



Оценка качества поиска

Чернов Евгений

Сегодня в программе:



- Постановка задачи
- Типы метрик
- Стандартные коллекции
- Оценка бинарного поиска
- Оценка ранжирующего поиска
- Маркерные тесты
- Асессоры
- Discounted Cumulative Gain
- A/B тестирование, сплиты

Задача поисковой системы



- Выдача пользователю информации, согласно его информационной потребности
 - Куда бы пойти вечером?
 - Запрос «афиша» поисковой системе
 - Просмотр результатов выдачи
 - Просмотр документа
 - Возврат к странице результатов
 - Покупка билетов, бронь столика ...
- Удовлетворение информационной потребности
- **Повышение счастья пользователя**

Как повысить счастье пользователя?



Счастье может быть измерено в котах

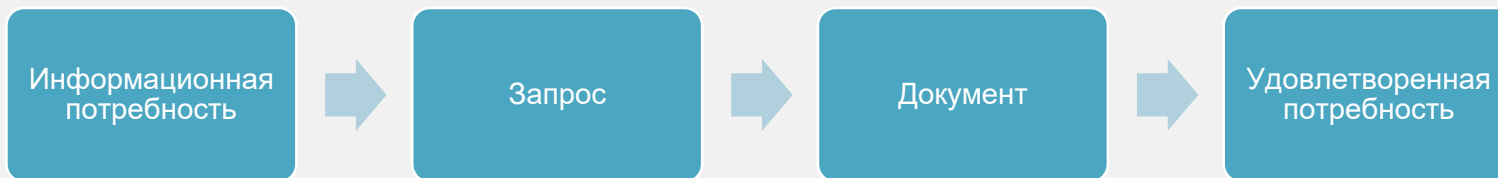
КОТЫ = СЧАСТЬЕ



Как повысить счастье пользователя?



С помощью информационного поиска:



- Информационная потребность переводится в запрос
- Подменяем счастье релевантностью
- Релевантность оценивается в соответствии с информационной потребностью, а не с запросом
- Запрос: [белевская пастила]
 - Ищем сайт производителя
- Запрос: [торт павлова]
 - Ищем рецепт

Как улучшить поиск



- Добавить еще 1 млрд. документов в индекс
- Улучшить поисковый алгоритм
- Перейти на Ajax интерфейс
- Сделать область с информационными карточками
- Улучшить поисковые подсказки
- ...

Как мы можем быть уверены что эти изменения сделают пользователя более счастливым?



“

То что вы не можете измерить – вы не можете улучшить



Лорд Кельвин

Сложность в оценках качества поиска



- Невозможно напрямую измерить счастье пользователя
 - хотя попытки есть
- Невозможно узнать информационную потребность
- Большинство рассуждений носят эмпирический характер, получены на основе практических наблюдений



- База нормативно-правовых документов
 - Важно не пропустить ни одного документа
 - Хотя результат можно и подождать, если не очень долго
- Корпоративная система
 - Безопасность поиска
 - Разграничения прав доступа
 - Типичные сценарии использования
- Торговая площадка
 - Пользователь находит то что ему нужно и покупает
 - Или не нужно, но все равно покупает
 - Максимизируем счастье пользователя или торговой площадки?
- Веб поиск
 - Пользователь находит то, что ему нужно – возвращается обратно
 - Иногда кликает на рекламу
 - Полнота не так важна



Основное предположение:

Спросим пользователей, что они думают

- ✓ Пользователи часть процесса оценки
- ✓ Возможность спросить пользователя обосновать свое мнение, ответ на вопрос «Почему?»
- Дороги и сложны в организации
- Много шумов
- Не воспроизводимы

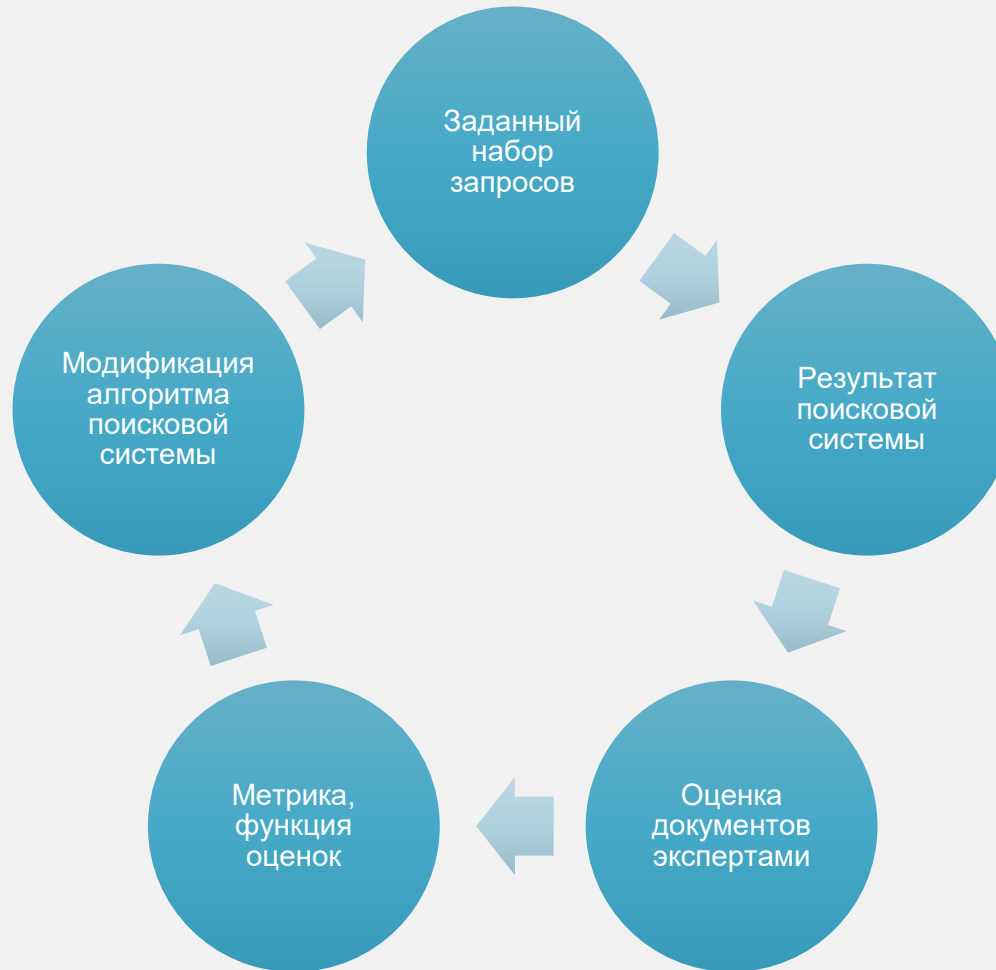


Основное предположение:

Заменим пользователей экспертами

- Нет пользователей
 - Система только разрабатывается
 - Не предполагается online-пользователей
- Формулировка задачи оценки
- Оценка экспертами (асессорами) документов
 - Использование оценок для улучшения поискового алгоритма
 - Предположение об изменении счастья пользователя на основании оценок экспертов

Offline метрики





- TREC – ежегодный конкурс от NIST (National Institute of Standards and Technology)
- Используются разные корпуса
- Указывается “Retrieval task”
 - Иногда в виде запросов
- Люди (эксперты, ассессоры) размечают каждый документ для каждого запроса – **Relevant** или **Nonrelevant**
 - Или берётся подмножество документов, которое было возвращено каким-нибудь поиском по этому запросу



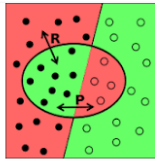
- Модифицируем алгоритм
- Разработаем эксперимент
- Собираем реакцию пользователя
 - Клики в выдачу
 - Время на странице поиска
 - Позиция последнего клика
 - Движение мыши
 - Выделения
 - ...
- Считаем и сравниваем метрики

Google

информационный поиск

информационный поиск **книга**
информационный поиск **и его виды**
информационный поиск **лекции**
информационный поиск **реферат**
Результатов: примерно 693 000 (0,34 сек.)

Информационный поиск (англ. information retrieval) — процесс **поиска** неструктурированной документальной информации, удовлетворяющей **информационные** потребности, и наука об этом **поиске**.



[Информационный поиск — Википедия](https://ru.wikipedia.org/wiki/Информационный_поиск)
https://ru.wikipedia.org/wiki/Информационный_поиск

[Подробнее...](#) • [Оставить отзыв](#)

[Информационный поиск — Википедия](https://ru.wikipedia.org/wiki/Информационный_поиск)
https://ru.wikipedia.org/wiki/Информационный_поиск ▾
Информационный поиск (англ. information retrieval) — процесс поиска неструктурированной документальной информации, удовлетворяющей ...
[История](#) · [Информационный поиск как процесс](#) · [Виды поиска](#) · [Методы поиска](#)

1.6.3. ИНФОРМАЦИОННЫЙ ПОИСК (ОБЩИЕ ТЕРМИНЫ)
www.gpntb.ru/win/book/1/Doc17.HTML ▾
ИНФОРМАЦИОННЫЙ ПОИСК (ОБЩИЕ ТЕРМИНЫ). ПОИСК [search] -. Совокупность операций, связанных с определением местонахождения объектов ...

Информационный поиск. - Поведение потребителей
www.pattern-cr.ru/Lectures/14.html ▾
Информационный поиск. Внутренний и внешний поиск. Типы искомой информации. Источники информации. Измерения и детерминанты поиска.

tpi-it - Информационный поиск.
tpi-it.wikispaces.com/Информационный+поиск ▾
Информационный поиск (information retrieval, data search) — процесс нахождения, отбора и выдачи определенной заранее заданными признаками ...

Основные понятия информационного поиска
koriolan404.narod.ru/tpis/25.htm ▾
Основные понятия информационного поиска. Релевантность, пертинентность и их отличие. Информационный поиск (ИП) (англ. Information retrieval) ...

Информационный поиск - это. Что такое Информационный поиск?



Основное предположение:

Наблюдаемое поведение пользователя отражает релевантность

- Пользователи действуют рационально (в какой-то мере)
 - Хотя это бывает не всегда
- Пользователи используют поисковую систему с какой-то целью
 - Они не тупо кликают в выдачу
- Пользователи последовательно достигают свою цель
 - Нерелевантные документы не привлекают пользователей

Online vs offline. Who wins?



	Online	Offline
Преимущества	<ul style="list-style-type: none">• Дешевизна• Взаимодействие с живыми пользователями• Большой объем данных	<ul style="list-style-type: none">• Контролируемые оценки• Контролируемый объем данных• Контролируемые расходы• Легче испытывать новые идеи
Недостатки	<ul style="list-style-type: none">• Не всегда применимо• Много шумов• Неповторяемость результатов	<ul style="list-style-type: none">• Тяжело спрогнозировать реакцию реальных пользователей• Дороги• Не быстры



- Например
 - Система поиска патентов
 - Корпоративные системы
 - Поиск задач в трекере
- Согласно запросу классифицируем документы
 - Подходят (релевантны)
 - Не подходят (не релевантны)
- Задача
 - Извлекать релевантные документы
 - Не извлекать нерелевантные
- Тестовая коллекция
 - Набор документов
 - Информационные потребности и запросы
 - Разметка документов



Пример запроса TREC 2014 Web Track

```
1.  <webtrack2014>
2.    <topic number="251" type="single">
3.      <query>identifying spider bites</query>
4.      <description>
5.        Find data on how to identify spider bites.
6.      </description>
7.    </topic>
8.    <topic number="252" type="single">
9.      <query>history of orcas island</query>
10.     <description>
11.       Looking for any historical information related to
12.       Orcas Island: geographical, buildings, people,
13.       infrastructure, etc.
14.     </description>
15.   </topic>
```

Точность, полнота

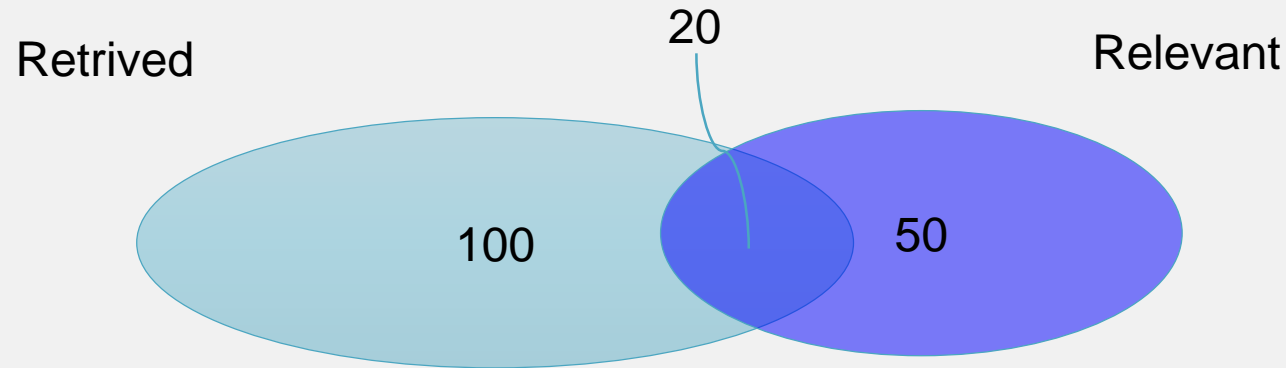


	Relevant	Not-relevant	Sum
Retrived	Tp (true positive)	Fp (false positive)	Tp+Fp
Not retrieved	Tn (false negative)	Fn (false negative)	Tn+Fn
Sum	Tp+Tn	Fp+Fn	Tp+Tn+Fp+Fn

$$Precision = \frac{Tp}{Tp + Fp} = \frac{|\{Relevant docs\} \cap \{retrived docs\}|}{|\{retrived docs\}|}$$

$$Recall = \frac{Tp}{Tp + Tn} = \frac{|\{Relevant docs\} \cap \{retrived docs\}|}{|\{relevant docs\}|}$$

Пример расчета



Let: $Sum = 1000$

$$Precision = \frac{Tp}{Tp + Fp} = \frac{20}{100} = 0,2$$

$$Recall = \frac{Tp}{Tp + Tn} = \frac{20}{50} = 0,4$$

Трудности с точностью и полнотой



- Нужно усреднять по большому количеству документов и запросов
- Нужны ручные оценки
 - И люди не очень надёжный источник оценок
- Оценки бинарные
 - Бывают градации
- Сильно зависит от коллекций
 - Качество поиска на одном корпусе практически ничего не говорит о качестве на другом
- Необходимость сопоставления значений точности и полноты



