

Информационный поиск

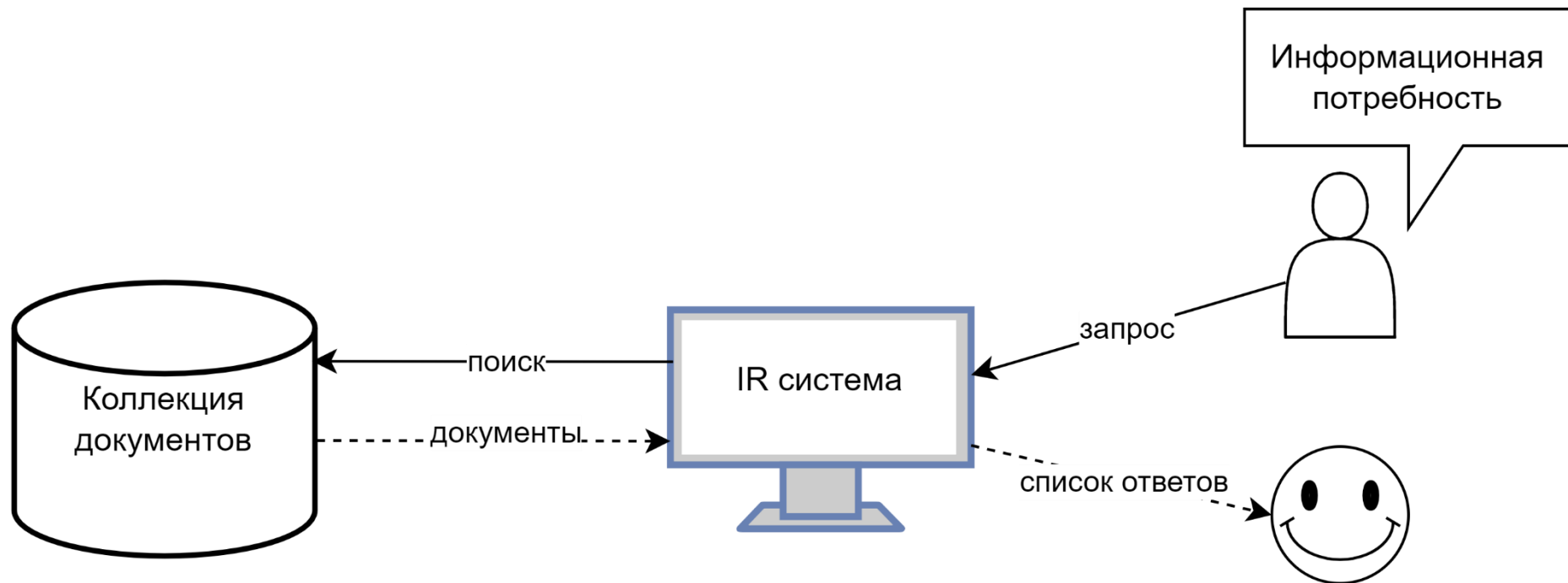
Лекция 11

Кудинова Светлана Дмитриевна

26 ноября 2021

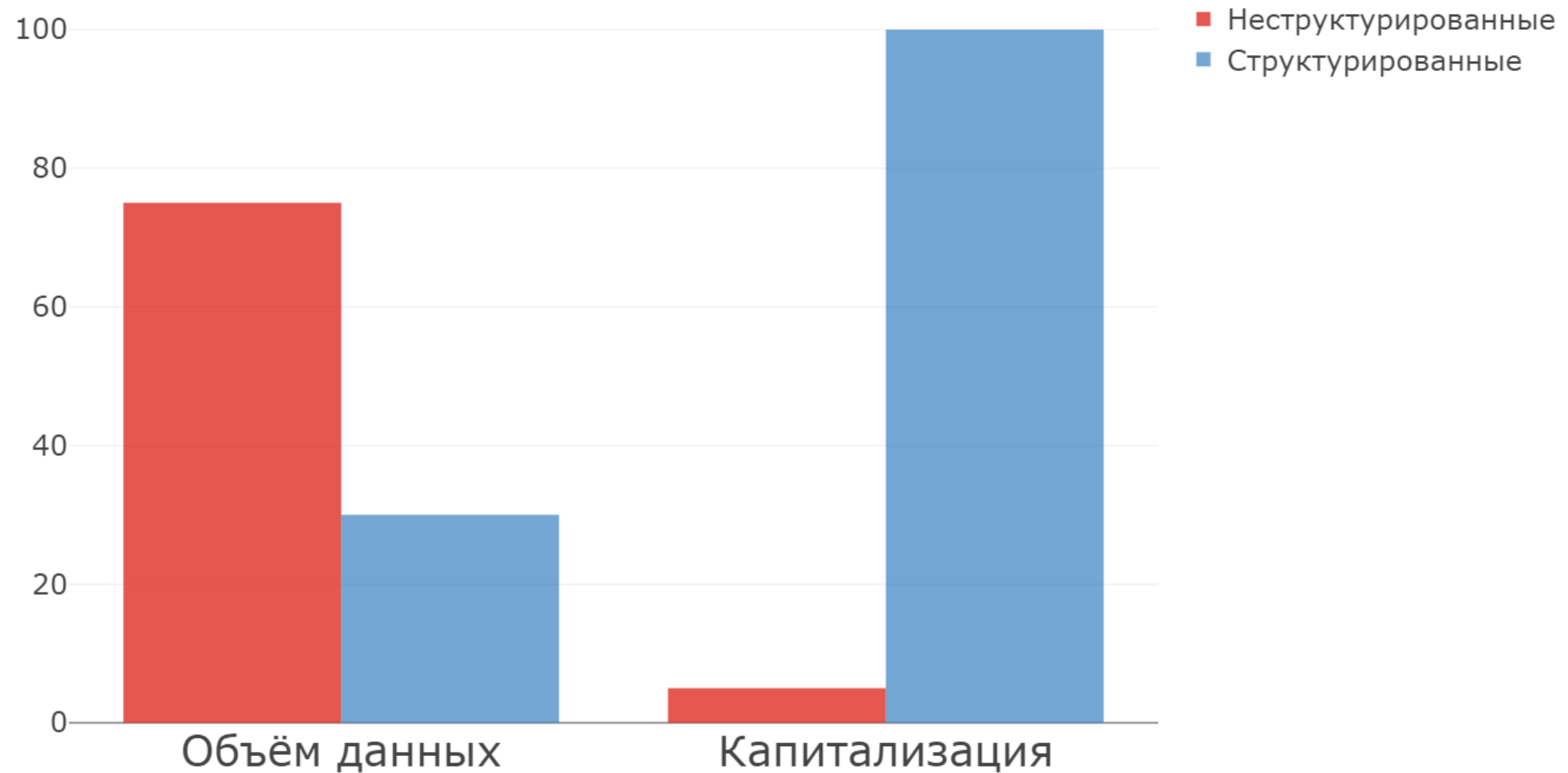
Определение

Информационный поиск (Information retrieval – IR) — это процесс поиска в большой коллекции (хранящейся, как правило, в памяти компьютеров) некоего неструктурированного материала, удовлетворяющего информационные потребности.

