safari browser window displaying the WestlawNext search results page. The search query is "advanced: 'adverse possession' & T('adverse possession')". The results are filtered to "California Jurisprudence" and show 13 items. The first two results are highlighted:

- § 461. Adverse possession**  
 California Jurisprudence 3d | CAJUR MUNPLTS § 461 | **Municipalities** | Romualdo P. Edevea, J.D., Alan J. Jacobs, J.D., Anne E. Melley, J.D., LL.M. and Mary Babb Morris, J.D. of the National Legal Research Group, Inc., Susan L. Thomas, J.D., of the Professional Publishing Association)  
 A city may acquire title to property by **adverse possession**. Thus, the fact that a city has, through its tenants, been in exclusive, continuous, and **adverse possession** of real property for the prescribed period prior to commencement of an action against it to quiet title fully supports a finding that the action is barred by limitations. Moreover, ...  
 ...1. Generally; Acquisition Topic Summary Correlation Table References § 461. **Adverse possession** West's Key Number Digest West's Key Number Digest, **Adverse Possession** 10 West's Key Number Digest, Municipal Corporations 224 A city may acquire title to property by **adverse possession** [ 1 ] Thus, the fact that a city has, through its tenants, been in exclusive, continuous, and **adverse possession** of real property for the prescribed period prior to commencement...  
 ...so as to create an interest in its favor by **adverse possession** [ 3 1 F.A. Hihn Co. v. City of Santa Cruz...  
 ...62 (1915) For discussion of the acquisition of property by **adverse possession** generally, see Cal. Jur. 3d, Real Estate §§ 779 to...
- § 874. Transfer or mortgage of property in adverse possession of another**  
 California Jurisprudence 3d | CAJUR REALEST § 874 | Real Estate | Eleanor L. Grossman, J.D., of the staff of the National Legal Research Group, Inc., Tammy E. Hinshaw, J.D., Leslie M. Larsen, J.D., William Lindsley, J.D., Alys Masek, J.D., Susan L. Thomas, J.D., of the staff of the Professional Publishing Association, Elizabeth Williams, J.D., and Nancy E. Yuenger, J.D.  
 Any person claiming title to real property in the **adverse possession** of another may transfer it with the same effect as if in actual **possession**. Also, a mortgage may be created on property held **adversely** to the mortgagor. However, one who buys land in the **adverse possession** of another is barred by limitations from commencing an action for its...  
 ...Elizabeth Williams, J.D., and Nancy E. Yuenger, J.D. Part Four. **Adverse Possession** XL. Title and Rights to Land Held Adversely Topic Summary...  
 ...Table References § 874. Transfer or mortgage of property in **adverse possession** of another West's Key Number Digest West's Key Number Digest, **Adverse Possession** 46.1 106(1) West's Key Number Digest, Mortgages 11 Any person claiming title to real property in the **adverse possession** of another may transfer it with the same effect 88...

The left sidebar contains filters for Jurisdiction (California), Date (All), Publication Type (Texts & Treatises), Publication Name (California Jurisprudence 3d), Author (Select), Viewed in the last 30 days (Client ID DESANTIS JOSEPH), and Documents in Folders (Saved to a Folder).

# Не-текстовый поиск.

## Twenty-two settings

[Download ABC](#)

X: 1  
T: Drowsy Maggie  
R: reel  
M: 4/4  
L: 1/8  
K: Edor  
|:E2BE dEBE|E2BE AFDF|E2BE dEBE|BABc dAFD:|  
d2fd c2ec|defg afge|d2fd c2ec|BABc dAFA|  
d2fd c2ec|defg afge|afge fdec|BABc dAFD|

# Added by *Jeremy* 17 years ago.

[Download](#)[Print](#)

# Контентные рекомендательные системы

Когда Каламу повесили на шею медаль победителя, он решил разделить эту победу с дочкой и отдал награду ей. Малышка даже попробовала медаль на зуб. «Это действительно важный день для меня, и я хочу разделить эту победу с младшей дочерью. То, что нам это удалось, говорит о том, что быть папой и работающим человеком вовсе не значит, что нужно отказываться от своих мечтаний, надежд и хобби», – заявил Калам.

Он рассчитывает стать примером для других семей, показывая, что активный образ жизни на свежем воздухе приносит пользу и детям, и взрослым.

Также смотрите [15 трогательных фото об отцовской любви](#).


Материал подготовила Татьяна Щеглова

11 февраля 2016 [материалы](#) [Все темы](#)


Обновлена в прошлый раз: 10.02.2016, 10:00. Выделите ее и нажмите Ctrl + Enter.