Объединённая метрика точности и полноты
(взвешенное гармоническое среднее):

$$F = \frac{1}{\alpha \left(\frac{1}{P} \right) + (1 - \alpha) \left(\frac{1}{R} \right)} = \frac{(1 + \beta^2)PR}{\beta^2 P + R}$$

$$F = \frac{2PR}{P + R}$$

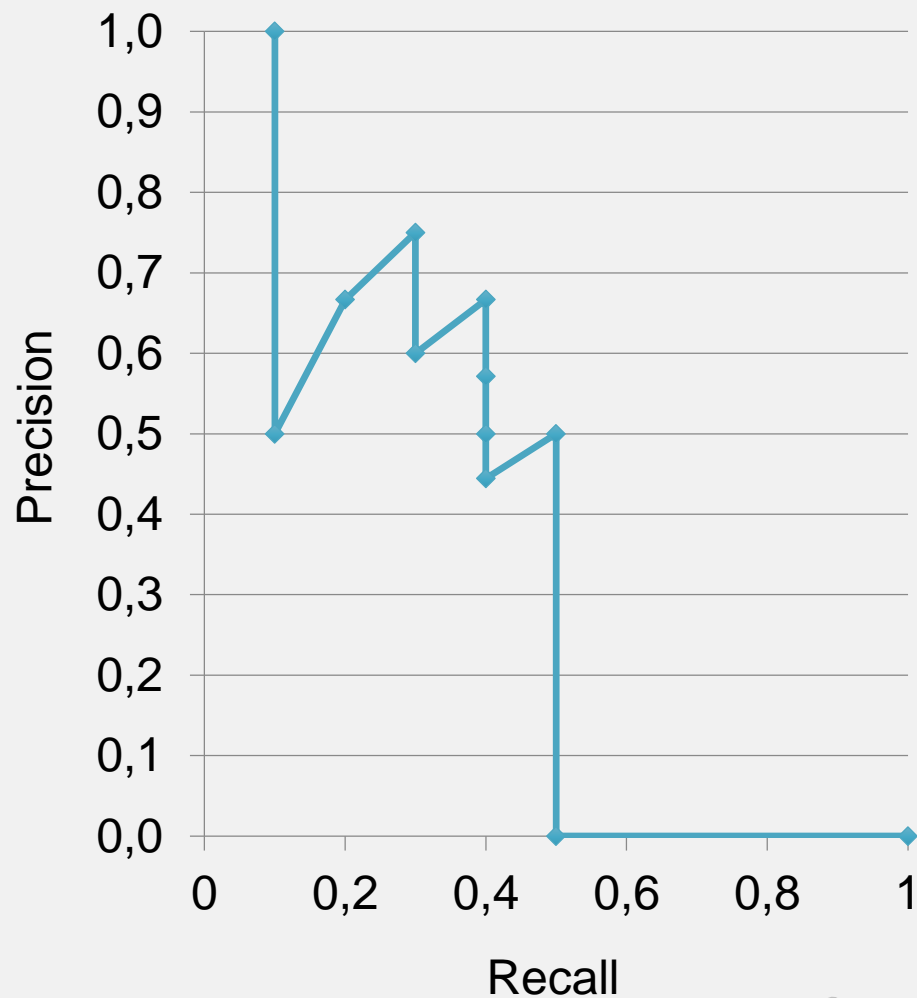


- Предположения:
 - Возвращает любое количество документов
 - Возвращает упорядоченный набор документов
 - Релевантные документы находятся выше нерелевантных
 - Пользователю показываем k первых результатов
- Методика – используя понятие точности и полноты, отбирая документы сверху, построим график точности-полноты

Оценка ранжирующего поиска



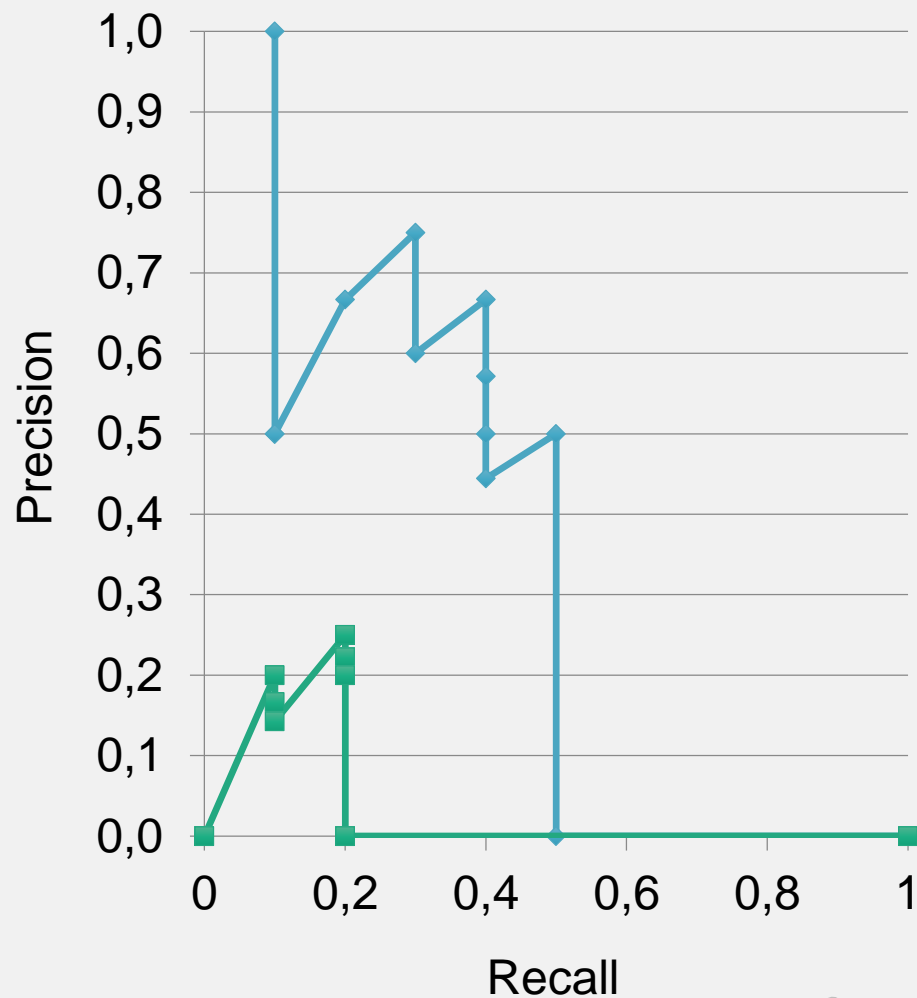
Rank	Rel.	Prec.	Recall
1	R	1.0	0.1
2	N	1/2	0.1
3	R	2/3	0.2
4	R	3/4	0.3
5	N	3/5	0.3
6	R	4/6	0.4
7	N	4/7	0.4
8	N	4/8	0.4
9	N	4/9	0.4
10	R	5/10	0.5
...
∞	R	0	10/10



Оценка ранжирующего поиска



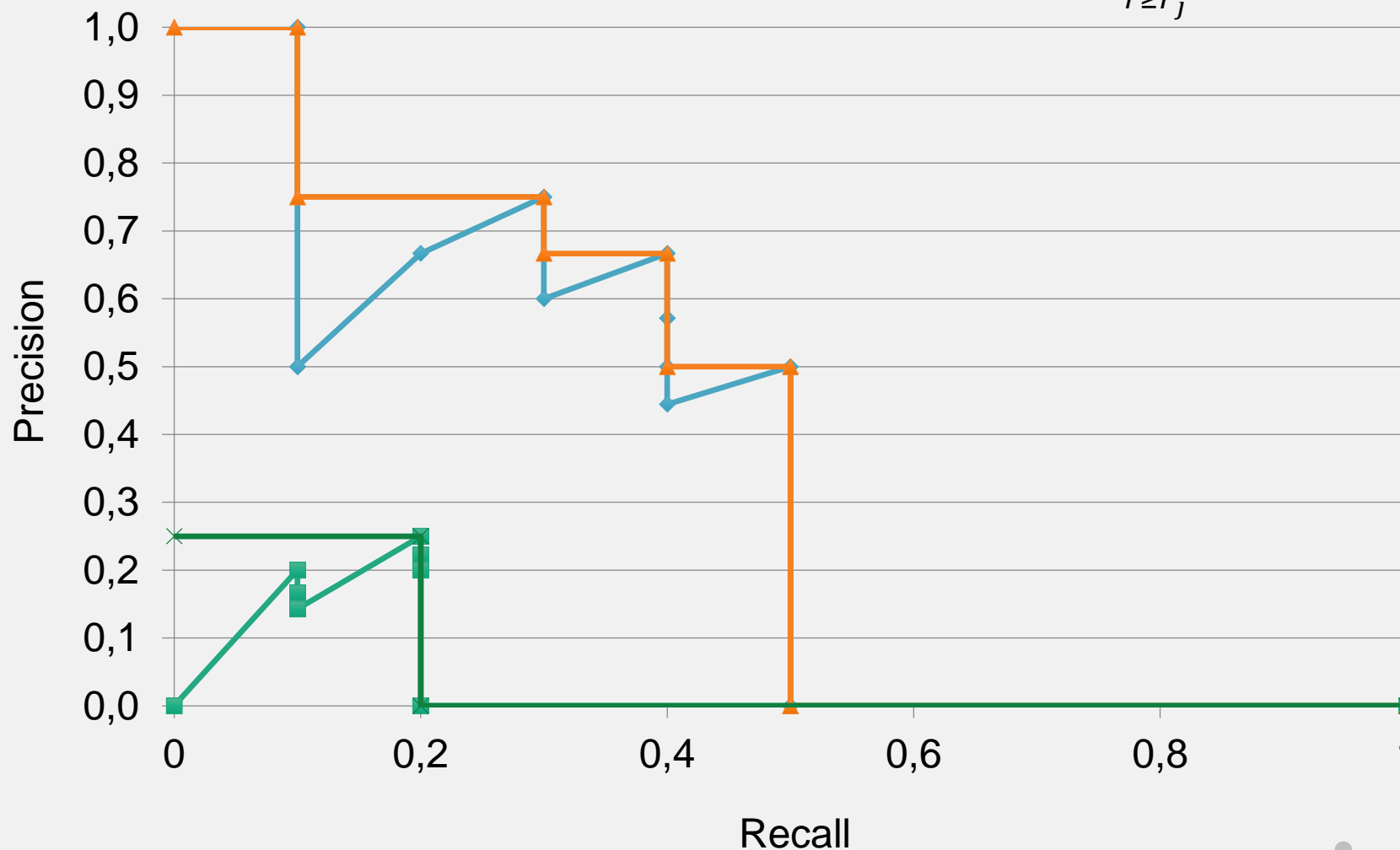
Rank	Rel.	Prec.	Recall
1	N	0	0
2	N		
3	N
4	N		
5	R	1/5	0.1
6	N	1/6	0.1
7	N	1/7	0.1
8	R	2/8	0.2
9	N	2/9	0.2
10	N	2/10	0.2
...
∞	R	0	10/10



Интерполированная точность



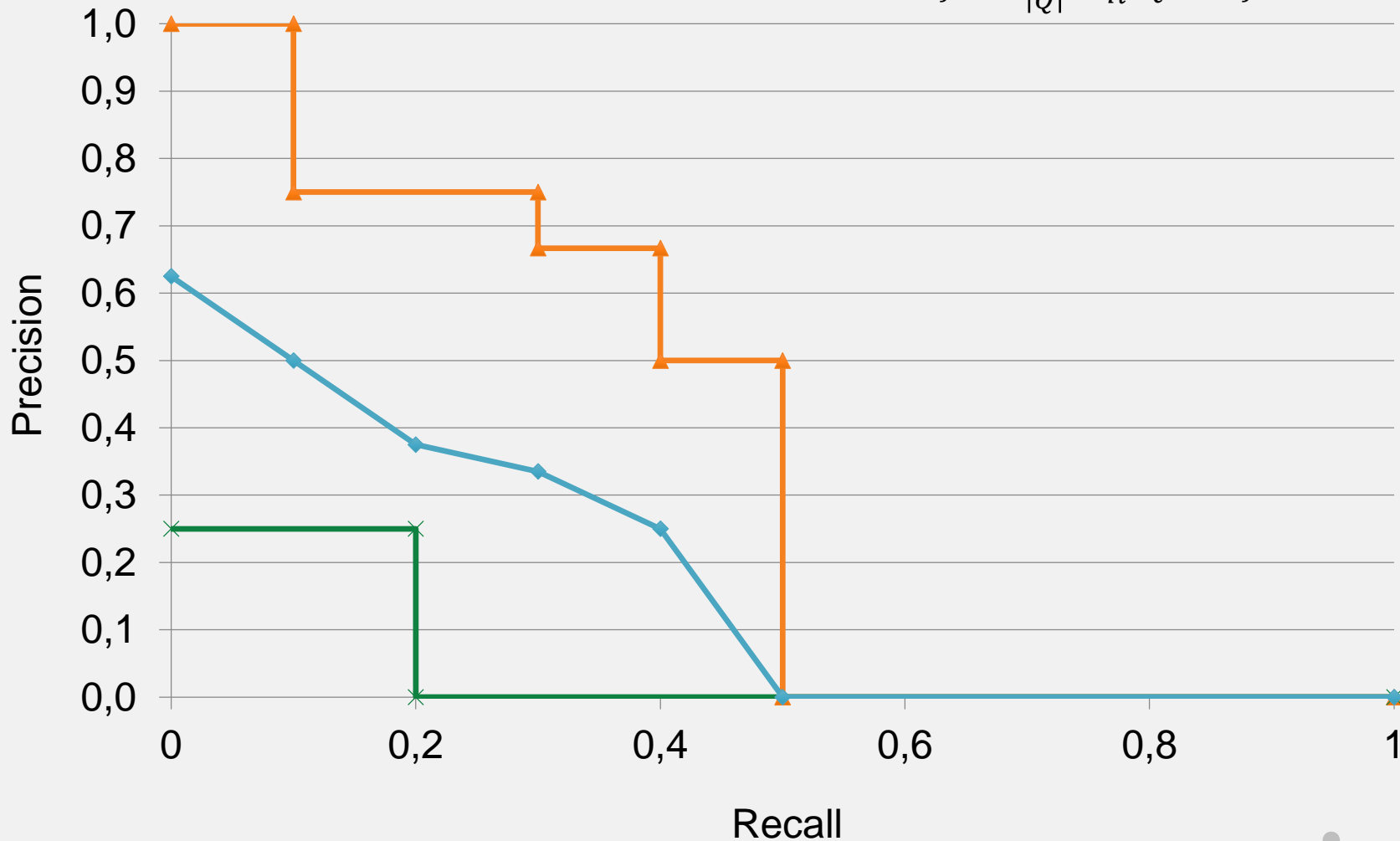
$$P(r_j) = \max_{r \geq r_j} P(r)$$



Средняя интерполированная точность



$$\overline{P(r_j)} = \frac{1}{|Q|} \sum_{q_i \in Q} P_i(r_j)$$





- R-precision, точность на уровне k , обычно $k = 5, 10, 20$. $P@k$ ($P@5, P@10, P@20$)
 - Основной посыл – пользователь просматривает небольшое количество документов первых документов
 - Стремится к единице
- Average precision, средняя точность. Пусть для запроса q_i найдено k релевантных документов, тогда

$$AP = \frac{1}{k} \sum_{j=1}^k P(j)$$

- Для второго случая

$$AP = \frac{1}{2} \left(\frac{1}{5} + \frac{2}{8} \right) = 0,225$$



- Представим популяцию из 10 000 животных
 - Какой процент из них здоровы?
- Применим практику социологических опросов для снижения количества оценок
- Случайным образом выбираем документы из результатов поиска
- Отдаем их на оценку
- Используем частичные оценки для подсчета метрик.