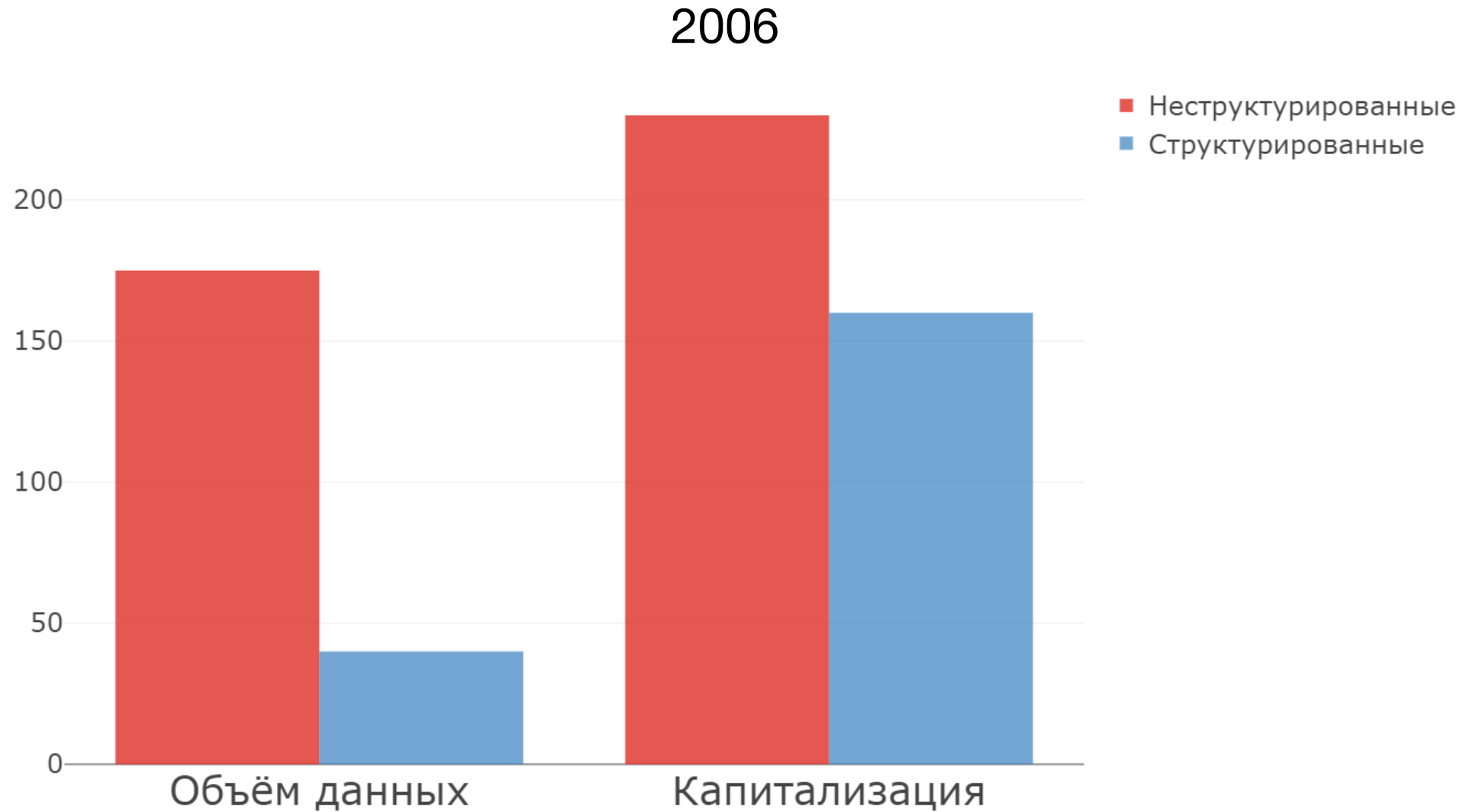


Распространение задачи

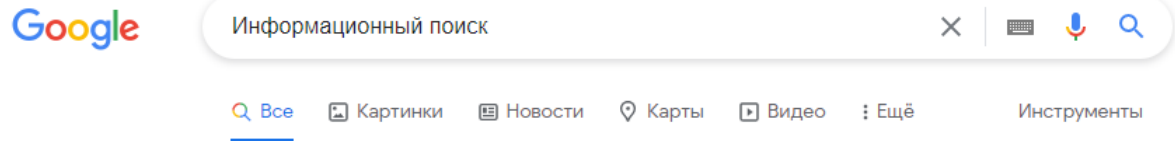
1996



Распространение задачи



Поисковые системы



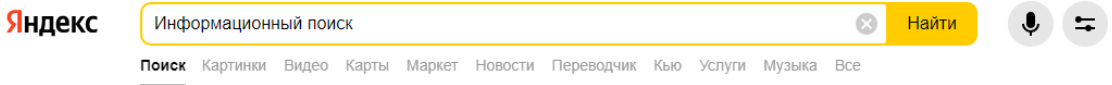
Результатов: примерно 107 000 000 (0,56 сек.)

https://ru.wikipedia.org/wiki/Информационный_поиск

Информационный поиск — Википедия

Информацио́нный по́иск (англ. information retrieval) — процесс поиска неструктурированной документальной информации, удовлетворяющей...
 История · Виды поиска · Методы поиска · Задачи информационного...

<http://www.aonfb.ru/win/book/Doc17>



Быстрый ответ

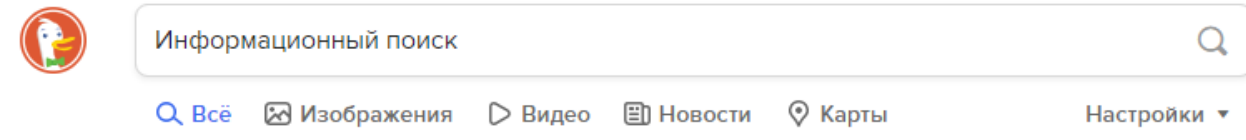
Технология **поиска** и отбора информации **Информационный поиск** — процесс выявления в некотором множестве документов (текстов) всех таких, которые посвящены указанной теме (предмету), удовлетворяют заранее определенному условию **поиска** (запросу) или содержат необходимые (соответствующие **информационной** потребности) факты, сведения, данные. В Интернете с каждым днем скапливается все больше информации когда-либо созданной и вновь создаваемой людьми.

edu.khsu.ru > Технология поиска и отбора информации.pdf ...
 Технология поиска и отбора информации

Результаты поиска

Информационный поиск — Википедия
ru.wikipedia.org > Информационный поиск ...
 Информацио́нный по́иск — процесс **поиска** неструктурированной документальной

Нашлось 7 млн результатов
[Показать только коммерческие](#) [Показать только не коммерческие](#)
[Разместить рекламу](#)



Россия · Безопасный поиск: умеренный · За всё время

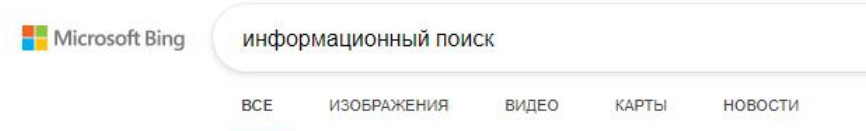
https://ru.wikipedia.org/wiki/Информационный_поиск

Информационный поиск — Википедия

Информацио́нный по́иск — процесс поиска неструктурированной документальной информации, удовлетворяющей информационные потребности, и наука об этом поиске.