Все новости


### Материалы по теме




Климова рассказала об интересной семейной традиции




Ученые предупредили об опасностях позднего отцовства



Какие подарки получили принцесса Шарлотта с братом




Анна Хильминович назвала дочь в честь себя и мужа




Константин Хабенский скоро снова станет папой

Директ

Системы Аквафор Морион от 5790 р.  
[tek.shop.aquaphor.ru](#)



Рекуператор воздуха квартирный  
[mailey-rus.ru](#)



Приточная вентиляция в квартире  
свежий-воздух.рф

Приточная и приточно-вытяжная вентиляция в квартире. Монтаж. Распродажа без%

Адрес и телефон

## Pandora's "Music Genome Project" explores the cold hard facts of how we interact with music

Twenty-five music analysts "grade" 10,000 songs a month. It's a mountainous job, writes Rob Pegoraro, but the results will be serious business.



# Поиск по структурированной информации и неструктурированным запросам



📄 📷 📊 📡

Web Apps Examples ↻ Random

Assuming "apple" is a food | Use the input as a [formula](#) instead

Assuming apple | Use [prepared apples](#) or [more](#) instead

Assuming any type of apple | Use [apple, with skin](#) or [apple, without skin](#) instead

Input interpretation:

apple	amount	1 apple	mass
-------	--------	---------	------

Result:

182 grams

Show non-metric

Unit conversions:

0.18 kg (kilograms)

0.4 lb (pounds)

6.4 oz (ounces)

Physical properties:

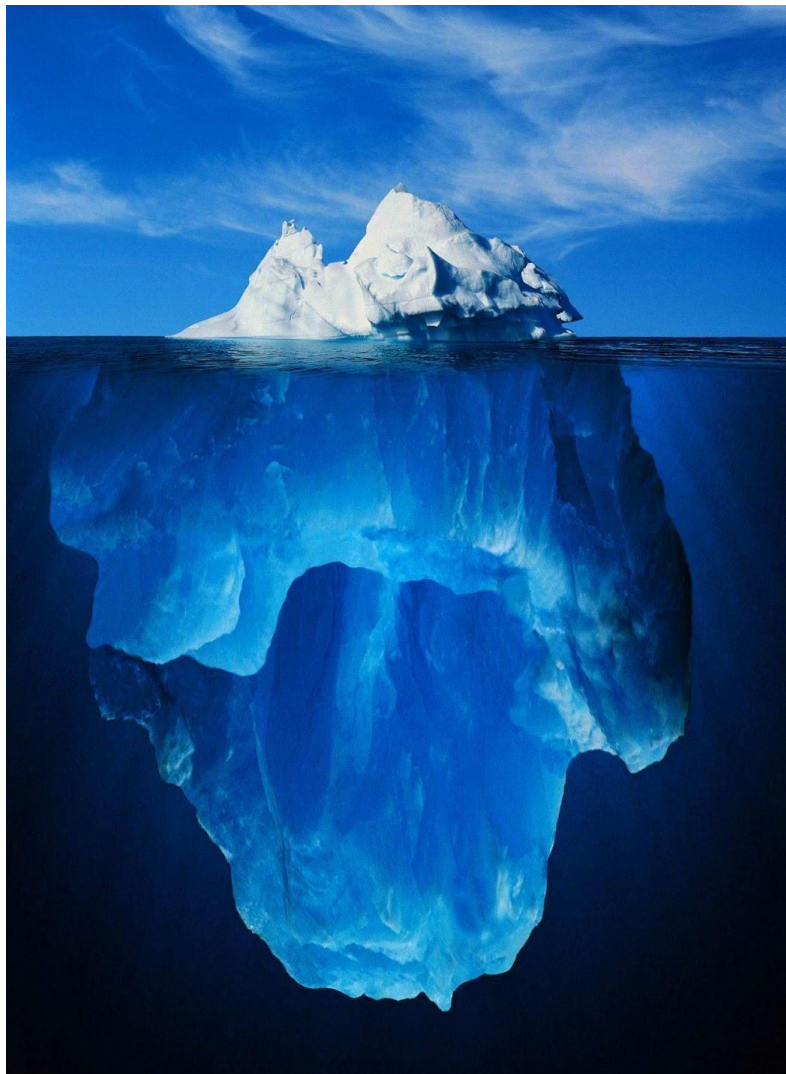
mass	182 grams
serving volume	233 mL (milliliters)