Критика чистой релевантности



- Общая релевантность выдачи
 - Документ может быть релевантным, но в выдаче – лишним
 - Текстовый дубль
 - Повтор информации (по другому написано, но всё равно повтор)
 - Правильнее оценивать всю выдачу, а не каждый отдельный документ
- Документы оценивать не совсем правильно – извлекаемые факты или объекты лучше оценивают релевантность.
- Но зато сильно усложняют процесс создания тестового корпуса

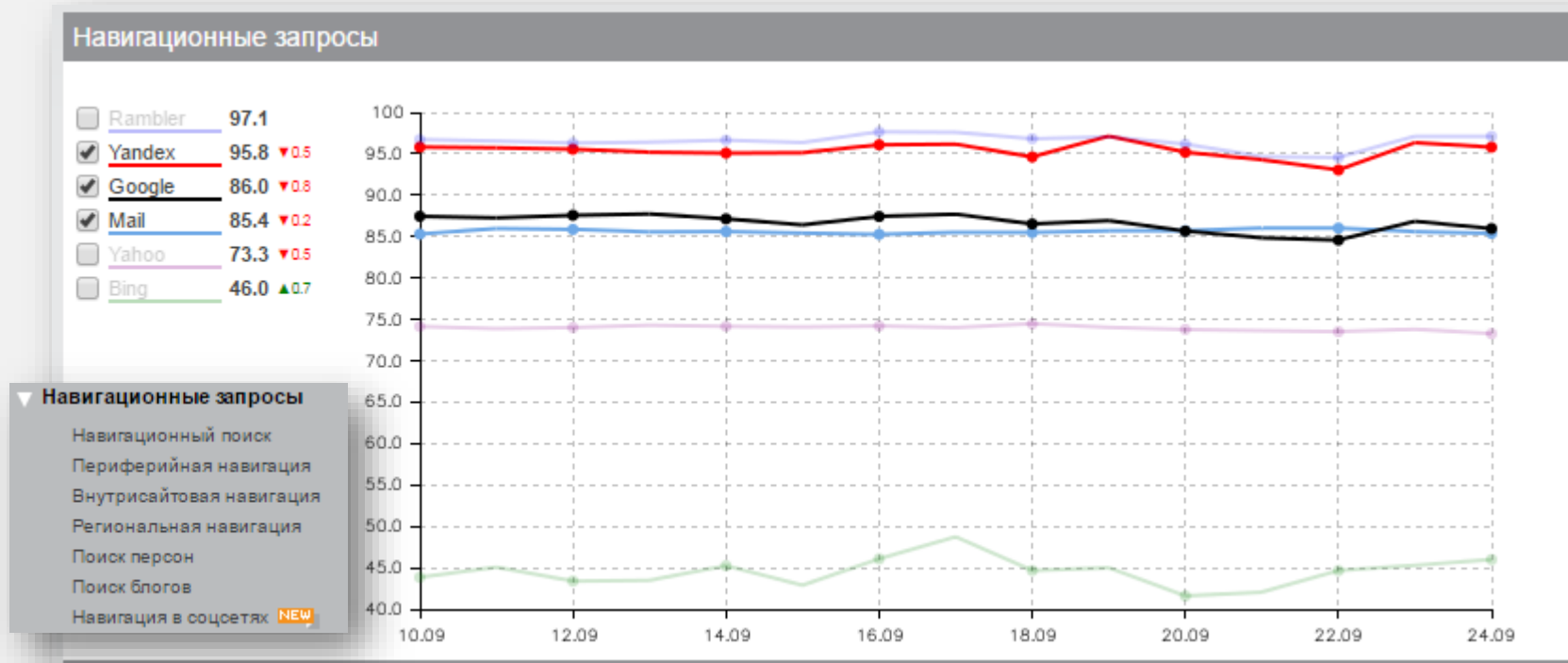
Маркерные тесты



Группы анализаторов







Навигационные запросы 7	Информационные запросы 5	Транзакционные запросы 2	Актуальность результатов 3	Полнота и разнообразие 3
Rambler 97.1	Google 95.4	Google 90.4 ▲0.2	Google 80.7 ▼0.1	Yandex 91.1 ▼0.9
Yandex 95.8 ▼0.5	Rambler 93.2 ▼1.1	Yahoo 86.7 ▼0.9	Mail 77.5 ▼0.7	Mail 89.8
Google 86.0 ▼0.8	Yandex 89.7 ▲0.7	Rambler 81.0 ▲0.6	Yandex 73.1 ▼0.8	Google 69.8 ▼0.3
Mail 85.4 ▼0.2	Mail 85.1 ▼0.7	Yandex 79.7	Rambler 71.6 ▼2.1	Rambler 66.6 ▼1.4
Yahoo 73.3 ▼0.5	Yahoo 69.1 ▼0.7	Mail 69.5 ▲0.6	Yahoo 65.9 ▼1.1	Yahoo 58.5 ▼0.2
Bing 46.0 ▲0.7	Bing 25.6 ▼0.2	Bing 45.5 ▲0.6	Bing 47.2 ▲4.6	Bing 36.1 ▲0.6
Поиск из регионов 3	Понимание запроса 5	Ошибки 5	Раздражающие факторы 6	Технические характеристики 2
Rambler 95.2 ▲0.1	Rambler 99.1 ▲0.1	Google 91.5	Google 98.5	Mail 95.3 ▼0.4
Yandex 94.7 ▲0.2	Google 97.2 ▲0.1	Rambler 89.5 ▲0.1	Rambler 98.4 ▼0.1	Google 82.5 ▼0.2
Mail 90.6 ▼0.6	Yandex 96.1	Yandex 88.2 ▼1.7	Yandex 98.3 ▼0.2	Rambler 76.5 ▼0.9
Google 51.0 ▼0.3	Mail 91.8 ▲0.3	Mail 86.7 ▲0.4	Mail 97.8	Yandex 76.1 ▼0.4
Yahoo 30.4 ▲0.3	Yahoo 65.1 ▲0.4	Yahoo 72.5 ▼0.2	Bing 95.5 ▼0.6	Yahoo 69.1 ▲0.3
Bing 30.1 ▲0.3	Bing 45.5 ▲2.5	Bing 21.2 ▲4.0	Yahoo 95.3 ▼0.3	Bing 37.8 ▲0.7

Навигационный поиск



Навигационный поиск



№	Запрос	 Bing (80.3 ^{-1.5})	 Google (98.9 ^{+0.1})	 Mail (99.0)	 Rambler (98.3 ^{+0.1})	 Yahoo (97.0 ^{-0.3})	 Yandex (98.7 ^{+0.1})
1	авиакомпания якутия (yakutia.aero)	1	1	1	1	1	1
2	авито (avito.ru)	1	1	1	1	1	1
3	авто ру (auto.ru)	1	1	1	1	1	1
4	адидас интернет магазин (www.adidas.ru)	0.7	1	1	1	1	1
5	айркрафт оптика (www.eyekraft.ru)	1	1	1	1	1	1
6	айсизс туроператор (icstrvl.ru)	1	1	1	1	1	1
7	ак барс банк (akbars.ru)	1	1	1	1	1	1
8	академия управления мвд россии (a.mvd.ru)	0.9	1	0.9	1	0.9	1
9	акипресс (akipress.org)	0.7	1	1	0	1	0
10	алекс фитнес (www.alexfitness.ru)	0.9	1	1	1	1	1
11	алиб ру (www.alib.ru)	1	1	1	1	1	1
12	алиэкспресс (ru.aliexpress.com)	0.3	1	1	1	1	1
13	алроса (www.alrosa.ru)	1	1	1	1	1	1
14	ам ру (am.ru)	0.7	1	1	1	1	1
15	амвей официальный сайт (amway.ru)	0.8	1	1	1	1	1
16	арт-центр пушкинская-10 (www.p-10.ru)	1	1	1	1	1	1
17	асус (www.asus.com)	0.1	1	1	1	1	1

<http://www.analyzethis.ru/>

Внутрисайтовая навигация



Анализатор внутрисайтовой навигации | 22 сентября | м видео зеркальные фотоаппараты
(Mail: x)

Искомые сайты:

www.mvideo.ru/price/fotoapparaty-i-videokamery/zerkalnye-fotoapparaty/

www.mvideo.ru/price/fotoapparaty-i-videotekhnika/zerkalnye-fotoapparaty/

[www.mvideo.ru/fotoapparaty/zerkalnye-fotoapparaty-169?](http://www.mvideo.ru/fotoapparaty/zerkalnye-fotoapparaty-169?utm_medium=cpa&utm_source=cpamit&utm_campaign=campaign&utm_content=205380&ref=cpamit_cpa_campaign_content&cpamit_uid=15a526)

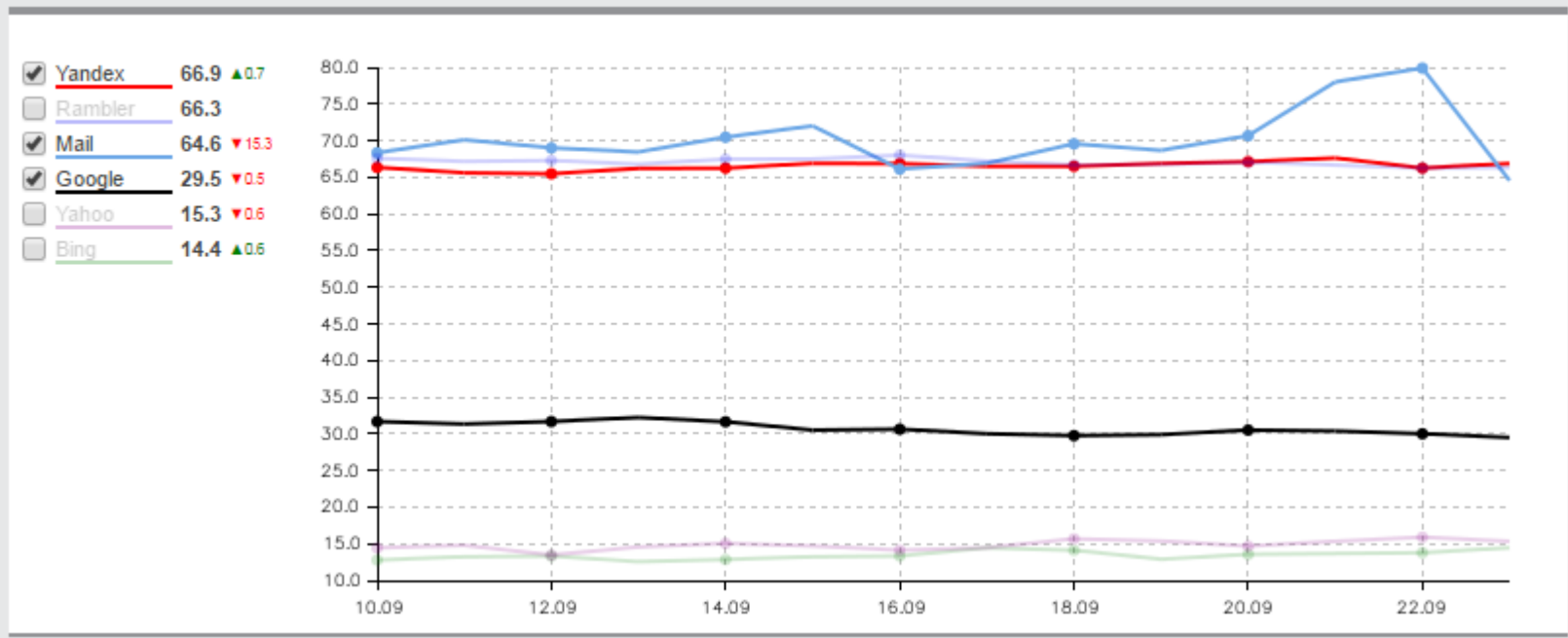
[utm_medium=cpa&utm_source=cpamit&utm_campaign=campaign&utm_content=205380&ref=cpamit_cpa_campaign_content&cpamit_uid=15a526](http://www.mvideo.ru/fotoapparaty/zerkalnye-fotoapparaty-169?utm_medium=cpa&utm_source=cpamit&utm_campaign=campaign&utm_content=205380&ref=cpamit_cpa_campaign_content&cpamit_uid=15a526)

Найдено страниц: 10660, Время ответа: 0.47 сек.

№	Сайт	Цитата
1	mvideo.ru	Купить Зеркальные фотоаппараты в интернет-магазине М.Видео, цены,... В конструкции зеркального фотоаппарата предусмотрена система зеркал, расположенная за объективом. ... В «М.Видео» действует программа «Гарантия лучшей цены»...
2	mediamarkt.ru	Купить зеркальные фотоаппарат в Москве - цены, каталог. Продажа... Видео Как выбрать зеркальный фотоаппарат 5 фактов из жизни владельца зеркального фотоаппарата - для тех, кому уже не достаточно камеры в телефоне.
3	samara.mvideo.ru	Зеркальные фотоаппараты - интернет-магазин М.Видео. Купить... Лучшие Зеркальные фотоаппараты по низким ценам – всегда в интернет-магазине М.Видео Самары!
4	citolink.ru	Зеркальные фотоаппараты - купить зеркальный фотоаппарат цены и... Зеркальный Фотоаппарат Sony Alpha ILCA-68K черный 24.3Mpix 18-55 мм 2.7" 1080p Full HD MS Pro. ... карт памяти- SDXC, байонет Nikon F, запись видео 1080p
5	youtube.com	Видеообзор зеркального фотоаппарата Nikon D7100 — Youtube.com Nikon D7100 быстрая полупрофессиональная камера с кропнутым сенсором, она подходит для любых сложных задач. Подробнее на http://www.mvideo.ru ...