https://info-farm.ru/alphabet_index/info-farm.ru/infomacionnyj-poisk.html

Информационный поиск



Результаты: 3 020 000 Дата

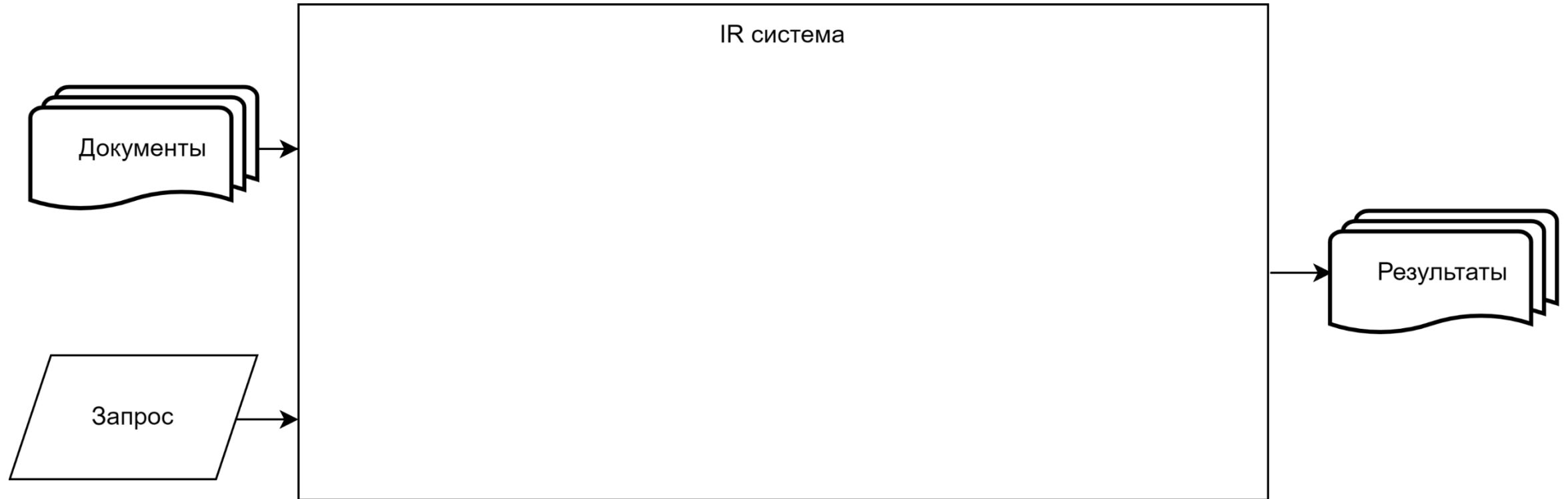
Библиотеки: Математическое образование: прошлое и настоящее

5. Библиографические источники и научный стиль

Содержание

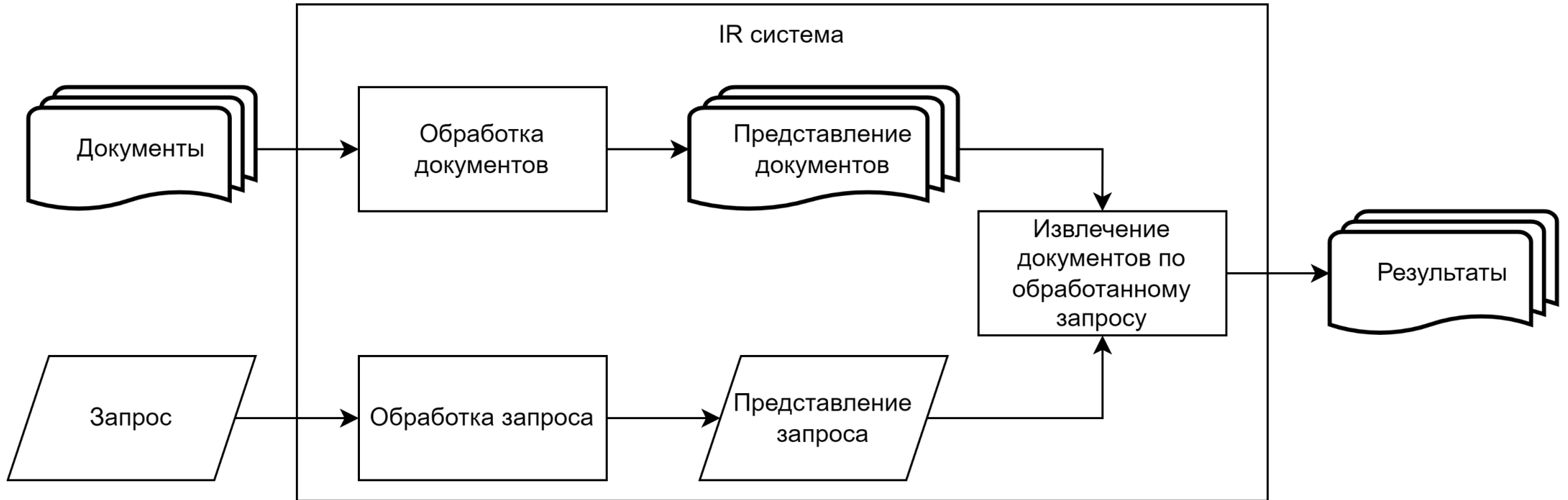
- Цели и задачи
- Минимум
- Особенности баз данных
- Стратегии
- Характеристика патентов
- Патентная инновация

Архитектура IR систем



Что входит в систему?

Архитектура IR систем



Задачи информационного поиска

- **Обработка документа**
В каком виде оптимально хранить информацию о документе.
В каком виде оптимально хранить документ в памяти
- **Обработка запроса**
Как представить запрос
- **Извлечение документов**
Какой документ удовлетворяет запросу
- **Оценка системы**
Как узнать, что система работает хорошо

Что такое документ?

Документ 1

This is first document with one sentence.

Документ 2

This is another document

Документ 3

Third document.

Матрица инцидентности

Документ 1

This is first document with one sentence.

Документ 2

This is another document

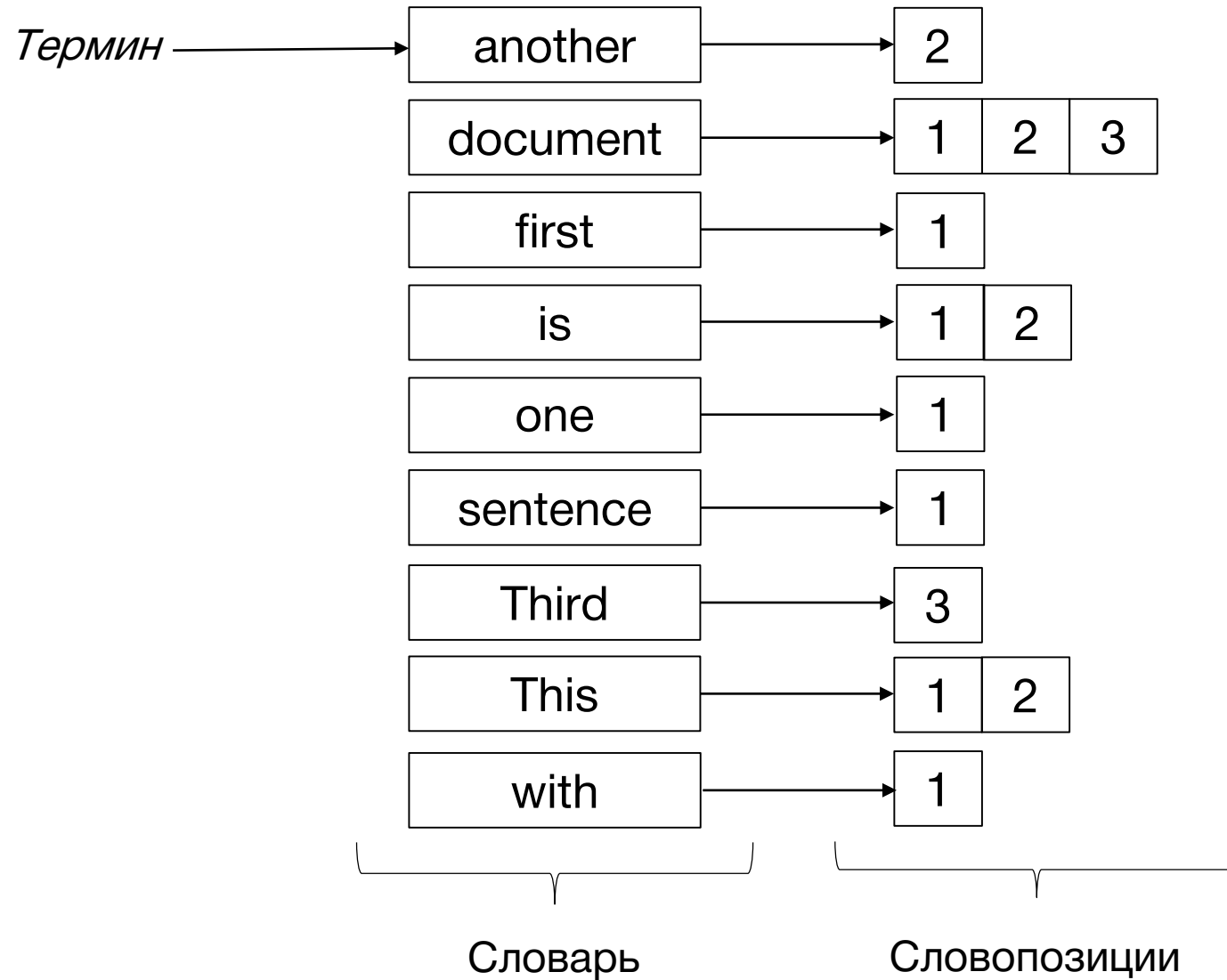
Документ 3

Third document.