# Обзор курса





# Обзор курса



# Преподаватели



Юлия  
Сергукова



Андрей  
Мурашев



Евгений  
Чернов



Владимир  
Гулин

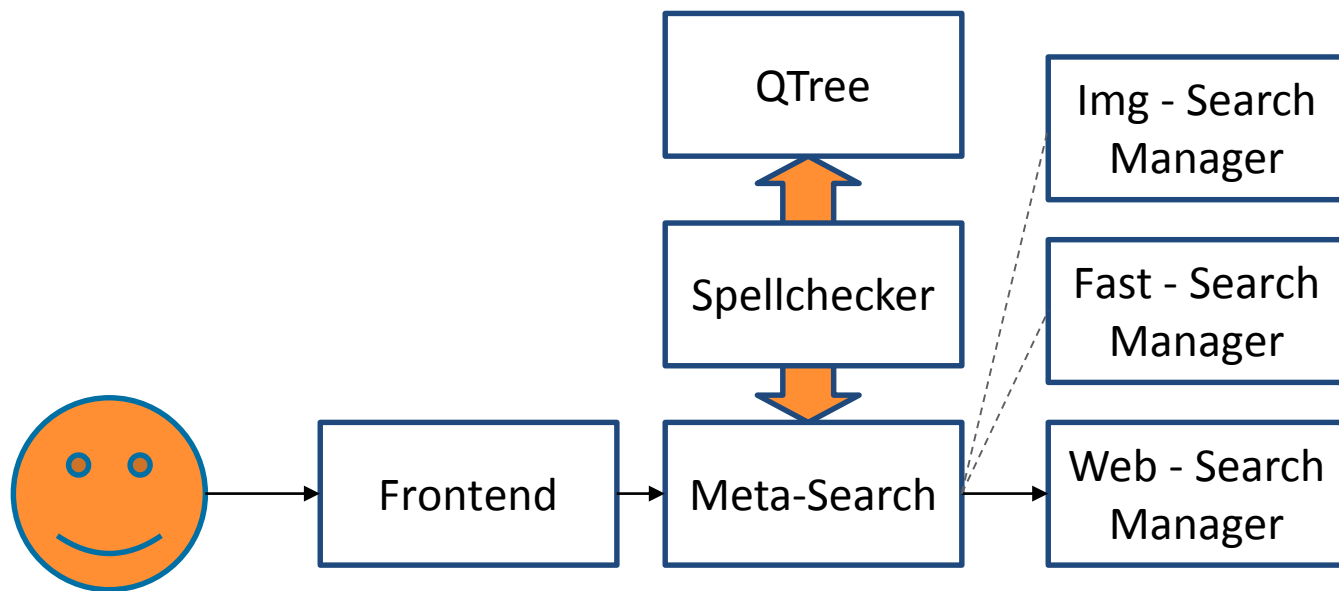


# Как работает поиск

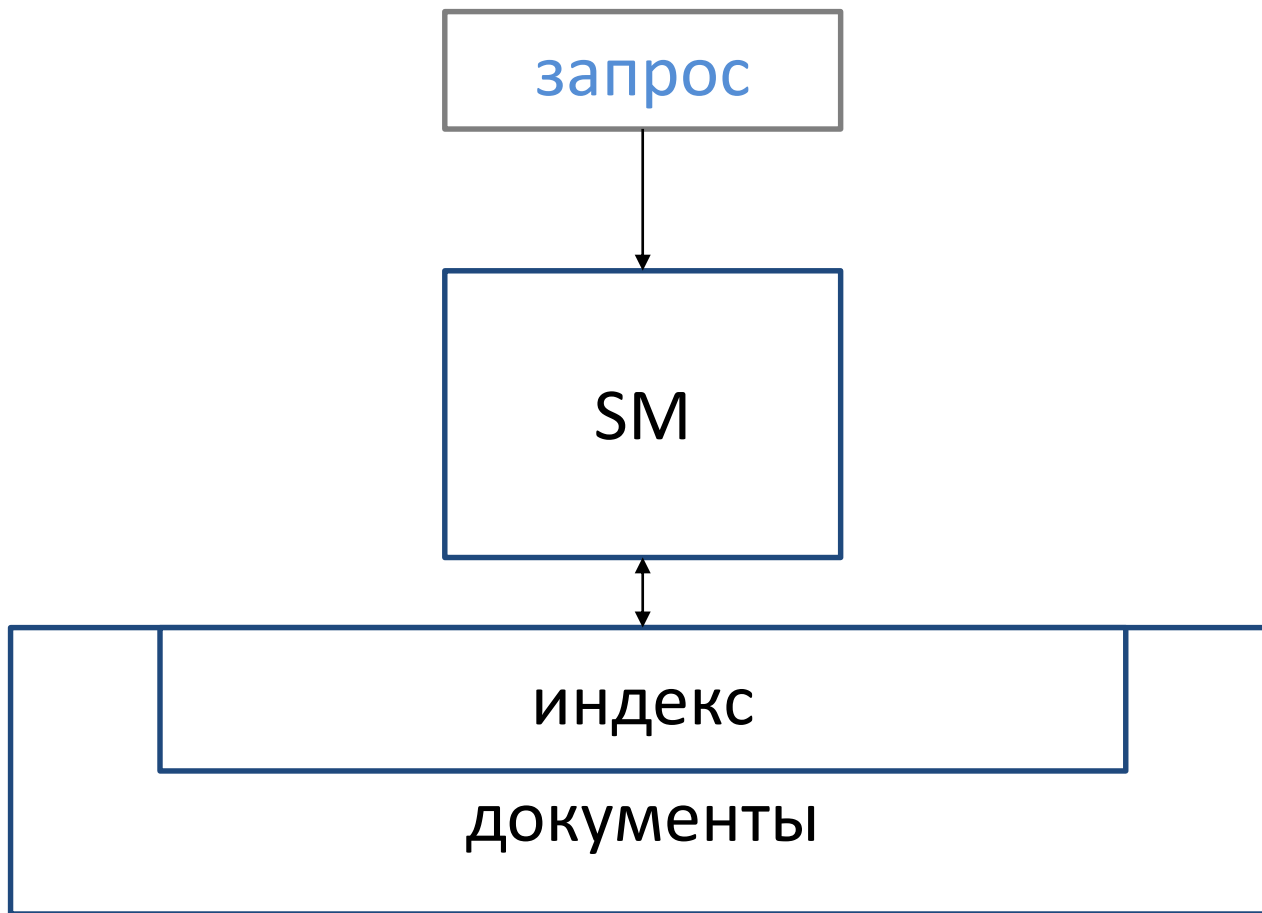
# Как работает поиск



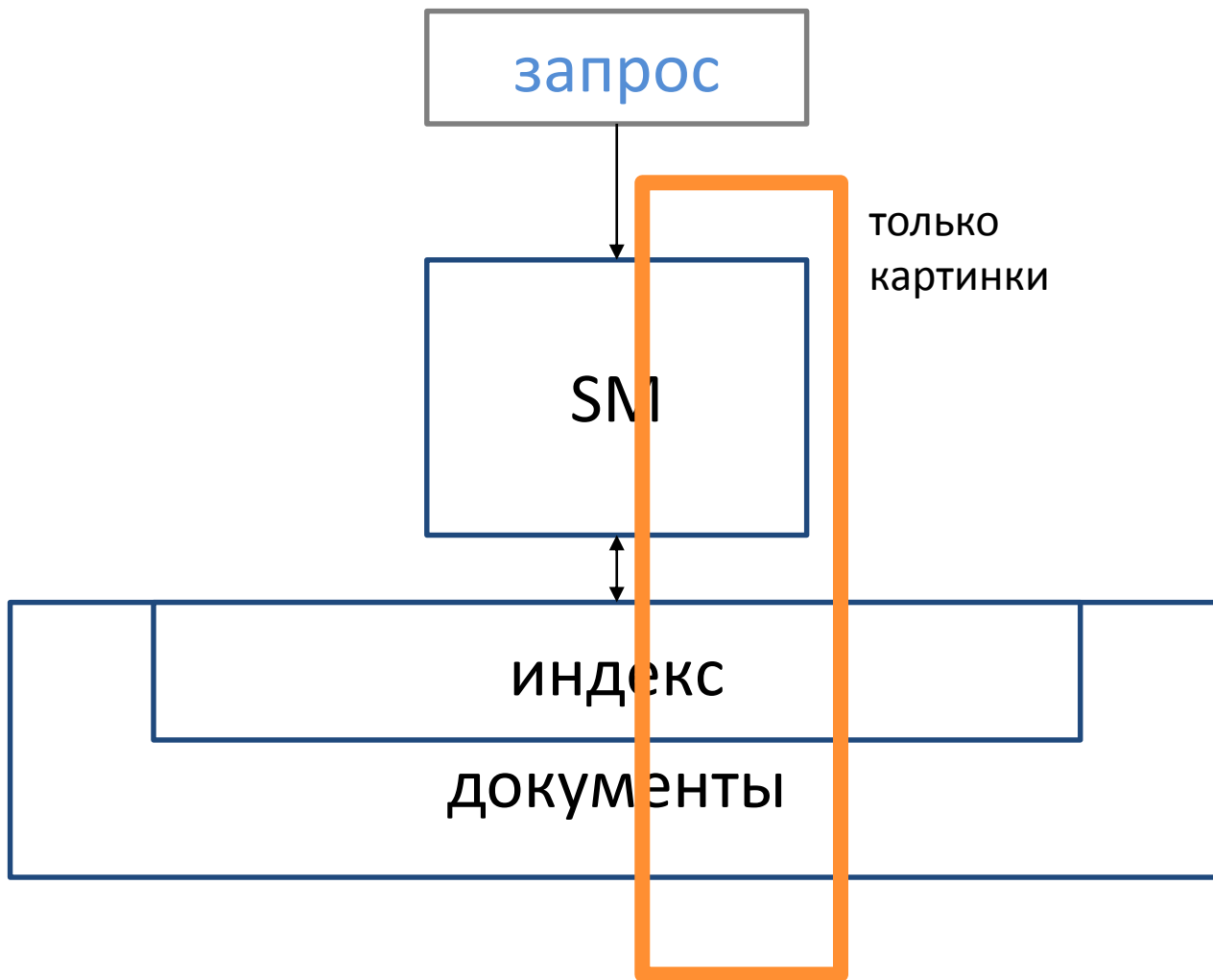