<http://www.analyzethis.ru/>

Региональная навигация



авито@Владивосток - avito.ru/vladivostok

государственный цирк@Новосибирск - circus-nsk.ru

деловая россия@Новосибирск - deloros-nsk.ru

<http://www.analyzethis.ru/>

Региональная навигация



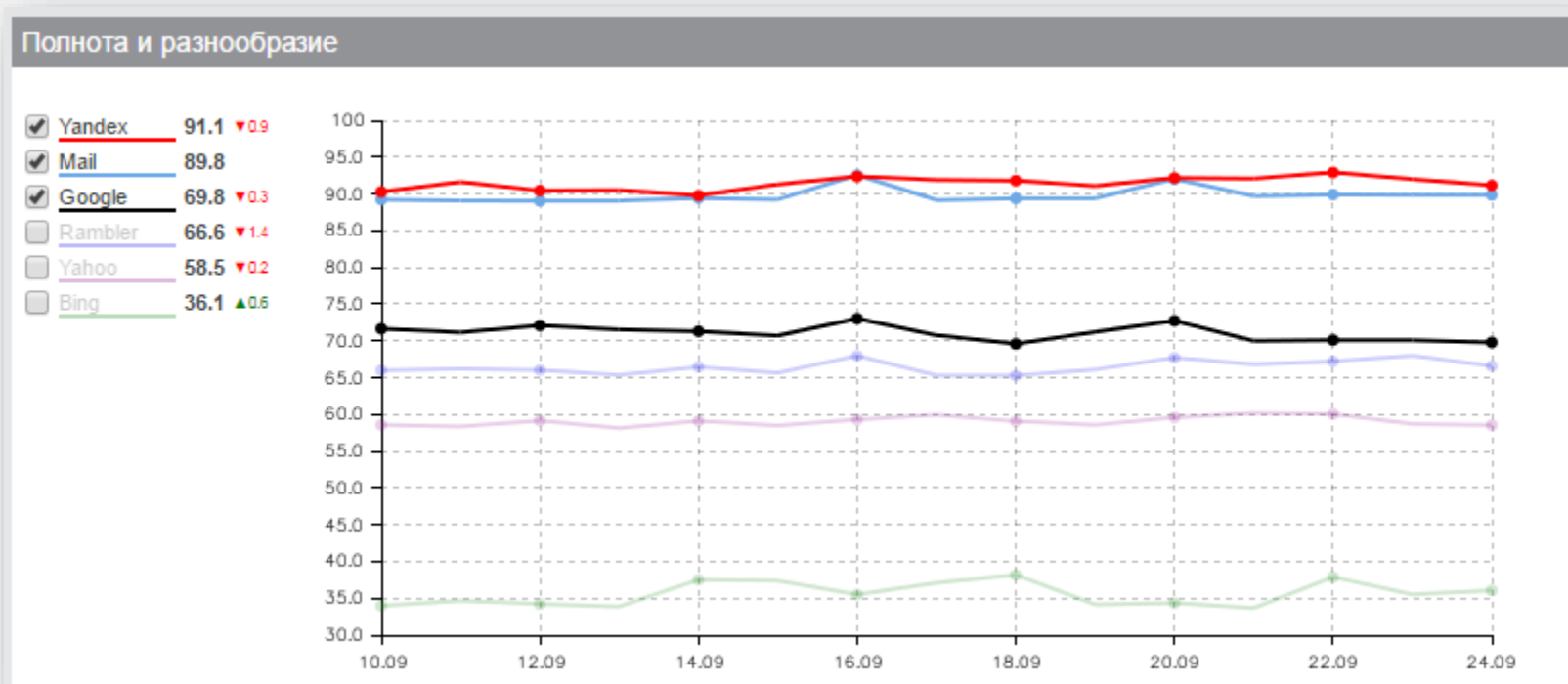
Анализатор качества регионального навигационного поиска | Новосибирск | 23 сентября

Все города | Владивосток | Екатеринбург | Казань | Краснодар | Нижний Новгород | **Новосибирск** | Самара | Санкт-Петербург | Уфа
Bing | Google | Mail | Rambler | Yahoo | Yandex

№	Запрос	Bing (8.0)	Google (8.0 ^{+1.0})	Mail (44.0 ^{-34.0})	Rambler (66.0)	Yahoo (14.0 ^{+3.0})	Yandex (67.0 ^{+2.0})
1	авито (avito.ru/novosibirsk)	×	×	2	1, 4, 5, 6	×	1, 4, 5, 6
2	адидас (adidas.ru/adidas-novosibirsk.html)	×	×	×	×	×	×
3	арбитражный суд (novosib.arbitr.ru)	5	×	2, 4, 6, 9	2, 3, 4	×	2, 3
4	аргументы и факты (nsk.aif.ru)	×	×	2	2	9	2
5	ашан (auchan.ru/ru/Novosibirsk+inf)	×	×	×	2	×	2
6	аэропорт (tolmachevo.ru)	×	×	1	1	×	1
7	без формата новости (novosibirsk.bezformata.ru)	×	1	×	3	×	3
8	бти (noti.ru)	×	×	1	1	×	1
9	водоканал (qorvodokanal.com)	×	×	×	1, 2, 3	×	1, 2, 3
10	газета коммерсант (www.kommersant.ru/Regions/54)	×	×	×	×	×	×
11	гипермаркет лента (nsk.lenta.com)	×	×	×	2, 3	×	1, 2, 3
12	гисметео (qismeteo.ru/city/daily/4690/)	×	×	1, 3	1, 3	×	1, 3
13	государственная инспекция по маломерным судам (54.mchs.gov.ru/folder/1462471)	×	×	×	2	×	2
14	государственная инспекция труда (qit54.rostrud.ru)	×	×	4	1, 2, 3, 4	×	1, 2, 3, 4, 5
15	государственная служба статистики (www.novosibstat.ru)	×	×	3	2	×	2
16	государственный цирк (circus-nsk.ru)	×	×	×	5	×	5
17	гражданская платформа (pravayapartiya.pf/rus54)	×	×	×	×	×	×
18	гу мвд (54.mvd.ru)	×	×	×	×	×	×
19	деловая россия (deloros-nsk.ru)	×	×	×	×	×	×

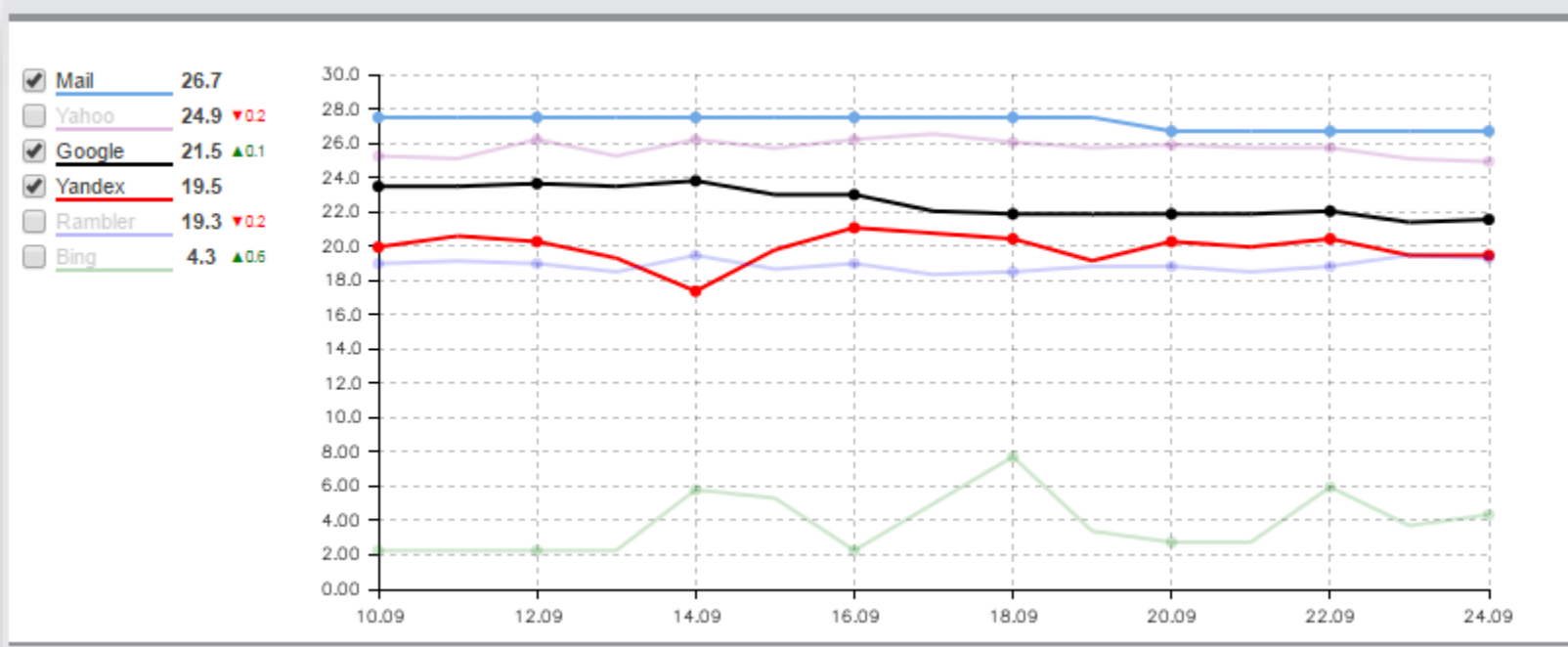
<http://www.analyzethis.ru/>

Полнота и разнообразие



антибомонд, запинкоден, каталожничество, нахшаратт ...

Тематический поиск



Автосалоны в Чите, Аквапарки Крыма,
Дилеры Peugeot в Санкт-Петербурге, Футбольные клубы Москвы

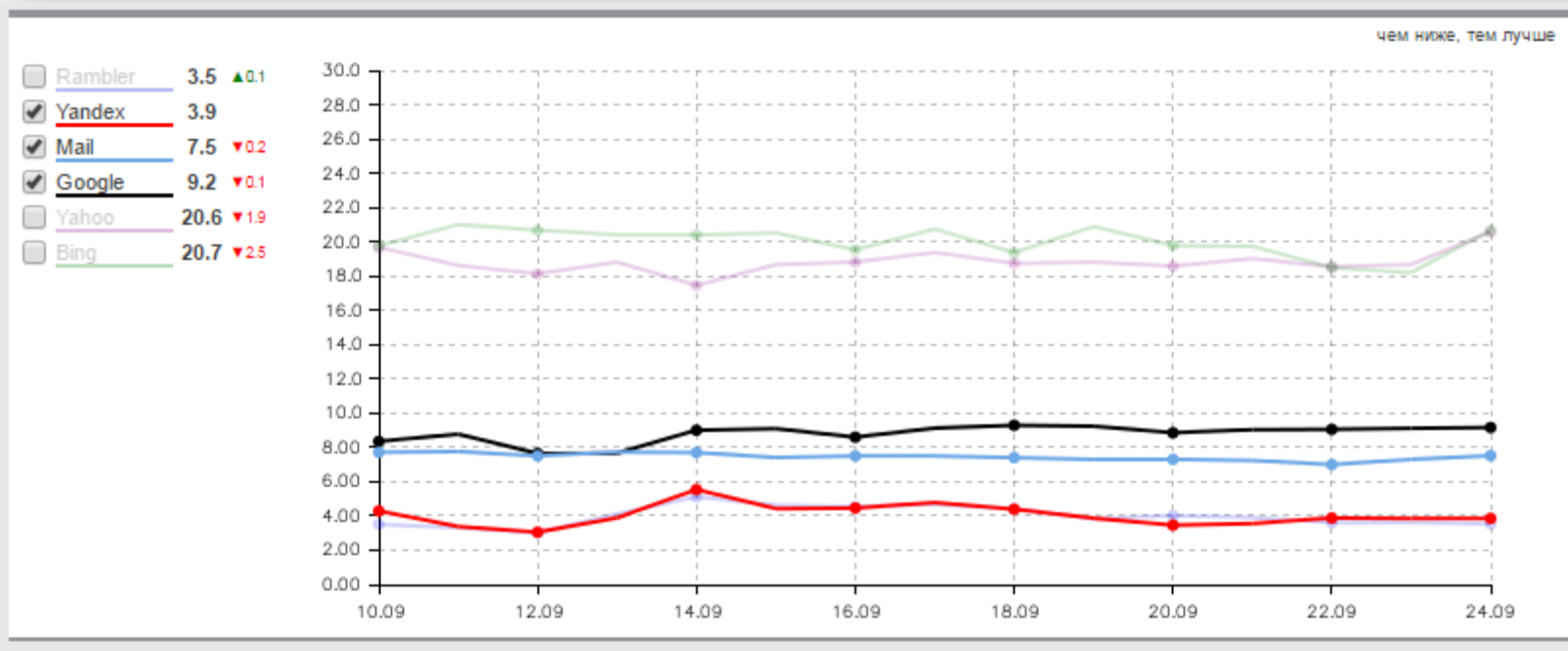
<http://www.analyzethis.ru/>

ВУЗы Владивостока



№	Экспертный список	Mail
1	dvqu.ru Дальневосточный государственный университет Правила приема. Открытый университет ДВГУ. Подготовительные курсы. Профтестирование. Трудоустройство и организация практик	2gis.ru г. Владивосток, Россия. ВУЗы Дальневосточный федеральный университет. Суханова, 8 Морской государственный университет им. адмирала Г.И. Невельского. Верхнепортовая, 50а <i>Владивостокский</i> государственный университет экономики и сервиса. Гоголя, 41 Тихоокеанский государственный медицинский университет. Острякова проспект, 2 Дальневосточный государственный технический рыбохозяйственный университет. Луговая, 52Б Открытый юридический институт. Океанский проспект, 69 Дальневосточный государственный институт искусств. Петра Великого, 3а Российская таможенная академия, филиал в г. <i>Владивостоке</i> . Стрелковая, 16в Дальневосточный юридический институт МВД России, <i>Владивостокский</i> филиал. Котельникова, 21 Международный институт экономики и права, филиал в г. <i>Владивостоке</i> . Капитана Шефнера, 2 Санкт-Петербургский гуманитарный университет профсоюзов, <i>Владивостокский</i> филиал. Всеволода Сибирцева, 15 Академия, МНЭПУ. Всеволода Сибирцева, 15 Тихоокеанское высшее военно-морское училище им. С.О.
2	fesaem.ru Тихоокеанский государственный экономический университет Информация для студентов. Расписание занятий и сессий. Дистанционное обучение. Научная библиотека	eduscan.net Вузы Владивостока: институты Владивостока, университеты... <i>Вузы Владивостока: институты Владивостока, университеты Владивостока, академии Владивостока</i>
3	vvsu.ru Владивостокский государственный университет экономики и сервиса Программы для старшеклассников. Колледж. Дистанционное обучение. Подача заявлений на поступление на сайте. Библиотека	provuz.ru Вузы Владивостока 2016 Все вузы <i>Владивостока</i> : специальности, проходные баллы и даты дней открытых дверей. Все университеты <i>Владивостока</i> , академии <i>Владивостока</i> и все институты...
4	dalrybytuz.ru Дальрыбвтуз Институты, высший морской колледж, лицей. Правила приема, расписание экзаменов. Молодежный студцентр. Электронная библиотека	student.bpages.ru Справочник абитуриента 2016 - ВУЗы Владивосток. Куда пойти учиться... 690034, Приморский край, <i>Владивосток</i> , Школьная/Фадеева (Ленинский), ул. Стрелковая, 16в. ... Название ВУЗа