	1	2	3
THIS	1	1	0
IS	1	1	0
FIRST	1	0	0
DOCUMENT	1	1	1
WITH	1	0	0
ONE	1	0	0
SENTENCE	1	0	0
ANOTHER	0	1	0
THIRD	0	0	1

Чем плоха матрица инцидентности?

Инвертированный индекс



Предварительная обработка

- Игнорирование распространенных терминов: стоп-слова
- Нормализация
 - Ударения и диакритические символы
 - Обработка без учета регистра
- Стемминг и лемматизация

Модели информационного поиска

- Документ D = множество взвешенных ключевых слов
- Запрос Q = множество невзвешенных слов

$$R(D, Q) = \sum_i w(t_i, D)$$

t_i — слова запроса

Булева модель

- Документ – конъюнкция слов
- Запрос – булево выражение (*AND*, *OR* и *NOT*)

$$R(D, Q) = Q \rightarrow D$$

$$D = t_1 \wedge t_2 \wedge \dots \wedge t_n$$

$$Q = (t_1 \wedge t_2) \vee (t_3 \wedge t_4)$$

$$Q \rightarrow D \text{ — то есть } R(D, Q) = 1$$

Булева модель с матрицей инцидентности

$$R(D, Q) = (THIS \textit{ OR THIRD}) \textit{ AND NOT WITH}$$

	1	2	3
THIS	1	1	0
IS	1	1	0
FIRST	1	0	0
DOCUMENT	1	1	1
WITH	1	0	0
ONE	1	0	0
SENTENCE	1	0	0
ANOTHER	0	1	0
THIRD	0	0	1

Булева модель с матрицей инцидентности

$$R(D, Q) = (THIS \textit{ OR } THIRD) \textit{ AND NOT } WITH$$

$$(110 \textit{ OR } 001) \textit{ AND NOT } 100 = 011$$



Результат: 2 и 3 документ

	1	2	3
THIS	1	1	0
IS	1	1	0
FIRST	1	0	0
DOCUMENT	1	1	1
WITH	1	0	0
ONE	1	0	0
SENTENCE	1	0	0
ANOTHER	0	1	0
THIRD	0	0	1

Какой алгоритм будет при работе с инвертированным индексом?

Найдите

1. (*NOT ANOTHER OR DOCUMENT*) *AND* (*IS OR THIS*)
2. (*NOT THIS AND WITH*) *OR* (*DOCUMENT AND THIRD*)
3. (*WITH AND DOCUMENT AND ANOTHER AND FIRST*) *OR* *SENTENCE*

	1	2	3	4	5	6
ANOTHER	0	1	0	0	0	0
DOCUMENT	1	1	1	1	0	1
FIRST	1	0	0	0	0	1
IS	1	1	0	0	0	0
ONE	1	0	0	0	0	0
SENTENCE	1	0	0	0	0	1
THIRD	0	0	1	1	1	0
THIS	1	1	0	1	0	0
WITH	1	0	0	1	0	1

Найдите

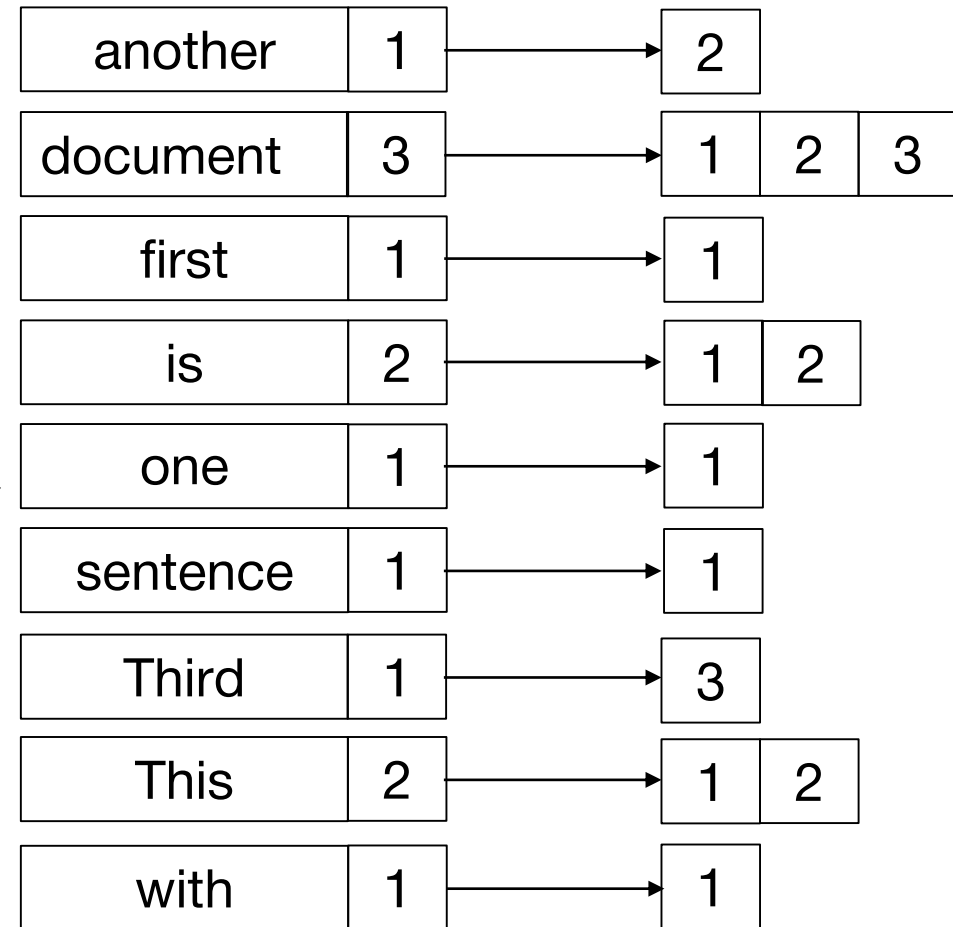
1. (*NOT ANOTHER OR DOCUMENT*) *AND* (*IS OR THIS*) – 1, 2, 4
2. (*NOT THIS AND WITH*) *OR* (*DOCUMENT AND THIRD*) – 3, 4, 6
3. (*WITH AND DOCUMENT AND ANOTHER AND FIRST*) *OR* *SENTENCE* – 1, 6

	1	2	3	4	5	6
ANOTHER	0	1	0	0	0	0
DOCUMENT	1	1	1	1	0	1
FIRST	1	0	0	0	0	1
IS	1	1	0	0	0	0
ONE	1	0	0	0	0	0
SENTENCE	1	0	0	0	0	1
THIRD	0	0	1	1	1	0
THIS	1	1	0	1	0	0
WITH	1	0	0	1	0	1

Булева модель с обратным индексом

Рассмотрим 2 варианта поиска:

- Уточнение запроса
(THIS OR THIRD) AND NOT WITH:
 1. *THIS OR THIRD*
 2. *NOT WITH*
 3. *(THIS OR THIRD) AND NOT WITH*
- Пересечение