# Как работает поиск



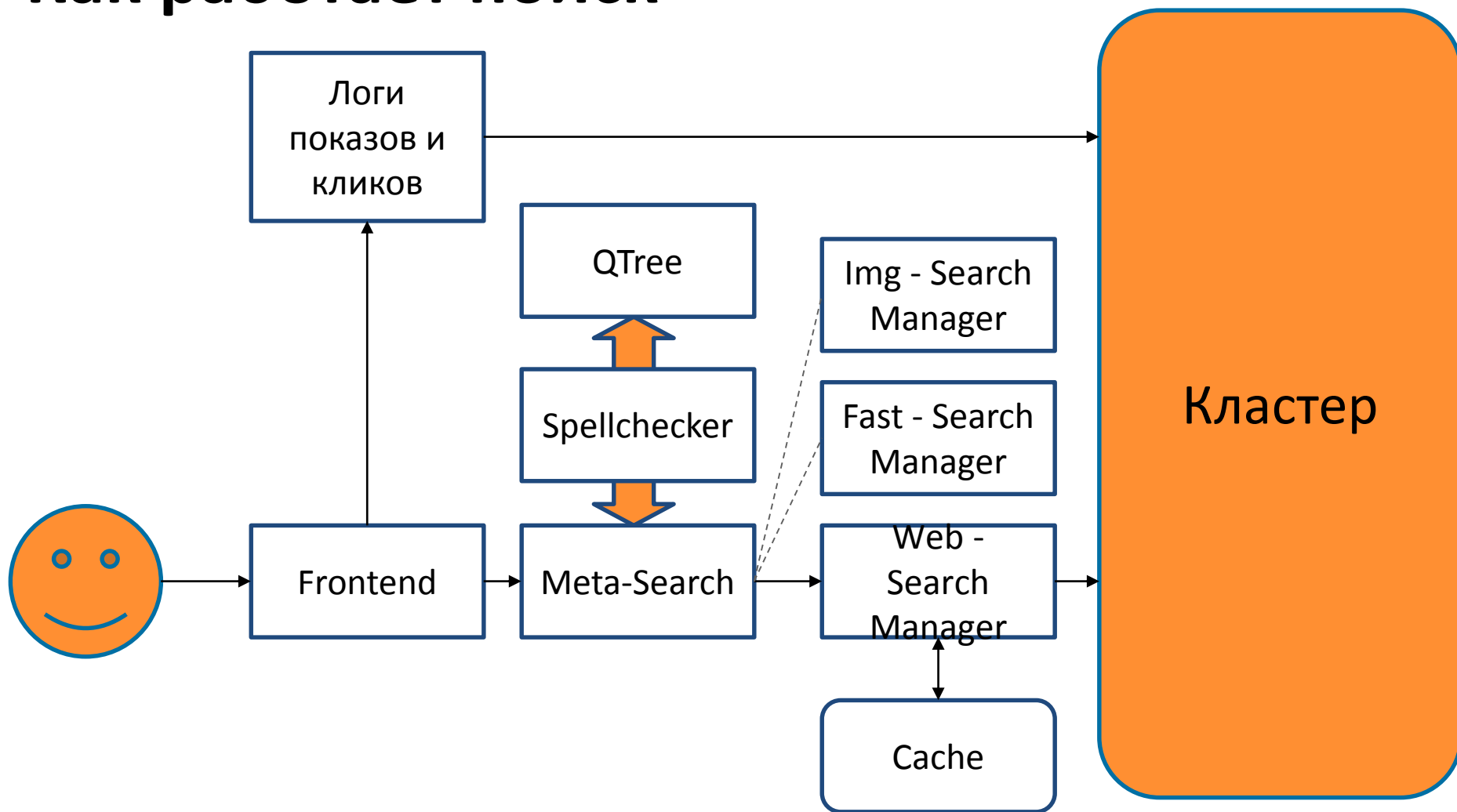
# Немного о вертикалях



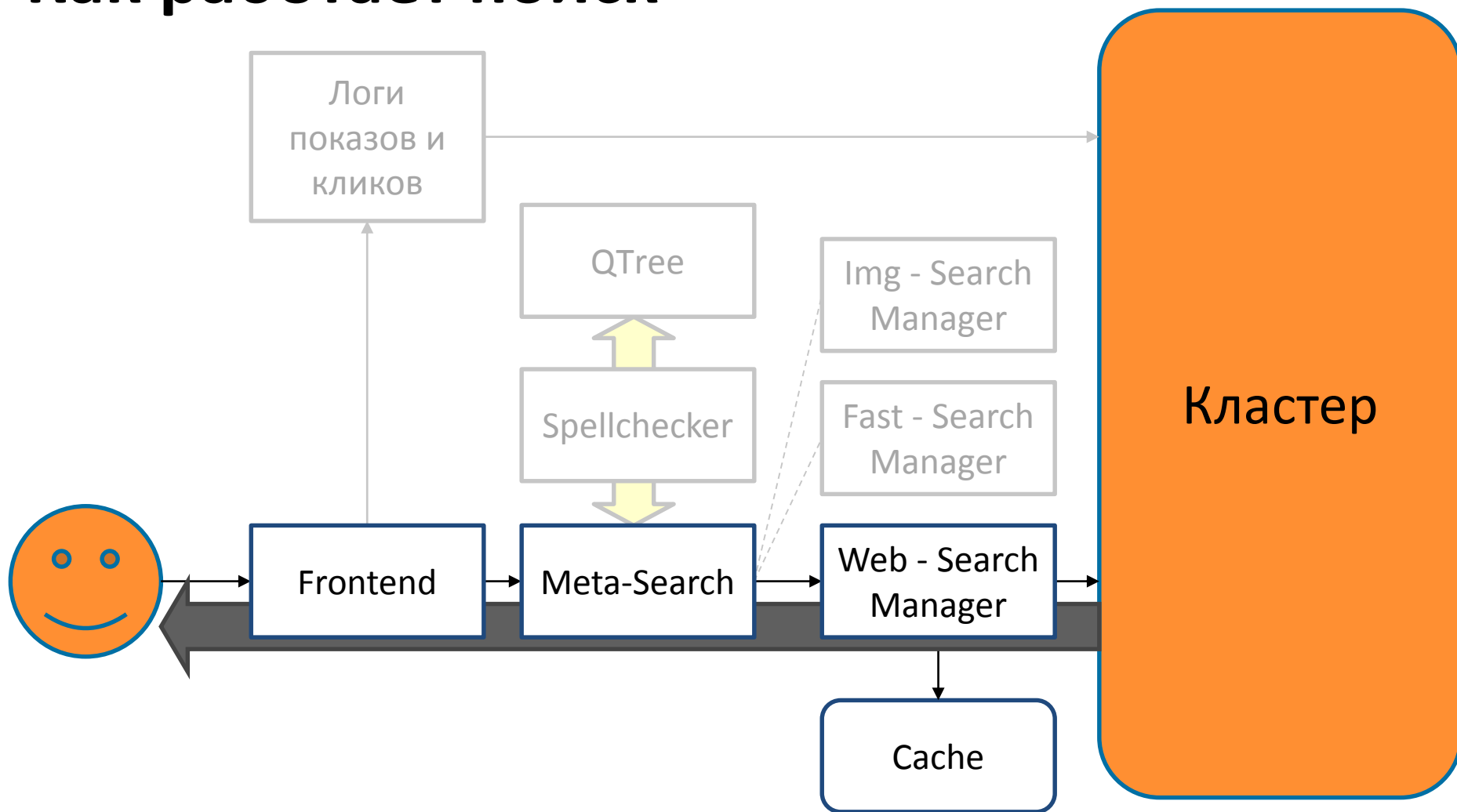
# Немного о вертикалях



# Как работает поиск



# Как работает поиск



# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.



# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.

КОЛЛОКВИУМ!

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.

## О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)
7. Обработка запросов. Опечатки. (ДЗ)

## О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)
7. Обработка запросов. Опечатки. (ДЗ)
8. Классификация запросов. Саджесты.  
Переформулировки.



# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)
7. Обработка запросов. Опечатки. (ДЗ)
8. Классификация запросов. Саджесты.  
Переформулировки.
9. Текстовые модели ранжирования.

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)
7. Обработка запросов. Опечатки. (ДЗ)
8. Классификация запросов. Саджесты. Переформулировки.
9. Текстовые модели ранжирования.
10. Суммаризация текста.

# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
2. Индексация и булев поиск. (ДЗ)
3. Методы оптимизации обратного индекса.
4. Вероятностные языковые модели.
5. Фильтрация. Дубликаты.
6. Фильтрация. Антиспам + антипорн. (ДЗ)
7. Обработка запросов. Опечатки. (ДЗ)
8. Классификация запросов. Саджесты. Переформулировки.
9. Текстовые модели ранжирования.
10. Суммаризация текста.
11. Граф знаний.

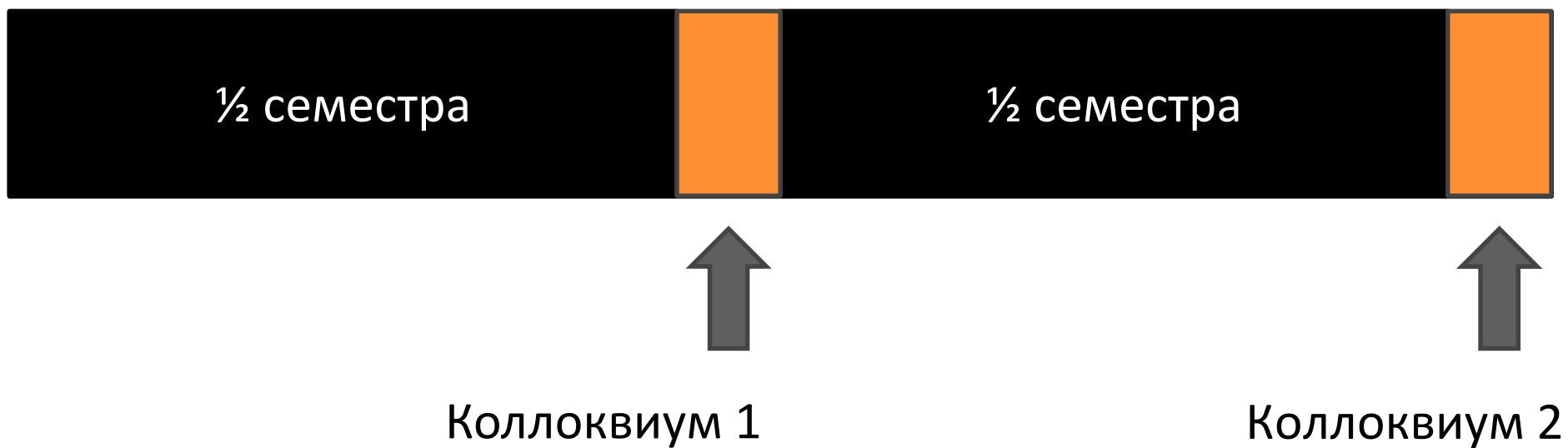
# О чем мы будем говорить на лекциях

1. Что такое поиск и с чем он работает.
  2. Индексация и булев поиск. (ДЗ)
  3. Методы оптимизации обратного индекса.
  4. Вероятностные языковые модели.
  5. Фильтрация. Дубликаты.
  6. Фильтрация. Антиспам + антипорн. (ДЗ)
  7. Обработка запросов. Опечатки. (ДЗ)
  8. Классификация запросов. Саджесты. Переформулировки.
  9. Текстовые модели ранжирования.
  10. Суммаризация текста.
  11. Граф знаний.
- + конкурс по текстовому ранжированию