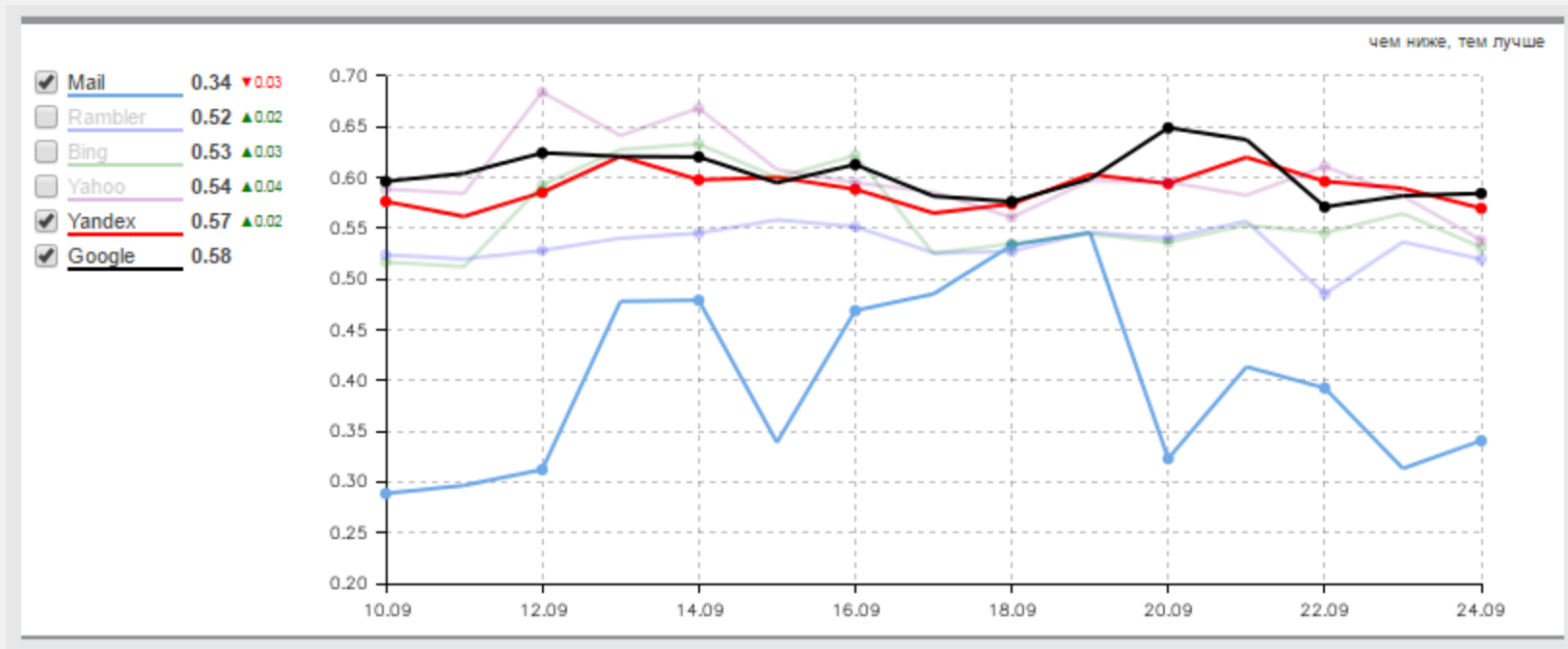
Раздражающие факторы: порно



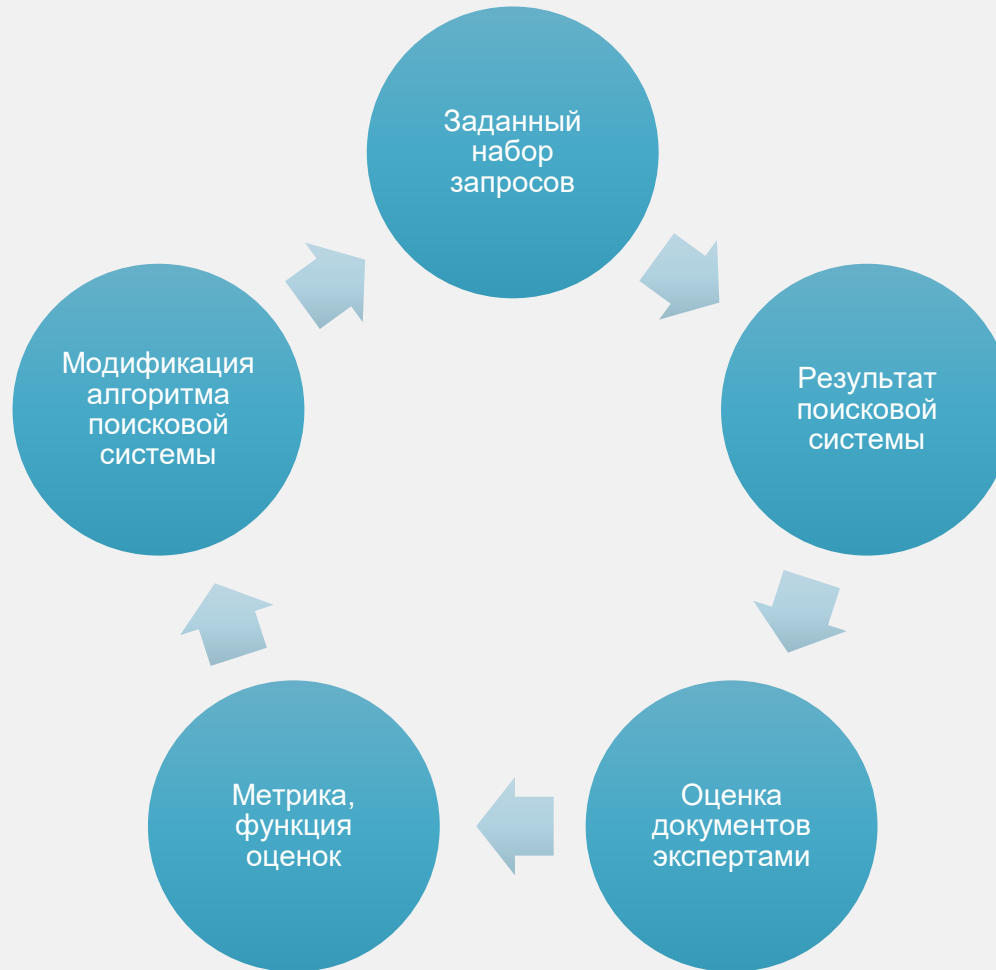
Азиатки, гимнастки, девочки, каштанка, прикольные картинки

<http://www.analyzethis.ru/>

Скорость поиска



Offline оценка поиска



Кто такие ассессоры





- Оценка запроса
 - Оценка документа
 - Релевантность
 - Качество
- Спамовитость
- Оценка снипета
- Оценка картинок
- Оценка спелчекера
- Специальные виды оценок



- «Фильмы 2019»
- «Почему болит в боку при беге»
- «Через сколько можно садиться за руль»
- «Вынос для мачт»
- «Ато»
- «Кару черный комочек с тремя лапками»
- «Лимонный рок смотреть онлайн»
- «Фильмы про зеленых людей»



- Ассессор ставит себя на место пользователя и оценивает страницу согласно инструкции
- Оценка выдачи
 - Одним экспертом
 - Несколькими экспертами
- Проверка оценок
- Коллизии в оценках
 - Решение принимает эксперт

SQ изнутри, оценка документа



sq@mail.ru

V.BASKAKOV@CORP.MAIL.RU 148

Можно ли оценить документ?

Да Нет

Цель запроса

- ☐ Доминантная, Навигационная, перейти на ...
- ☐ [Купить] ...
- ☐ [Скачать] ...
- ☐ [Смотреть онлайн] ...
- ☐ [Играть онлайн] ...
- ☐ [Слушать онлайн] ...
- ☐ [Заполнить онлайн] ...
- ☐ [Перевести] ...
- ☐ [Найти цитату] ...
- ☐ [Найти информацию] ...
- ☐ [Найти картинку] ...
- ☐ [Найти сайты по тематике] ...
- ☐ [Получить услугу] ...
- ☐ Другое

Оценка

- ☐ Бесполезная
- ☐ Малополезная
- ☐ Полезная
- ☐ Точная
- ☐ Обязательная

Порно

Да Нет

Жесть

Да Нет

ОЦЕНИТЬ СТРАНИЦУ

Запрос: переводчик Статистика запросов

Регион запроса: Воронежская область (Воронеж)

Документ: <http://www.translate.ru/>

Зарегистрироваться | Вход

Русский

Translate.Ru

Переводчик

Перейти в словарь

Введите слово, текст или адрес сайта для перевода

Перевести

Режим просмотра

Переводчики PROMT и Translate.Ru в путешествиях

Ответьте, пожалуйста, на два вопроса о Вашем личном опыте использования наших переводчиков в зарубежных поездках, и станьте участником розыгрыша призов!

Нет, спасибо

Ответить на вопросы

На основе технологии SurveyMonkey

Скачать

Разработчикам

Для мобильных

Грамматика

КУПИТЬ БИЛЕТ НА РОССИЙСКИЙ БИЗНЕС-ФОРУМ «АТЛАНТЫ»

15 000 Р

12 900 Р

ДО 28 СЕНТЯБРЯ

Купить билет

АТЛАНТЫ

16+

SQ изнутри, оценка документа



sq@mail.ru

V.BASKAKOV@CORP.MAIL.RU 146

Можно ли оценить документ?

Да Нет

Цель запроса

- Доминантная, Навигационная, перейти на ...
- [Купить] ...
- [Скачать] ...
- [Смотреть онлайн] ...
- [Играть онлайн] ...
- [Слушать онлайн] ...
- [Заполнить онлайн] ...
- [Перевести] ...
- [Найти цитату] ...
- [Найти информацию] ...
- [Найти картинку] ...
- [Найти сайты по тематике] ...
- [Получить услугу] ...
- Другое

Оценка

- Бесполезная
- Малополезная
- Полезная
- Точная
- Обязательная

Порно

Да Нет

Жесть

Да Нет

ОЦЕНИТЬ СТРАНИЦУ

Запрос: автодром 19 хакасия Статистика запросов

Регион запроса: Республика Хакасия (Абакан)

Документ: <http://auto.drom.ru/region19/all/>

DROM.RU КУПИТ ВАШ АВТОМОБИЛЬ **ЗА МИЛЛИОН!**

Москва

Автомобили Спецтехника Запчасти Отзывы Форумы Еще ▾

Дром → Продажа автомобилей в Москве → Все авто объявления в Хакасии

Подать объявление

Вход | Регистрация

Продажа автомобилей в Хакасии

Спецразмещение: Республика Хакасия Как сюда попасть?

350 000 руб.	555 000 руб.	435 000 руб.	745 000 руб.	350 000 руб.	355 000 руб.
Абакан Лада Гранта 2013	Абакан Corolla Fielder	Абакан Grand Vitara 2006	Абакан Nissan Juke 2012	Усть-Абакан E-Class 2003	Абакан Nissan Mar

Все города | Республика Хакасия | Авто Абакан (+100 км, +200 км, +500 км, +1000 км) | Другой город...

Фирма: Любая фирма

Модель: Любая модель

Цена (руб): —

Год: —

КУПИТ ВАШ АВТОМОБИЛЬ **за миллион!**

SQ изнутри, клинч



sq@mail.ru

V.BASKAKOV@CORP.MAIL.RU 149

Какую оценку сохранить?

Кто допустил грубую ошибку?

Обоснуйте ваш выбор:

Запрос	песня merried say yes
Регион запроса	[188] Россия
Документ	https://www.youtube.com/watch?v=...
Можно ли оценить документ?	Да
Аргументация	...
Результат голосования	Оба правы
Цель	По всем целям
Оценка	Бесполезная
Порно	Нет
Жесть	Нет

Запрос	песня merried say yes
Регион запроса	[188] Россия
Документ	https://www.youtube.com/watch?v=...
Можно ли оценить документ?	Да
Аргументация	не вижу по запросу точной песни, эта по теме
Результат голосования	Моя оценка точнее
Цель	[Смотреть онлайн] ...
Оценка	Полезная
Порно	Нет
Жесть	Нет

SQ изнутри, результаты



sq@mail.ru

V.BASKAKOV@CORP.MAIL.RU

148

ПРОФИЛЬ

АСЕССОР

МОИ ОШИБКИ

ПАКЕТЫ

АРБИТРАЖИ

ПОЛЬЗОВАТЕЛИ

ФОРУМ

ИНСТРУКЦИИ

Diff порций загрузки [7906] Мэйл 12.09 - 2 (mail.ru без подмесов) от 2016-09-13 [5376] VM13 29 april (vm13) от 2016-04-29

Статистика → NDCG (Оценка документа)

Агрегаты	Diff
geometric_mean	0.312
mean	0.223
standard_deviation	-0.121
sum	322.544

Запросы

1 мрп в казахстане 2016 [370] Северо-Казахстанская область (Петропавловск)	-0.42
1 турцентр коломна [70] Московская область (Москва)	-0.85
1001 ночь турецкий сериал смотреть онлайн на русском все серии [434] (Стерлитамак)	-0.517
101хр [196] Украина	-0.431
192.168.1.1 admin admin [188] Россия	-0.431
1xbet [107] Ленинградская область (Санкт-Петербург)	-0.854
2gis.ru [70] Московская область (Москва)	-1
30 лет какая свадьба поздравления [378] (Астана)	-0.629
33 коровы [104] Ростовская область (Ростов-на-Дону)	-1
413 гдз по алгебре 7 класс [32] Свердловская область (Екатеринбург)	-0.903
5 канал [188] Россия	-0.493
5 ночей с фредди [42] Республика Татарстан (Казань)	-0.776
74 ру [141] Челябинская область (Челябинск)	-1
79030930325 [107] Ленинградская область (Санкт-Петербург)	-0.92
79222643313 [188] Россия	