Нечёткий поиск. Запросы с джокерами

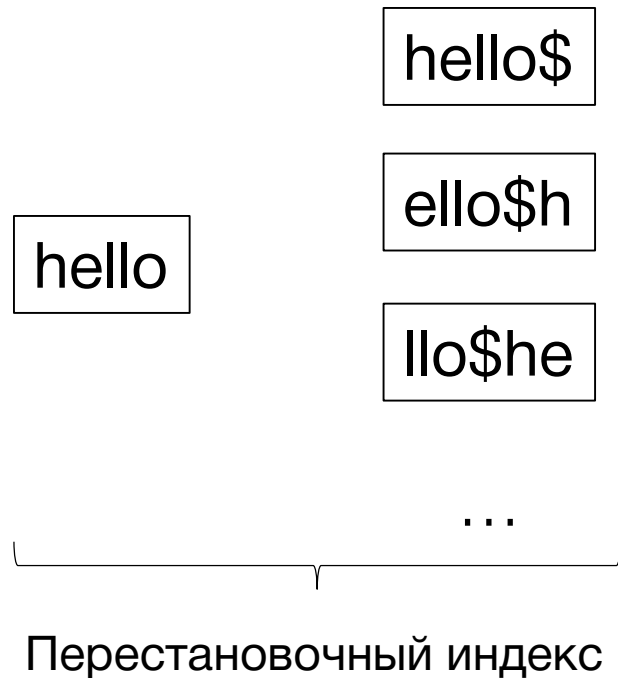
Часть запроса содержит джокер или wildcard, обозначенный символом " * "

Используются в следующих ситуациях:

1. пользователь не знает, как правильно пишется слово
2. пользователь знает о существовании нескольких вариантов написания
3. пользователь ищет документы, содержащие вариант термина, унифицированный в результате стемминга
4. пользователь не уверен в правильности воспроизведения иностранного слова или фразы

Перестановочные индексы

Введём специальный символ \$, обозначающий конец термина.
 Перестановочный индекс терминов получается циклической перестановкой символов исходного термина



Для поиска может потребоваться некоторая модификация запроса:

$$m^*n \rightarrow n\$m^*$$

$$fi^*mo^*er \rightarrow er\$fi^* \rightarrow er\$mo^*$$

к-граммный индекс

При использовании к-граммного индекса словарь содержит все к-граммы, образованные из всех терминов лексикона. Каждый инвертированный список ставит в соответствие к-грамме все термины лексикона, содержащие данную к-грамму.

3-грамма: castle → \$ca – cas – ast – stl – tie – le\$

Модификация запроса (для 3-граммного индекса):
re*ve → \$re AND ve\$

Недостатки булева поиска

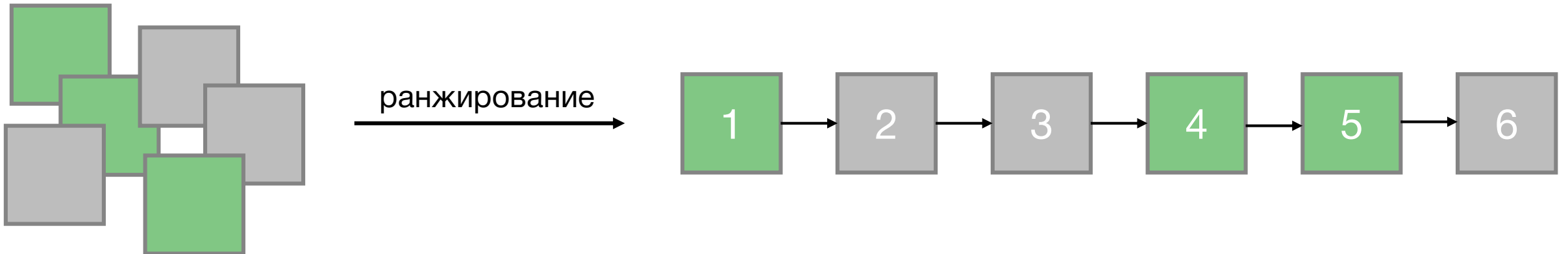
- $R \in [0,1]$ – неупорядоченные документы
- Обычно результатов на запрос либо слишком много, либо слишком мало
 - Запрос 1: “скачать бесплатно” (4 млн. результатов)
 - Запрос 2: “скачать бесплатно без регистрации без смс без кидалава без хурмы” (0 результатов)
- Писать булевы запросы сложно

Задача ранжирования

- Для некоторого запроса $q \in Q$ найдено множество документов \rightarrow имеем выборку пар (d, q)
- Y — упорядоченное множество рейтингов, более высокое значение соответствует более высокой релевантности
- $y : X \rightarrow Y$ — оценки, предоставленные ассессорами
- Правильный порядок определён только на множестве документов, найденных по одному запросу

Ранжирование

- Параметрические и зонные индексы
- Взвешивание на основе комбинации частоты и обратной документной частоты термина
- Вероятностные методы
- Определение весов на основе машинного обучения



Релевантность документа

Взвешивание слов (терминов)

- tf – term frequency - частота встречаемости термина в документе
- df – document frequency - число документов, содержащих термин
- idf – inverse document frequency - специфичность термина
- $weight(t, D) = tf(t, D) \cdot idf(t)$

Варианты tf-idf

$$idf(t) = \log\left(\frac{N}{n}\right)$$

n – количество документов, содержащих t

N – количество документов в корпусе

- $tf(t, D) = freq(t, D)$
- $tf(t, D) = \log[freq(t, D)]$
- $tf(t, D) = \log[freq(t, D)] + 1$
- $tf(t, D) = \frac{freq(t, d)}{Max[f(t, d)]}$

Пример расчёта tf-idf

Документы:

- маленький котик ест еду
- большой щенок ест еду
- маленький котик большой котик и маленький щенок едят еду

еду	1
ест	1
щенок	0
маленький	1
большой	0
и	0
котик	1
едят	0

Прямой индекс 1

еду	1
ест	1
щенок	1
маленький	0
большой	1
и	0
котик	0
едят	0

Прямой индекс 2

еду	1
ест	0
щенок	1
маленький	2
большой	1
и	1
котик	2
едят	1

Прямой индекс 3

еду	3
ест	2
щенок	2
маленький	2
большой	2
и	1
котик	2
едят	1

Обратный индекс

Пример расчёта tf-idf

Документы:

- маленький котик ест еду
- большой щенок ест еду
- маленький котик большой котик и маленький щенок едят еду

еду	0
ест	0.176
щенок	0
маленький	0
большой	0
и	0
КОТИК	0.176
едят	0

tf-idf 1

еду	0
ест	0.176
щенок	0.176
маленький	0
большой	0.176
и	0
КОТИК	0
едят	0

tf-idf 2

еду	0
ест	0
щенок	0.176
маленький	0.352
большой	0.176
и	0.477
КОТИК	0.352
едят	0.477

tf-idf 3

Векторная модель

- Векторное пространство всех слов

$$\langle t_1, t_2, t_3, \dots, t_n \rangle$$

- Документ

$$D = \langle a_1, a_2, a_3, \dots, a_n \rangle$$

a_i — вес t_i в документе D

- Запрос

$$Q = \langle b_1, b_2, b_3, \dots, b_n \rangle$$

b_i — вес t_i в запросе Q

Матричное представление

Пространство
документов

Пространство
терминов

	t_1	t_2	t_3	...	t_n
D_1	a_{11}	a_{12}	a_{13}	...	a_{1n}
D_2	a_{21}	a_{22}	a_{23}	...	a_{2n}
D_3	a_{31}	a_{32}	a_{33}	...	a_{3n}
...
D_m	a_{m1}	a_{m2}	a_{m3}	...	a_{mn}
Q	b_1	b_2	b_3	...	b_n

Разреженная матрица (!)