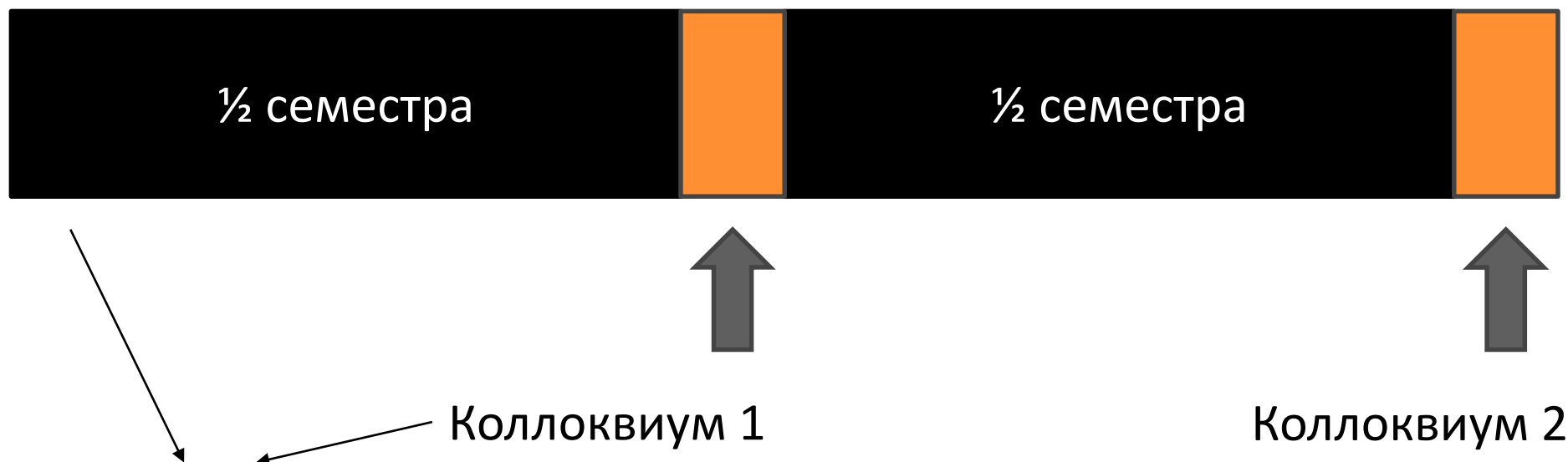
# Структура курса

СЕМЕСТР

# Структура курса

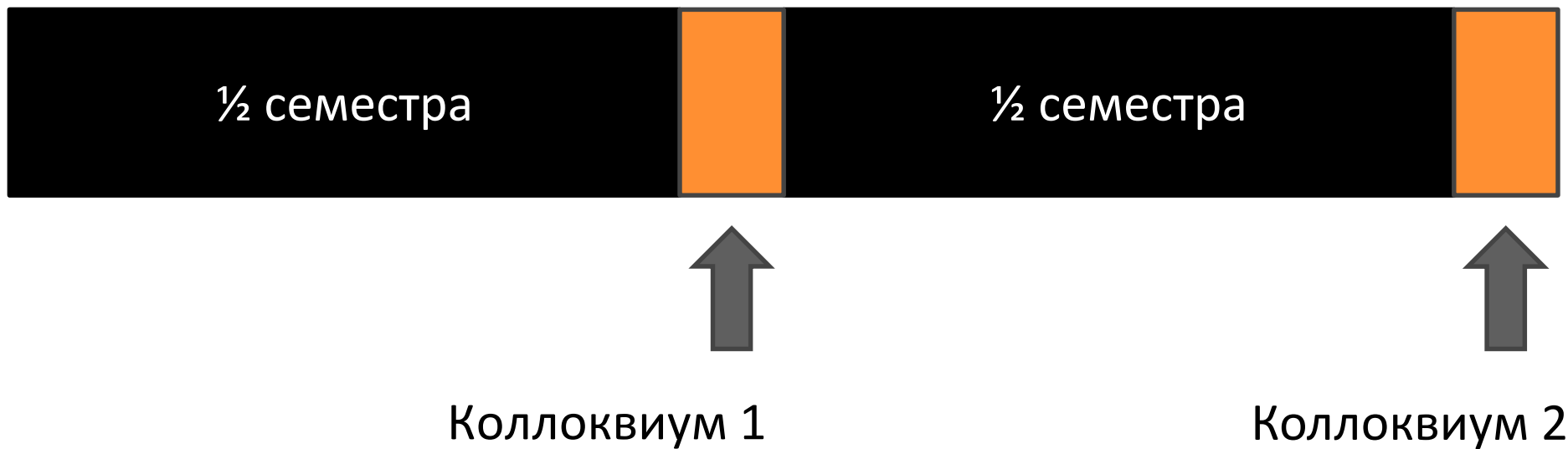


# Структура курса



1 домашка (до 25 баллов)  
работа на занятиях (0-20 баллов)  
коллоквиум (опционально, 0-10 баллов)  
порог выживания = 10 баллов

# Структура курса



Весь курс:

3 домашние работы

2 коллоквиума (опционально, 0-20 баллов)

1 конкурсное задание (обязательно!!!)



# Делайте домашки!

После первого коллоквиума у вас должно быть 10 баллов. Любыми средствами.

Не набрали – прощаемся с вами в середине семестра.

# Оценки

За что можно получить баллы:

- домашние работы
- активная работа на лекция и семинарах
- сдача коллоквиумов

# Домашние работы

- проверяются преимущественно автоматически
- списывание и/или непроходимая глупость – 0 баллов
- дедлайн – обычно 2 недели, но уточняйте

2 недели

до конца курса

N баллов

N/2 баллов

Если что-то непонятно:

в первую очередь спрашивайте преподавателя, который выдал домашку/читал лекции.

# Блог – ваш друг

<https://sphere.mail.ru/blog/view/40/>

Зачем?

- Выкладываются лекции
- Перед семинаром даются указания
- Выкладываются домашки (и иногда результаты)
- Иногда бывают изменения в расписании

# Лайки!



...а еще отзывы и посещаемость