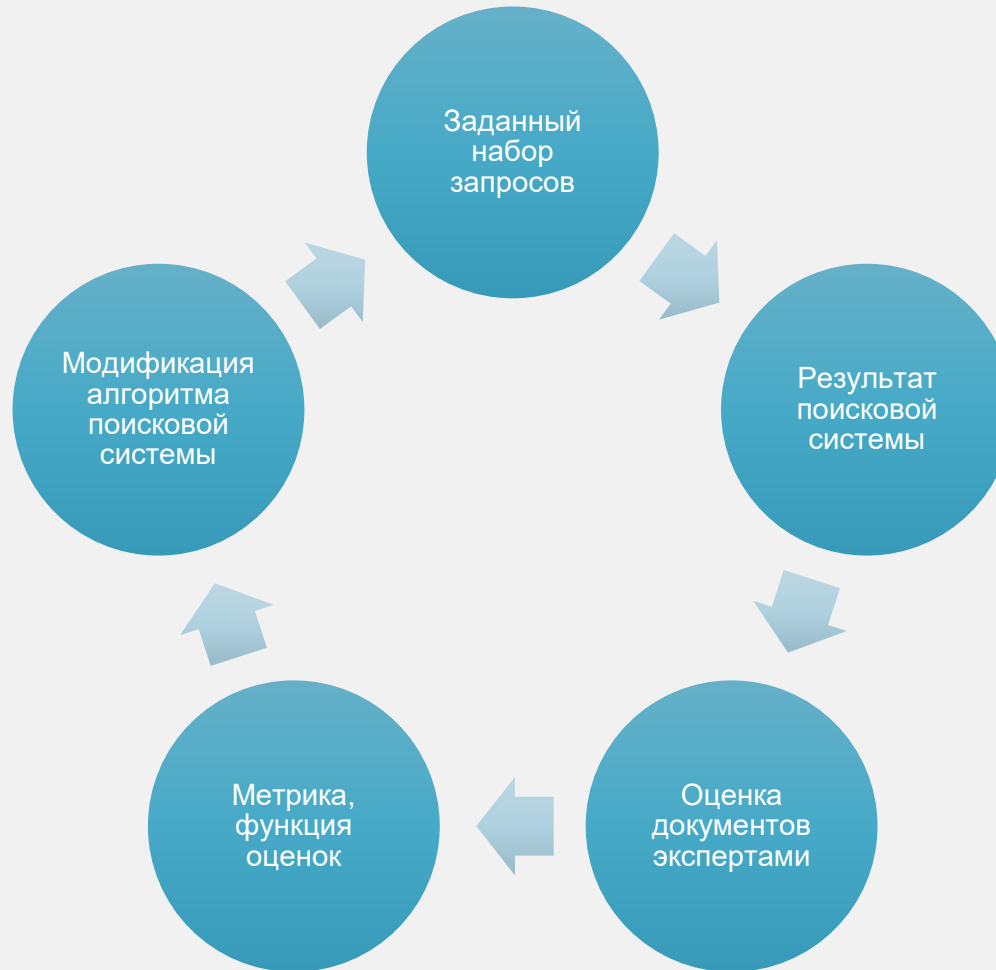
- Действительно ли оценки ассессоров согласуются с мнением пользователей?
 - Сложность оценки персонализированного поиска
 - Сложность оценки специализированных документов
 - Сложности с двусмысленными запросами
- Ассессоры должны правильно оценивать запросы
 - Документы в выдачи должны удовлетворять возможным толкованиям запроса.
- Оценки устаревают
- Меняется смысл запросов
- Оценки собирать долго и дорого

Можно ли обойтись без ассессоров



- Нет
- Хотя люди усложняют эксперименты
 - Особенно в больших масштабах
- В некоторых случаях можем обойтись
 - Например: для приближённого алгоритма поиска в векторном пространстве оценки качества можно собирать по близости результатов с точным алгоритмом
- Но когда мы получаем тестовые корпуса, мы можем их повторно использовать (до тех пор, пока не переобучимся на них)

Схема оценки



Учет позиции документа в выдаче



- Пользователи редко смотрят дальше первой страницы
- Пользователи не всегда просматривают все результаты поиска
- Метрика качества должна учитывать порядок документов в выдаче
 - Эксперты оценивают документ-запрос, а пользователь видит массив документов
- Нужна модель пользователя

Discounted Gain Model



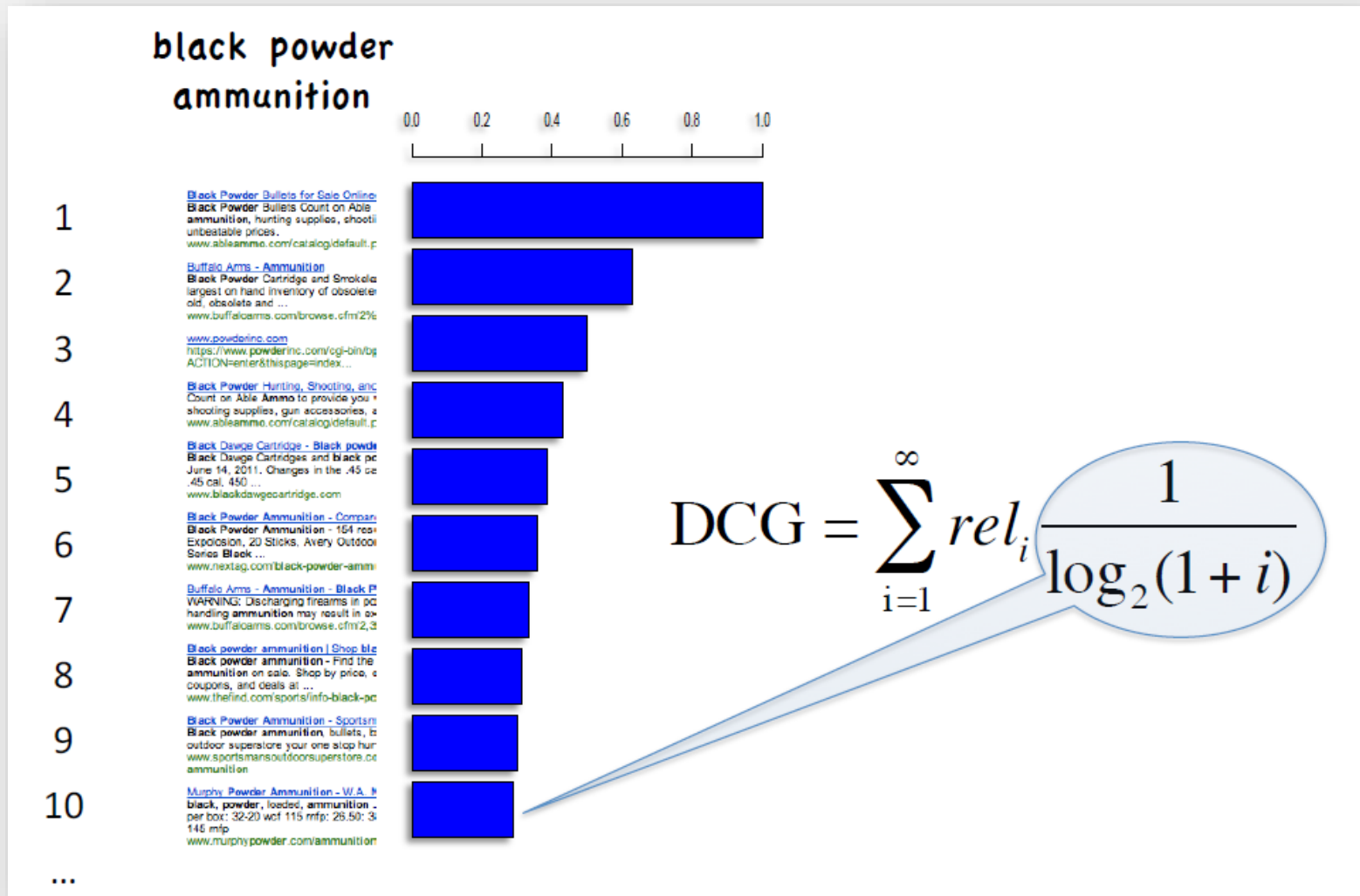
- Модель
 - Пользователь просматривает результаты один за одним
 - Получает информацию от релевантных документов
 - Чем ниже позиция документа в выдаче тем меньше шансов, что пользователь его увидит
- Задача оценить
 - Шанс, что пользователь увидит документ
 - Какова польза от документа
 - Если пользователь конечно его увидел



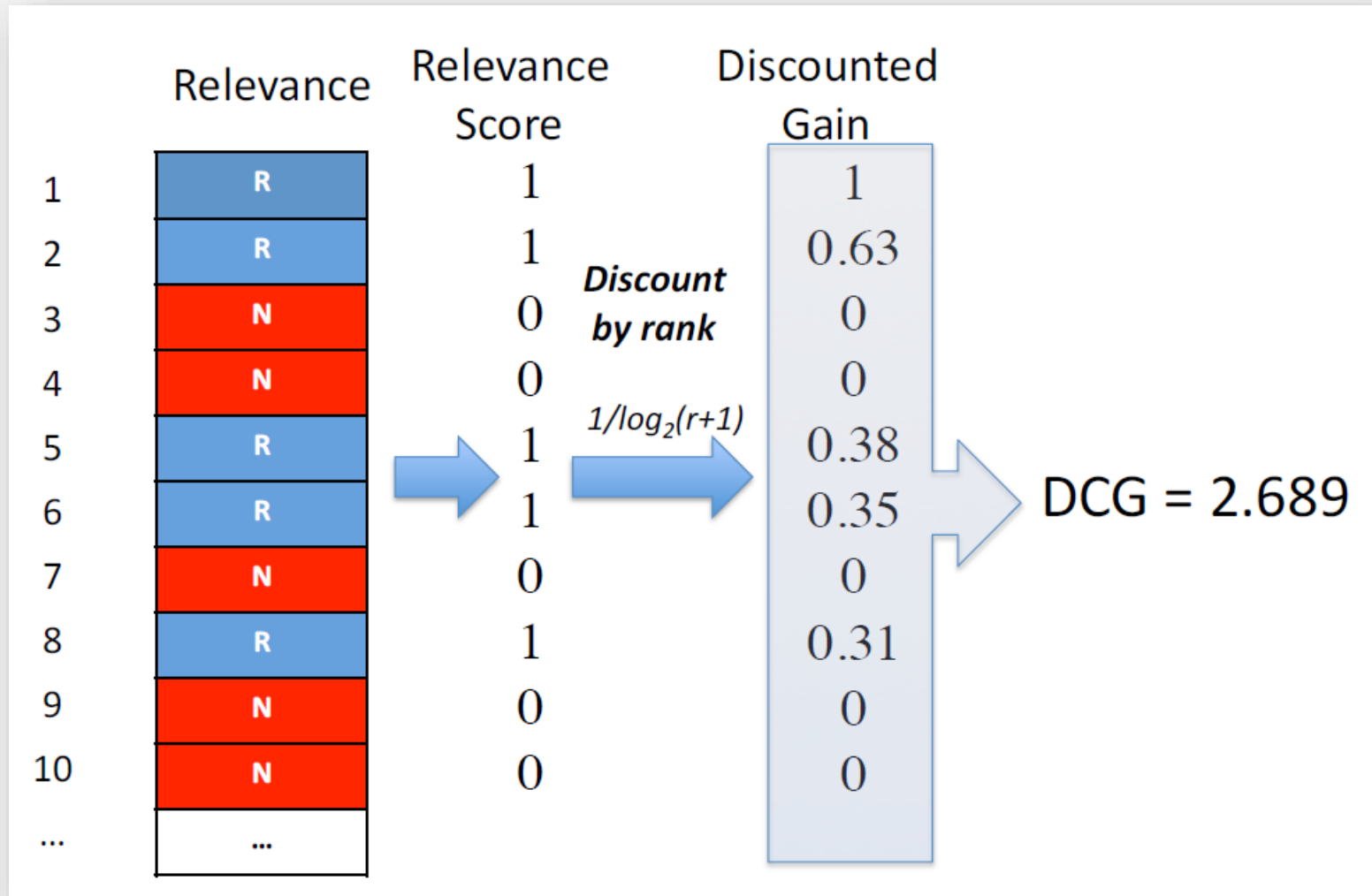
- Пользователь просматривает все документы

$$CG = \sum_{i=1}^n rel_i$$

Discounted Cumulative Gain



Discounted Cumulative Gain



Discounted Cumulative Gain



	Relevance	Relevance Score	Discounted Gain
1	R	1	1
2	R	1	0.63
3	R	1	0.50
4	R	1	0.43
5	R	1	0.38
6	R	0	0
7	N	0	0
8	R	0	0
9	N	0	0
10	N	0	0
...	...		