Подсчет близости

Скалярное произведение:

$$Sim(D, Q) = \sum_{i=1}^n a_i \cdot b_i = \langle a, b \rangle$$

Косинусная мера:

$$Sim(D, Q) = \frac{\langle a, b \rangle}{\sqrt{\sum_{i=1}^n a_i^2} \cdot \sqrt{\sum_{i=1}^n b_i^2}}$$

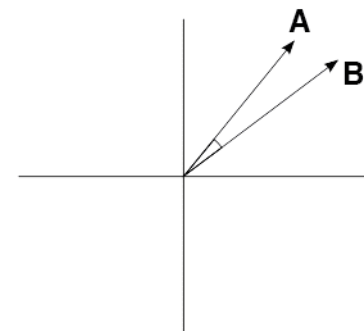
Мера Дайса:

$$Sim(D, Q) = \frac{2\langle a, b \rangle}{\sum_{i=1}^n a_i^2 + \sum_{i=1}^n b_i^2}$$

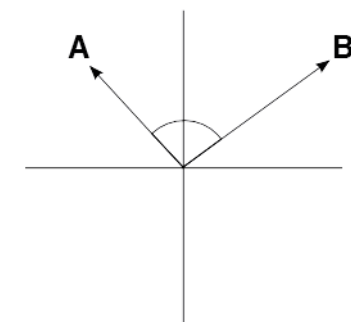
Мера Жаккара:

$$Sim(D, Q) = \frac{\langle a, b \rangle}{\sum_{i=1}^n a_i^2 + \sum_{i=1}^n b_i^2 - \langle a, b \rangle}$$

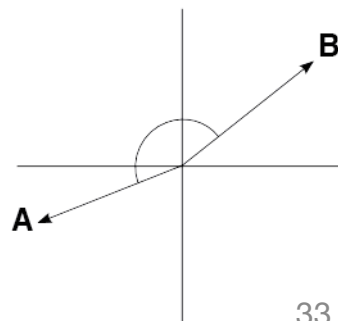
Similar



Unrelated



Opposite



Вероятностная модель

Принцип бинарной независимости

Если поисковая система в ответ на запрос пользователя отдает документы в порядке уменьшения вероятности их релевантности запросу, то качество такой поисковой системы будет максимальным (Robertson, 1977)

Модель бинарной независимости

- Документ представляет собой бинарный вектор термов
- Запрос представляет собой бинарный вектор термов
- Предположение о независимости появления термов в документе

$$P(D|R) = \prod_{i=1}^t P(D_i|R)$$

Окапи BM25

Классический вероятностный алгоритм ранжирования,
 Основанный на модели бинарной независимости

$$RSV_d = \sum_{t \in q} \log \left[\frac{\left(|VR_t| + \frac{1}{2} \right) / \left(|VNR_t| + \frac{1}{2} \right)}{\left(df_t - |VR_t| + \frac{1}{2} \right) / \left(N - df_t - |VR| + |VR_t| + \frac{1}{2} \right)} \right] \\
 \times \frac{(k_1 + 1)tf_{rd}}{k_1((1 - b) + b(L_d/L_{ave})) + tf_{td}} \times \frac{(k_3 + 1)tf_{tq}}{k_3 + tf_{tq}}$$

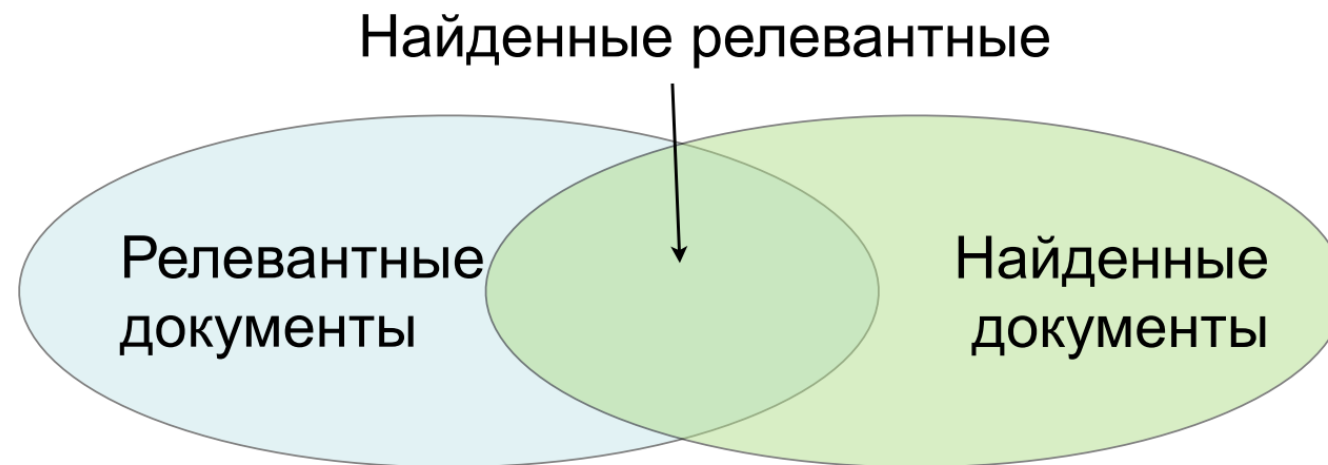
ML методы

- **Point-wise** — поточечный
Каждой паре (q, d) проставлена некоторая численная оценка, и задача сводится к построению регрессии (или классификации, если оценок всего несколько)
- **Pair-wise** — попарный
Для двух документов, соответствующих одному запросу, решается задача бинарной классификации (какой из них релевантнее)
- **List-wise** — списочный
На вход поступает список всех документов, на выходе — их перестановка. Модель напрямую оптимизирует одну из метрик (точнее, её гладкую аппроксимацию)

Оценка информационного поиска

$$\text{Точность } (P) = \frac{\text{количество найденных релевантных}}{\text{количество найденных}}$$

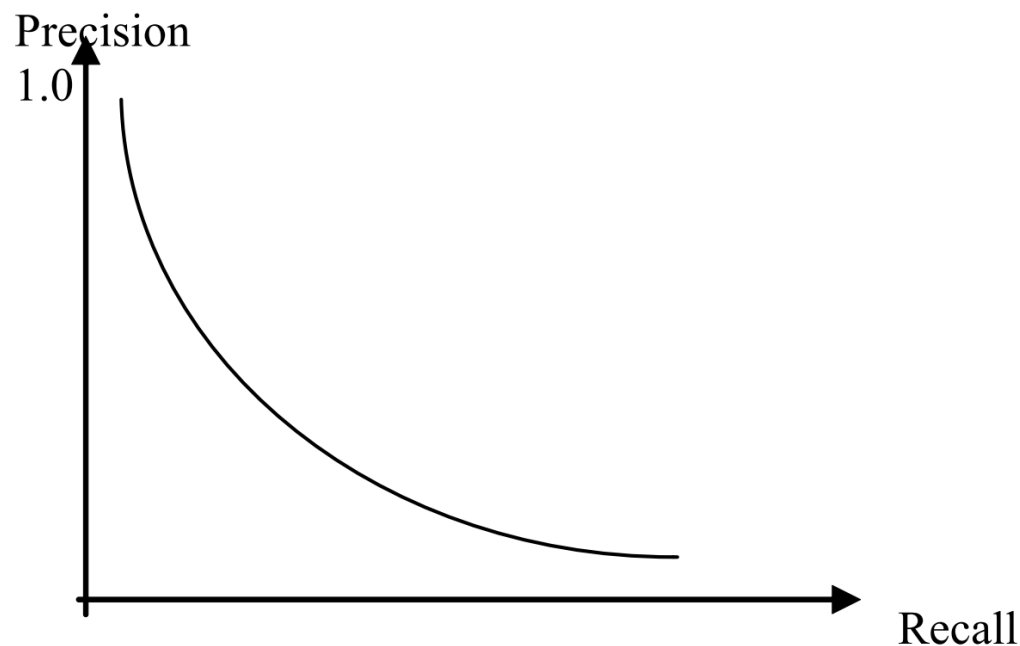
$$\text{Полнота } (R) = \frac{\text{количество найденных релевантных}}{\text{количество релевантных}}$$