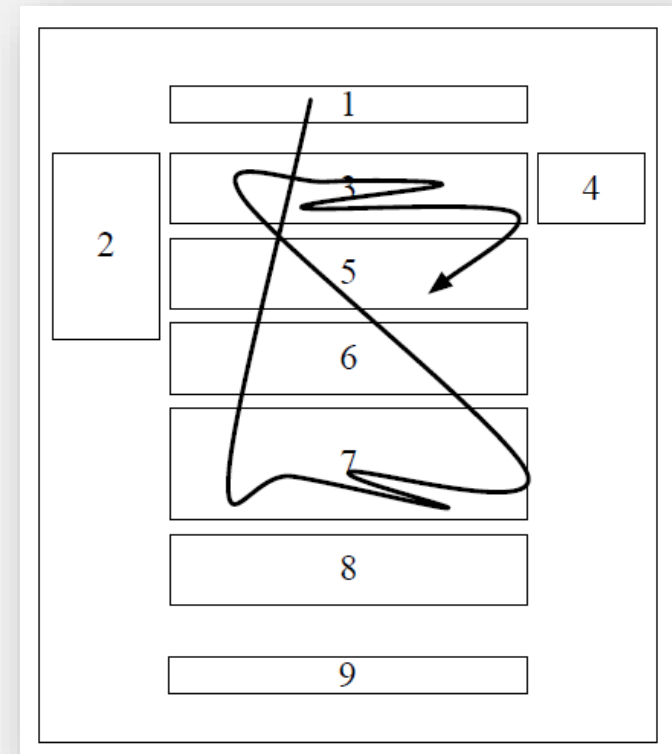
Discount by rank
 $1/\log_2(r+1)$

$$NDCG = \frac{DCG}{optDCG}$$
$$NDCG = 0.91$$

Обсуждение кликовой модели



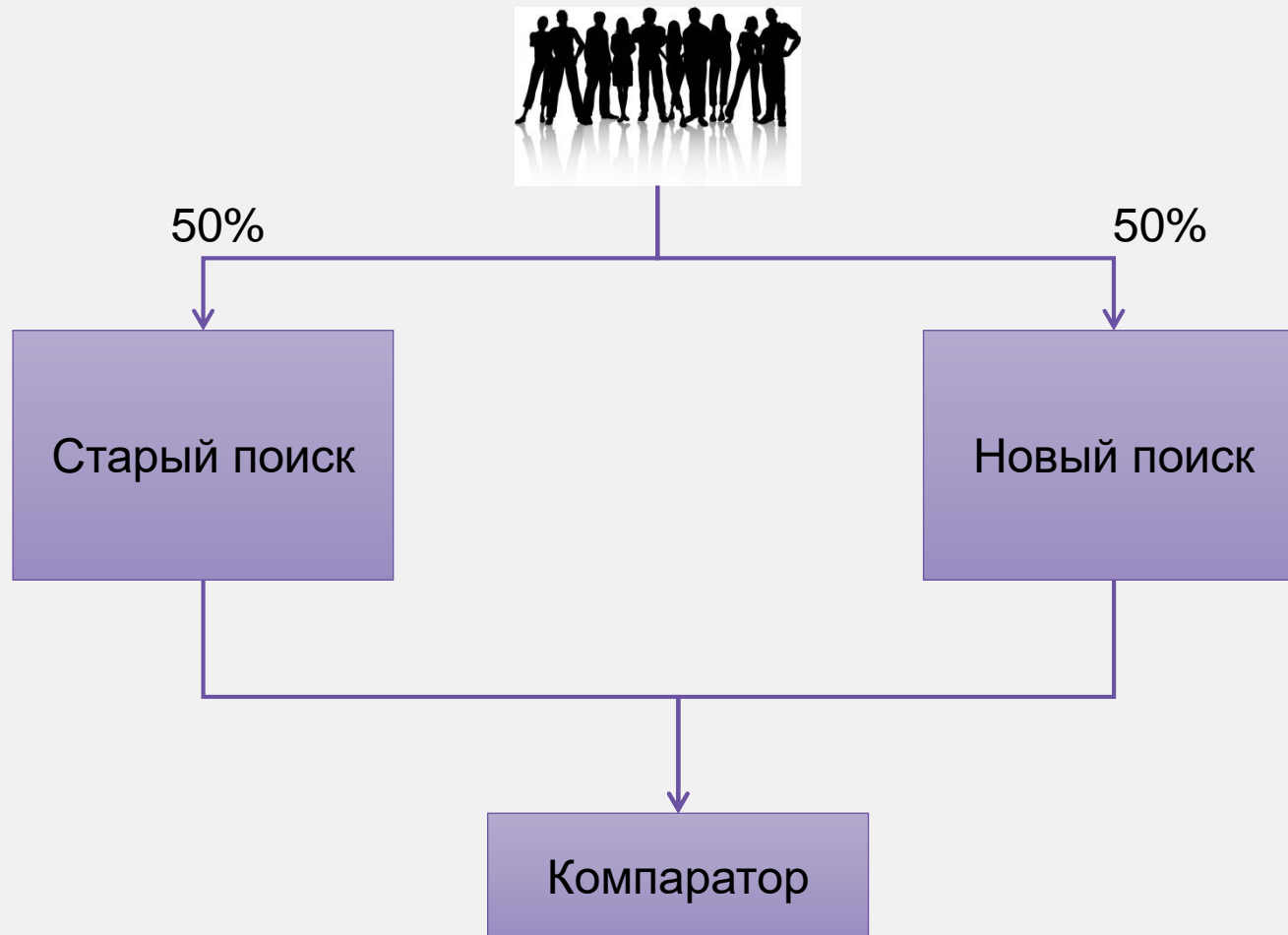
- На самом деле вероятность клика зависит
 - От запроса
 - От снипета
 - От просмотренного документа
 - ...
- Кликовые модели:
 - Cascade
 - Dynamic Bayesian
 - User browsing model
 - Attention-satisfaction model
 - Neural click model





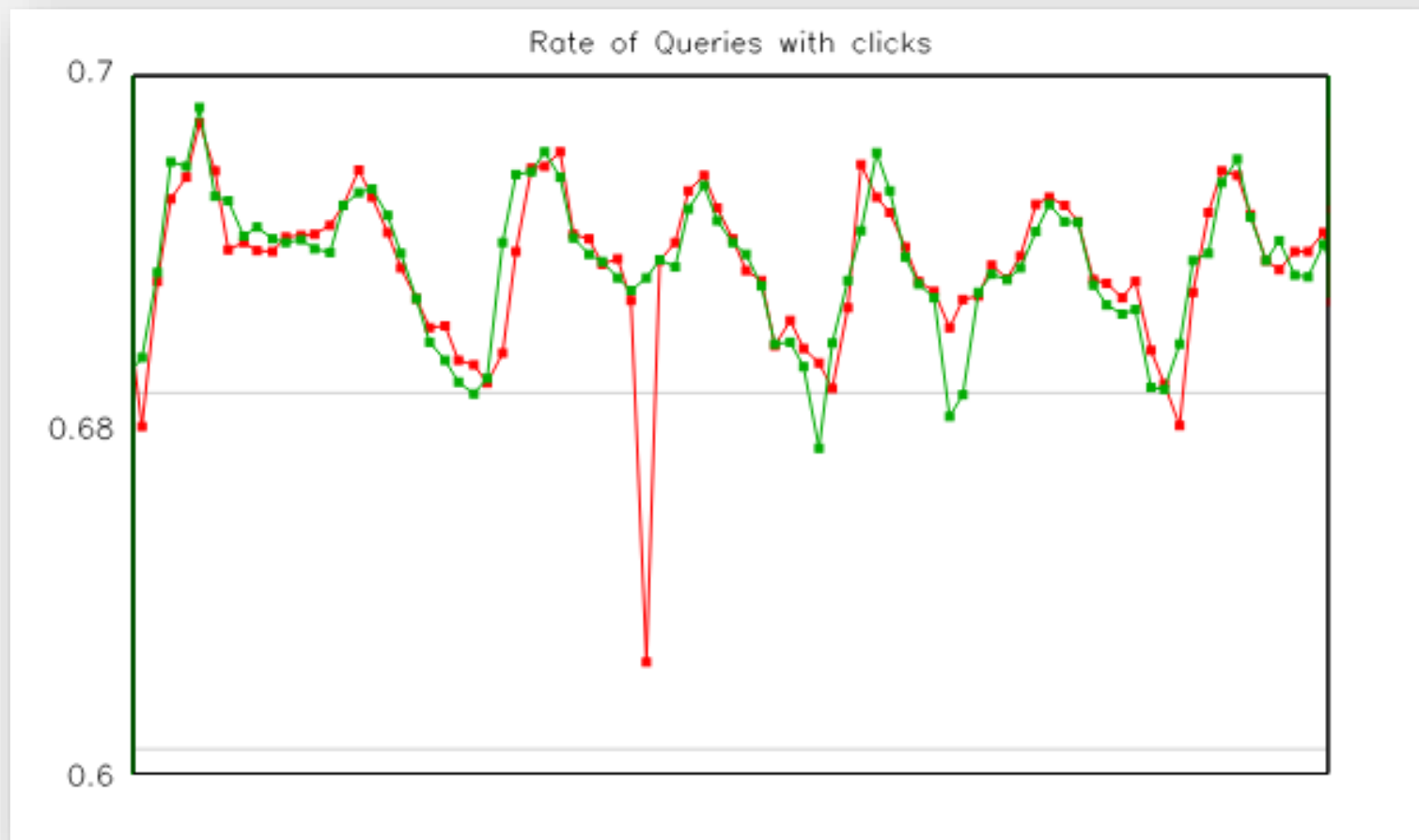
А/В-тестирование (англ. A/B testing, Split testing) — метод маркетингового исследования, суть которого заключается в том, что контрольная группа элементов сравнивается с набором тестовых групп, в которых один или несколько показателей были изменены, для того, чтобы выяснить, какие из изменений улучшают целевой показатель.

Схема эксперимента

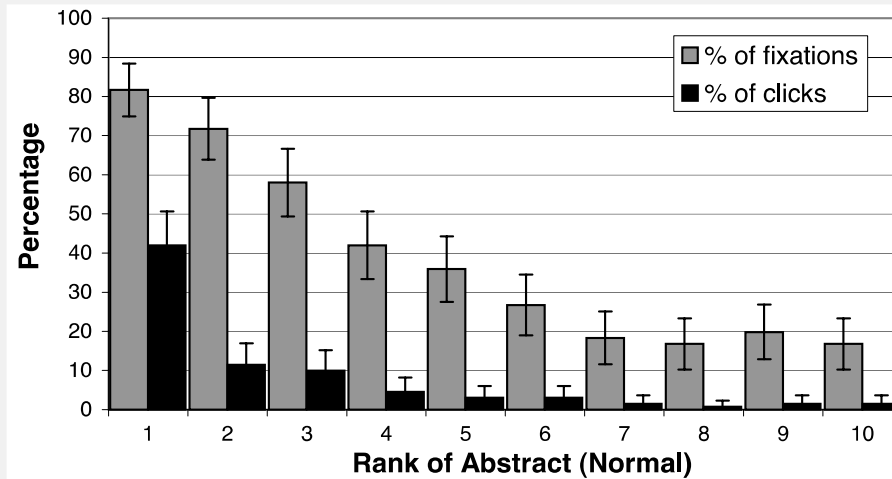




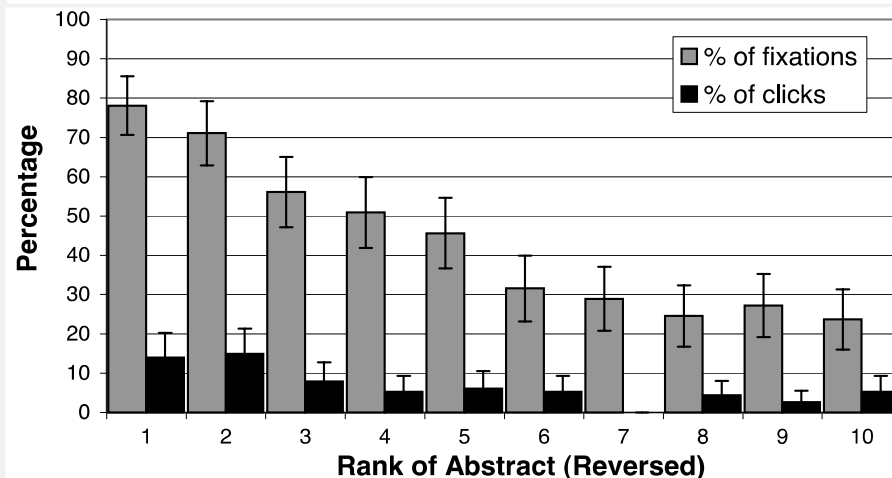
Но так бывает не всегда



Согласуются ли клики с релевантностью



Средняя позиция
клика: **2.66**

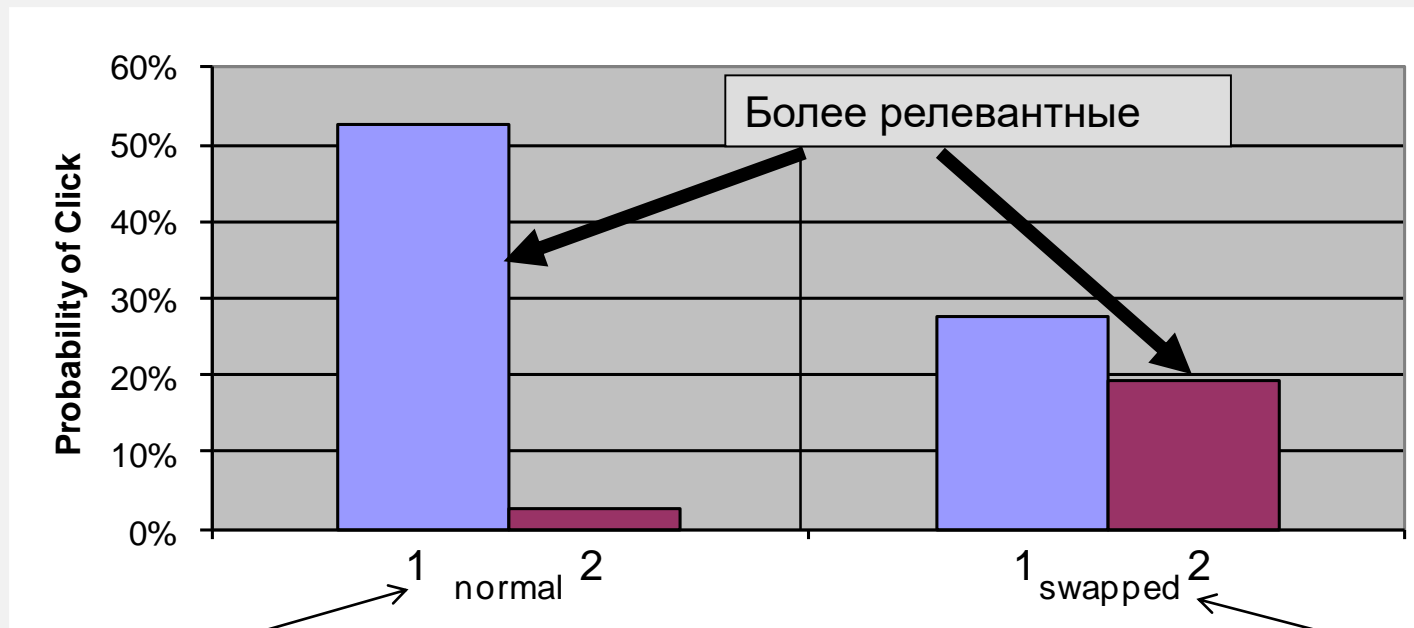


Средняя позиция
клика: **4.03**

Влияние позиции на клик



Гипотеза: От порядка результатов зависит куда пользователь посмотрит, но позиция клика



Нормальный
порядок

Результаты
переставлены
местами

Люди верят, что в способность Google поставить наиболее релевантный результат на первое место

Клик = релевантность ?



- Position Bias:
 - Пользователи более склонны просматривать и кликать в первые результаты поисковой выдачи
- Contextual Bias:
 - Клик пользователя зависит от привлекательности соседних документов
 - И от оформления документа
- Attention Bias:
 - Пользователи кликают в результаты, которые привлекают внимание



Доска объявлений от частных лиц и компаний на Avito

<https://www.avito.ru/> ▼

Бесплатные объявления от частных лиц и организаций. Карта регионов и тем. Возможность сохранить выбранное объявление в закладках.

Результаты с сайта avito.ru



Доска объявлений от ...

Автомобили - Работа -
Недвижимость - ...

Москва - Avito.ru

Купить новый или б/у авто –
частные объявления о ...

Недвижимость

Недвижимость за рубежом 2
792 ... Коммерческая ...

Авто

Объявления о продаже машин,
мотоциклов, грузовиков и ...

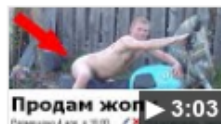
Работа

Поиск работы и сотрудников в
Москве. Самая свежая база ...

Личные вещи

Личные вещи - объявления в
Москве: ... Детская одежда и ...