Точность и полнота

Общая форма зависимости

- Точность и полнота зависимы
- Системы нельзя сравнивать в одной точке
- Вычисляют среднюю точность (например, в 11 точках с полнотой: 0.0, 0.1, ..., 1.0)



$$AveP = \int_0^1 p(r) dr$$

$$AveP = \sum_{k=1}^n P(k) \Delta r(k)$$

$$AveP = \frac{\sum_{k=1}^n (P(k) \times rel(k))}{\text{число релевантных документов}}$$

$rel(k) \in \{0,1\} = 1$, если k -ый документ релевантен запросу

MAP

Mean Average Precision

$$MAP = \frac{1}{n} \sum_{Q_i} \frac{1}{|R_i|} \sum_{D_j \in R_i} \frac{j}{r_{ij}}$$

- r_{ij} – ранг j -го релевантного документа для Q_i
- $|R_i|$ – число релевантных документов для Q_i
- n – количество тестовых запросов

Ранг	1	4	1-ый релевантный документ
	5	8	2-ый релевантный документ
	10		3-ый релевантный документ

$$MAP = \frac{1}{2} \left[\frac{1}{3} \left(\frac{1}{1} + \frac{2}{5} + \frac{3}{10} \right) + \frac{1}{2} \left(\frac{1}{4} + \frac{2}{8} \right) \right] \approx 0.408$$

Обратная связь

- Явная оценка пользователей
- Неявные показатели релевантности:
 - Оценка времени просмотра документа
 - Клики по документам
 - Другие метаданные

Обратная связь

Проблемы обратной связи

- Пользователи часто отказываются от участия в сборе обратной связи
- Неявная оценка менее надежна
- Поисковая потребность может не соответствовать запросу

Web-поиск

Особенности:

- Постоянный рост документов
- Поиск не только текста, но и мультимедиа
- Спам
- Реклама
- Наличие полуструктуры (теги, микроразметка)

Какие задачи решаются в web-поиске?

- Переиндексация (поисковые роботы)
- Хранение
- Генерация сниппетов
- Работа с дубликатами
- Фасеты
- Фильтрация нежелательного контента

Реклама

Поисковая система Goto предоставляла только рекламные объявления, ранжирование было по цене.

Сейчас используется аукционная система для ранжирования рекламы

Реклама

Яндекс

купить пальто

Найти

Поиск Картинки Видео Карты Маркет Новости Переводчик Кью Услуги Музыка Все

Женское пальто ElectraStyle

[sbermegamarket.ru](#) > Пальто-женские-Elect... реклама

Большой выбор женских платьев. Удобная доставка и примерка перед **покупкой!** · Реальные отзывы. Вход по Сбер ID. Доставка в срок. Безопасная оплата

Контактная информация · +7 (800) 600-XX-XX Показать · круглосуточно · м. Бауманская · Москва

Доставка

География доставки – более 450 городов

+3% кэшбэка

При оплате онлайн

Программа лояльности

Оплачивай до 50% бонусными рублями

Акция "Гарантия лучших цен"

Хорошо, когда выгодно

Нашлось 15 м

[Показать топы](#)

596 тыс. показ

[Разместить ре](#)

Купить женское пальто цена. Бесплатная доставка!

[msk.dreamwhite.ru](#) > Женское-пальто реклама

Трендовые модели. Размеры 38-50. Шоу-румы с огромным выбором. Скидка на 1-ю **покупку!** · 20 лет на рынке. Трендовые модели. Скидки и Акции

Женские плащи · Женские куртки · Женские **пальто** · Доставка и оплата

Контактная информация · +7 (495) 777-XX-XX Показать · ежедневно 12:00-21:00 · Москва

Премиальные пальто. Доп скидка -15% к Black Friday

[noone.ru](#) > NO-ONE реклама

Дарим -15% по коду Night15 и -5% за оплату онлайн. С 20:00 25.11 до 10:00 26.11. · Доставка по России. Более 60 бутиков. 100+ брендов. -3% при оплате онлайн

Женская обувь · Мужская обувь · Ботинки · Кроссовки

Купите пальто. Скидки до 30%

[finn-flare.ru](#) > купить-пальто ★ 4,9 реклама

Только до 30.11 успеете **купить** верхнюю одежду и трикотаж со скидкой до 30%. · Практичная одежда. Качественные материалы. Образы от стилистов

Женские **пальто** · Мужские **пальто** · Базовый гардероб · Одежда из вельвета

Контактная информация · 8 (800) 700-XX-XX Показать · пн-вс 8:00-22:00 · Москва

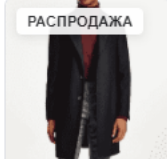




Google

купить пальто

Все Покупки Картинки Карты Новости Ещё Инструменты

Результатов: примерно 27 700 000 (0,88 сек.)

Реклама · Купить: купить пальто

 <p>РАСПРОДАЖА</p>				
<p>Tommy Hilfiger - ...</p> <p>35 992 ₺ 4...</p> <p>Tommy Hilfiger</p>	<p>HUGO</p> <p>50 500 ₺</p> <p>Стокманн</p>	<p>Пальто полушерс...</p> <p>5 499 ₺</p> <p>Каляев</p>	<p>Пальто HENDER...</p> <p>29 999 ₺</p> <p>Henderson</p>	<p>Полупальто мужское...</p> <p>5 192 ₺</p> <p>sbermegam...</p>

Реклама · <https://www.asos.com>

ASOS: пальто - Чёрная Пятница уже близко

У нас большие планы. Будь готов к огромным скидкам в Чёрную Пятницу. Новые **пальто** от именитых дизайнеров и в линейке ASOS Design! Скидки до 80% Аутлет: до -70% Скидки 25% на все товары. Типы: Платья, Обувь, Косметика, Сумки.

Реклама · <https://market.yandex.ru>

Покупайте женские пальто - Покупайте на Яндекс.Маркете

Кешбэк до 5% баллами и экономия до 99% для подписчиков Яндекс Плюса. Месяц...

Ранжирование web-страниц – Page Rank (1998-2016)

Page Rank (PR) – алгоритм ссылочного ранжирования, основанный на вычислении веса страницы путем подсчета важности ссылающихся на нее документов. Выражается числом.

Являлся одним из показателей авторитетности сайта для поисковой системы Google.

Page Rank

На странице P_j есть l_j ссылок. Если одна из этих ссылок ведет на страницу P_i , то P_j передаст $\frac{1}{l_j}$ своей важности P_i :

$$I(P_i) = \sum_{P_j \in B_i} \frac{I(P_j)}{l_j}$$

Матрица гиперссылок:

$$H_{ij} = \begin{cases} \frac{1}{l_j}, P_j \in B_i \\ 0, \text{ иначе} \end{cases}$$



Page Rank

Вектор значений Page Rank - рейтинг важности всех страниц:

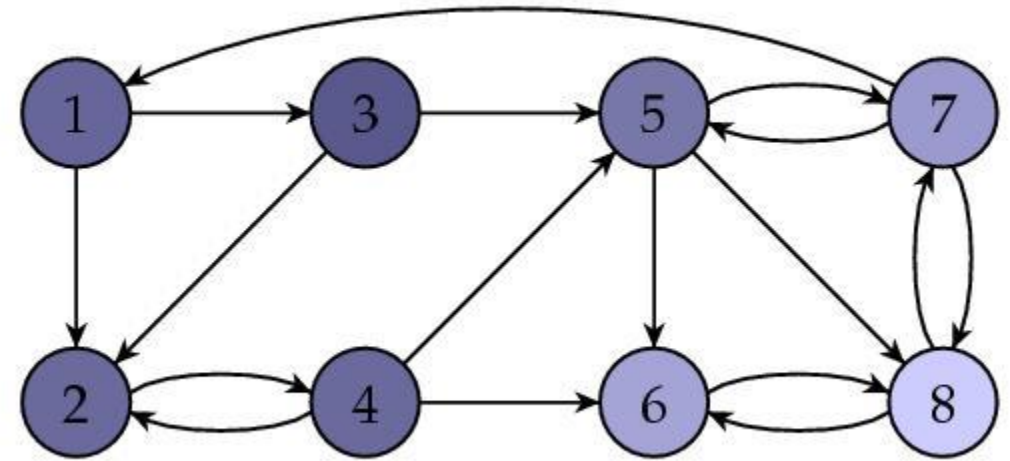
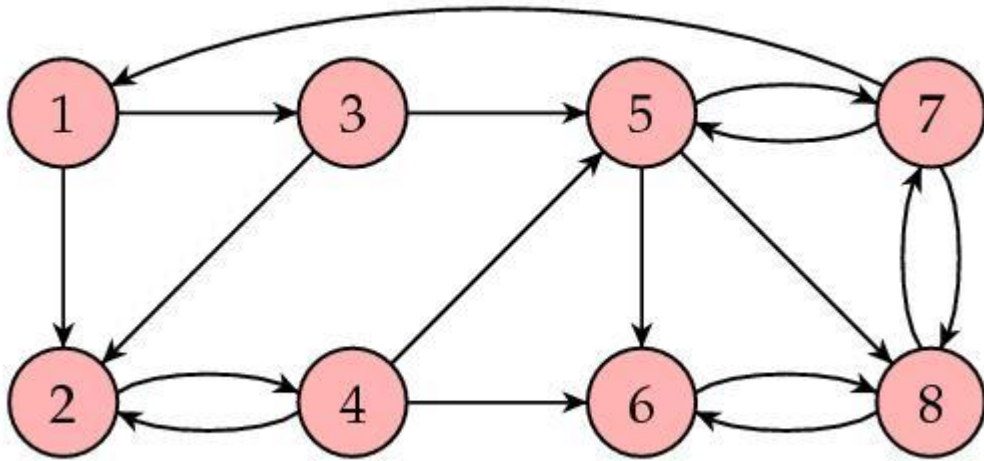
$$I = [I(P_i)]$$

Условие определения PageRank может быть выражено как

$$I = HI$$

Вектор I является собственным вектором матрицы H с собственным значением 1.

Page Rank



$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix} \quad I = \begin{bmatrix} 0.06 \\ 0.0675 \\ 0.03 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.18 \\ 0.295 \end{bmatrix}$$

Page Rank

$$I^{k+1} = HI^k$$

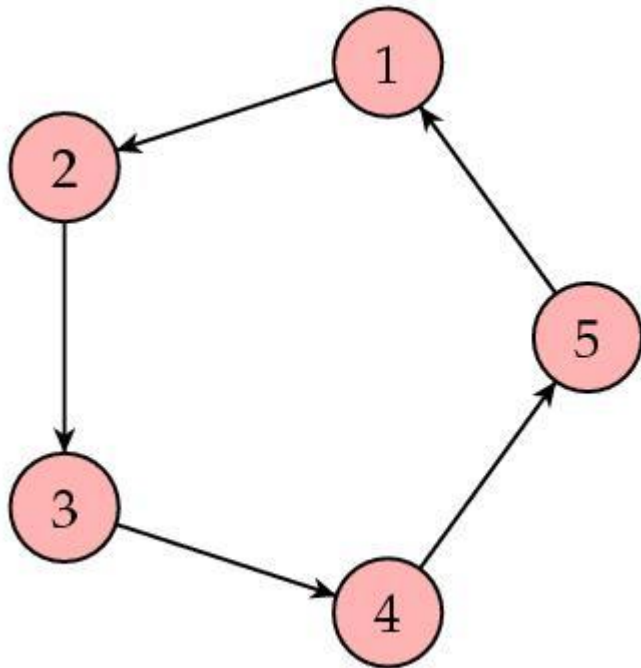
I^0	I^1	I^2	I^3	I^4	...	I^{60}	I^{61}
1	0	0	0	0.0278	...	0.06	0.06
0	0.5	0.25	0.1667	0.0833	...	0.0675	0.0675
0	0.5	0	0	0	...	0.03	0.03
0	0	0.5	0.25	0.1667	...	0.0675	0.0675
0	0	0.25	0.1667	0.1111	...	0.0975	0.0975
0	0	0	0.25	0.1806	...	0.2025	0.2025
0	0	0	0.0833	0.0972	...	0.18	0.18
0	0	0	0.0833	0.3333	...	0.295	0.295

Page Rank. А оно работает вообще?

- Всегда ли последовательность I^k сходится?
- Независим ли конечный вектор от начального приближения I_0 ?
- Содержат ли рейтинги важности необходимую информацию?

Описанный выше метод даёт отрицательные ответы на все вопросы. Можно модифицировать метод для удовлетворения всех требований.

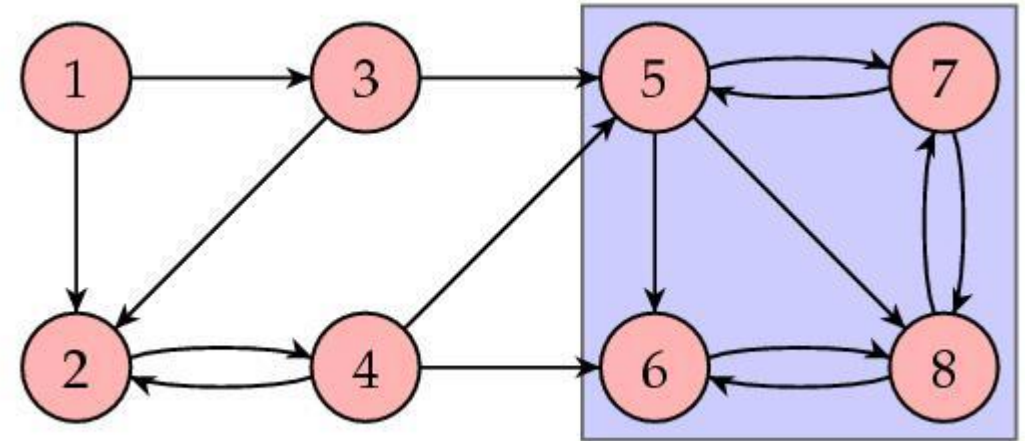
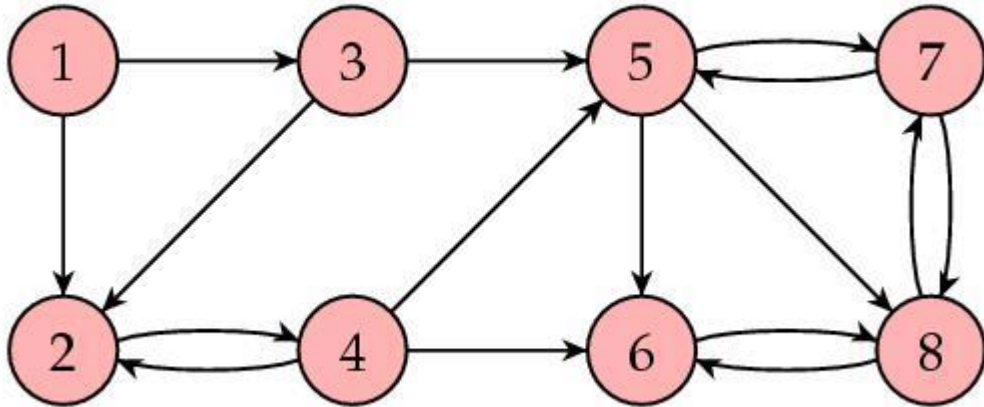
Page Rank. Плохие ситуации



$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

I^0	I^1	I^2	I^3	I^4	I^5
1	0	0	0	0	1
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0

Page Rank. Плохие ситуации



$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/2 & 0 \end{bmatrix} \quad I = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.12 \\ 0.24 \\ 0.24 \\ 0.4 \end{bmatrix}$$

Вопросы?