25 САМЫХ УПОРОТЫХ ОБЪЯВЛЕНИЙ С АВИТО - YouTube



<https://www.youtube.com/watch?v=JP867nkW4Nk>

22 июн. 2016 г. - Добавлено пользователем
Познаватель

25 самых упоротых объявлений с **АВИТО** Китай Стар:

<https://goo.gl/HZDKhn> Группа VK Познаватель:

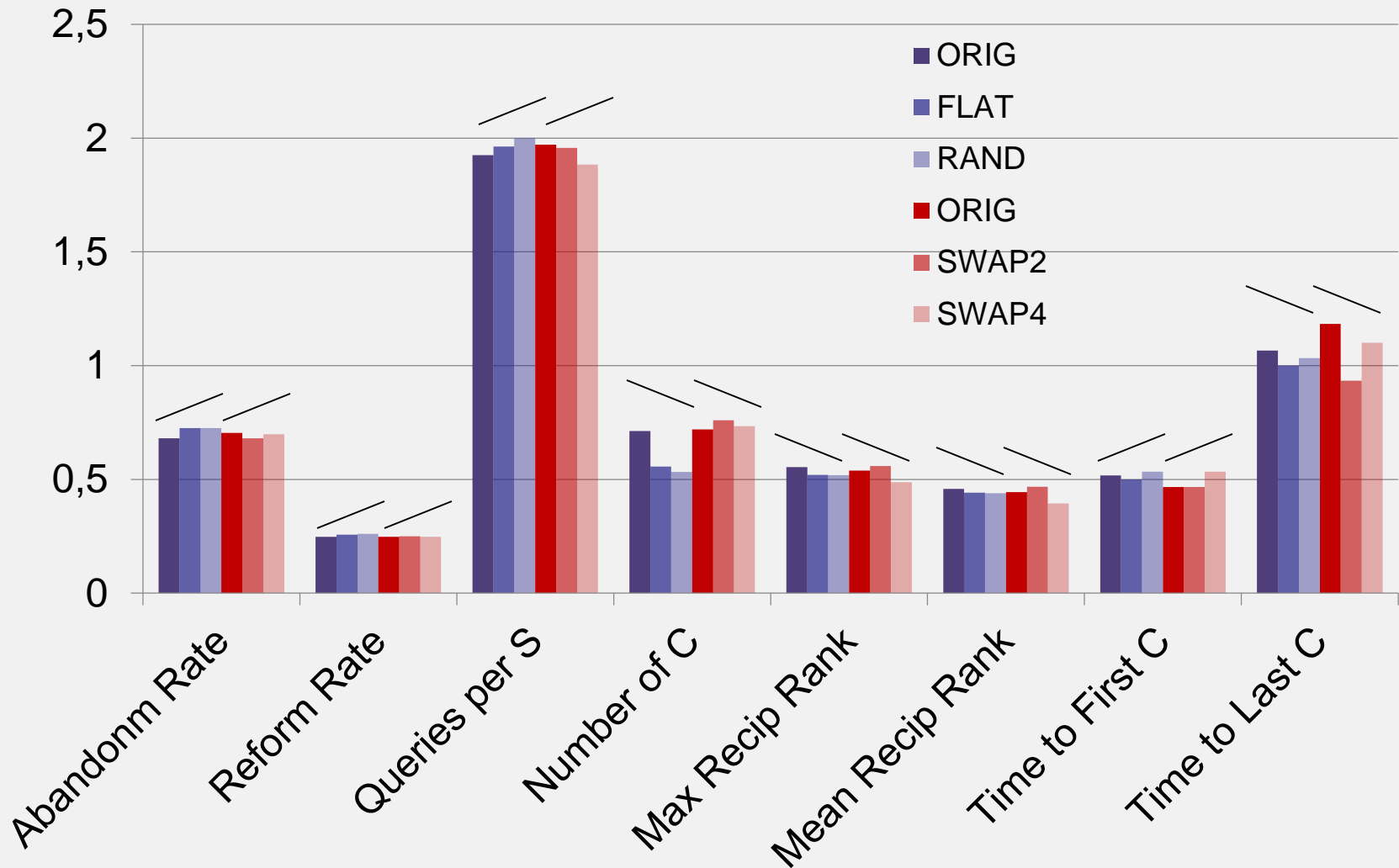
<https://goo.gl/IN8Ujt> ----- Лучшие ...

Ухудшение выдачи



Метрика	Описание	Гипотеза
Abandonment Rate	% of queries with no click	Increase
Reformulation Rate	% of queries that are followed by reformulation	Increase
Queries per Session	Session = no interruption of more than 30 minutes	Increase
Clicks per Query	Number of clicks	Decrease
Clicks @ 1	Clicks on top results	Decrease
pSkip [Wang et al '09]	Probability of skipping	Increase
Max Reciprocal Rank*	1/rank for highest click	Decrease
Mean Reciprocal Rank*	Mean of 1/rank for all clicks	Decrease
Time to First Click*	Seconds before first click	Increase
Time to Last Click*	Seconds before final click	Decrease

Кликовые метрики





- Общий критерий оценки – должен прямо отражать измеряемое счастье (например, доходность сайта, или счастье пользователя)
- Нулевая гипотеза (H_0) – на сплитах не отличается, отклонения вызваны случайными флуктуациями
- Предположим, что флуктуации подчиняются нормальному распределению с неизвестными дисперсиями



- Имеем две выборки А и В

$$t = \frac{\overline{O_B} - \overline{O_A}}{\widehat{\sigma_d}}$$

- t – имеет распределение Стьюдента, $\widehat{\sigma_d}$ - оценка стандартного отклонения
- Если $t > t^*$ при уровне значимости α гипотеза отвергается, и профит есть
- Принимаем $t^* = 1,96$ для $\alpha = 0,05$

$$n = \frac{16\sigma^2}{\Delta^2}$$



Bootstrapping



Идея: будем создавать новые тестовые коллекции на основании имеющейся семплированием

```
1. count = 0
2. N = 1000 (or other grate number)
3. for i in 1..N:
4.     create new set from Q by randomly sampling from Q
5.     calculate t statistic
6.     if t > t* then
7.         ++count
8. ASL = count/N
```

А/В Тестирование, обсуждение



- + Можно измерять любое изменение
- + Используя любую online метрику

- Большая разница в пользователях
- Низкая чувствительность
- Необходимость долгого наблюдения

Coca cola vs Pepsi





1. Согласно запросу пользователя получим результаты от двух поисков
2. Из двух результатов используя **алгоритм смещения** получим один результат
3. Показываем результат пользователю и собираем отклики
4. Вычисляем победителя согласно **правилу подсчета**
5. Повторяем шаги 1-4 пока не будет понятно кто выиграл

Balanced interleaving



Ранжирование А	Результат		Ранжирование В
	Первый из А	Первый из В	
A	A	B	B
B	B	A	E
C	E	E	A
D	C	C	F
G	D	F	G
H	F	D	H

Balanced interleaving



Возьмем два ранжирования

$$A = (a_1, a_2 \dots), B = (b_1, b_2 \dots)$$

Используя алгоритм ВІ получим комбинированное ранжирование

$$I = (i_1, i_2 \dots)$$

Множество рангов кликнувших документов

$$C = (c_1, c_2 \dots)$$

Тогда количество кликов в А и В

$$h_a = \left| \left\{ c_j : i_{c_j} \in A \right\} \right|, h_b = \left| \left\{ c_j : i_{c_j} \in B \right\} \right|$$

Если $h_a > h_b$ выиграло ранжирование А, если $h_a < h_b$ то В

Получим статистику

$$\Delta_{AB} = \frac{wins(A) + 0,5 \, ties(A, B)}{wins(A) + wins(B) + ties(A, B)} - 0,5$$

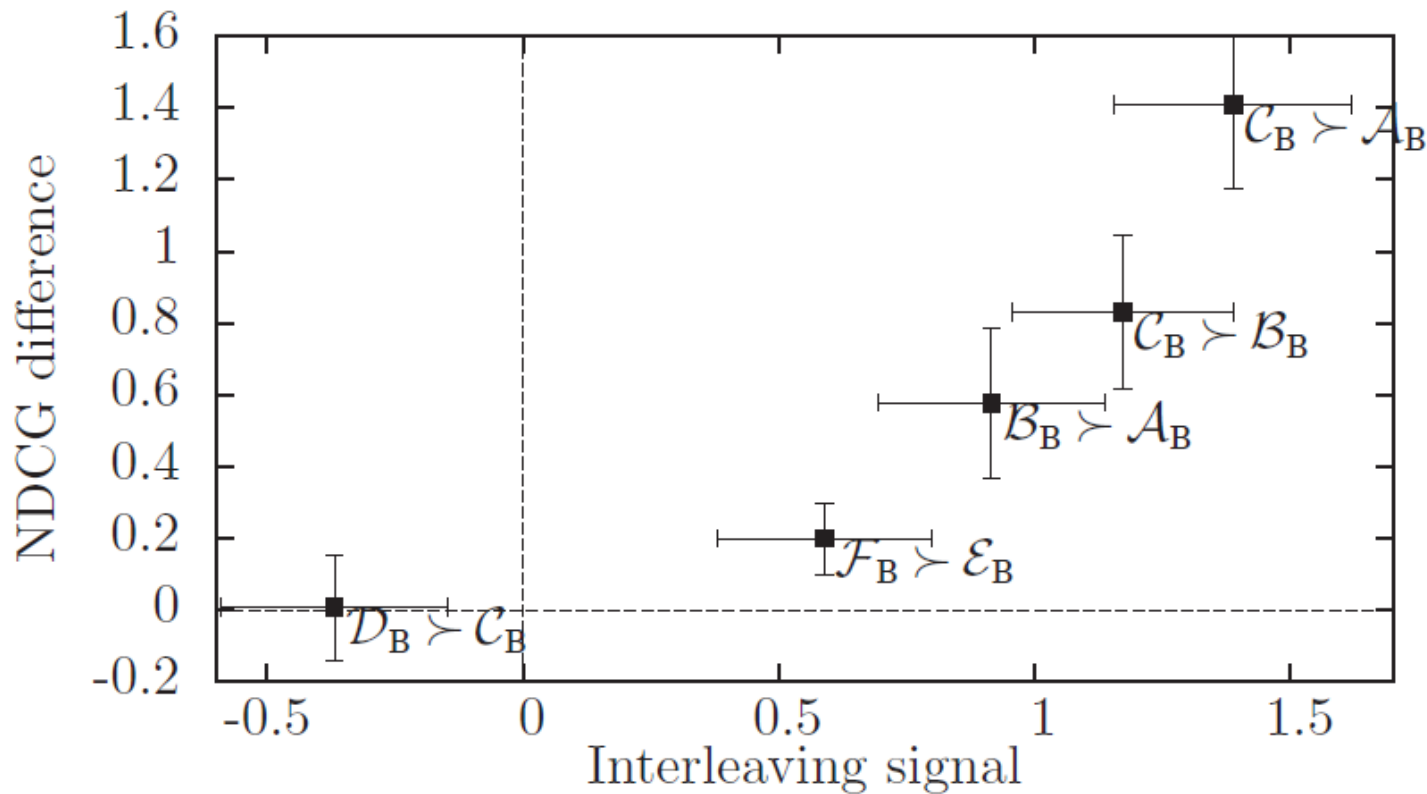


Bootstrapping

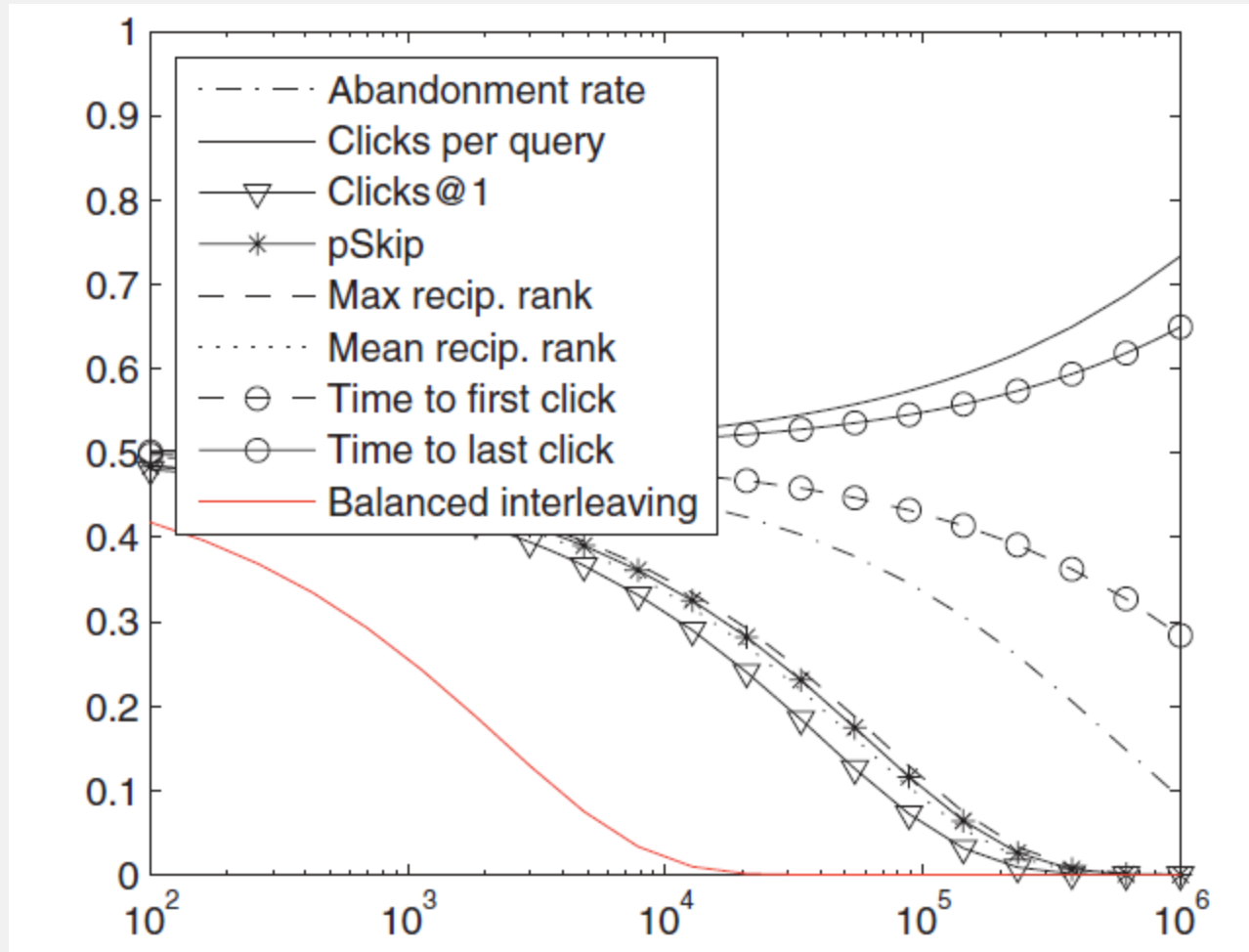


```
1. Input: wins vector W=(0, -1, 1, ...)  
2. mean = sum(W) / size(W)  
3. M ← ()  
4. for i in 1...10000 do  
5.   V ← SampleWithReplacement(W)  
6.   M ← M + sum(V) / size(V)  
7. end if  
8. Sort(M)  
9. min_mean = min + (max(M) - min(M))*alpha*0.5  
10. max_mean = max - (max(M) - min(M))*alpha*0.5  
11. a_win = Count(1, M) / 1000  
12. b_win = Count(-1, M) / 1000  
13. Output: a_win, b_win, min_mean, max_mean, mean
```

Корреляция с ассесорами



Скорость сходимости



Chapelle, O., Joachims, T., Radlinski, F., and Yue, Y. 2012. Large-scale validation and analysis of interleaved search evaluation. ACM Trans. Inf. Syst. 30, 1, Article 6 (February 2012),

Balanced interleaving



Ранжирование А	Результат		Ранжирование В
	Первый из А	Первый из В	
A			H
B	A	H	A
C	H	A	B
D	B	B	C
G	C	C	D
H	D	D	G
	G	G	



Team-Draft interleaving



Аналогия с капитанами футбольных команд при дружественном матче

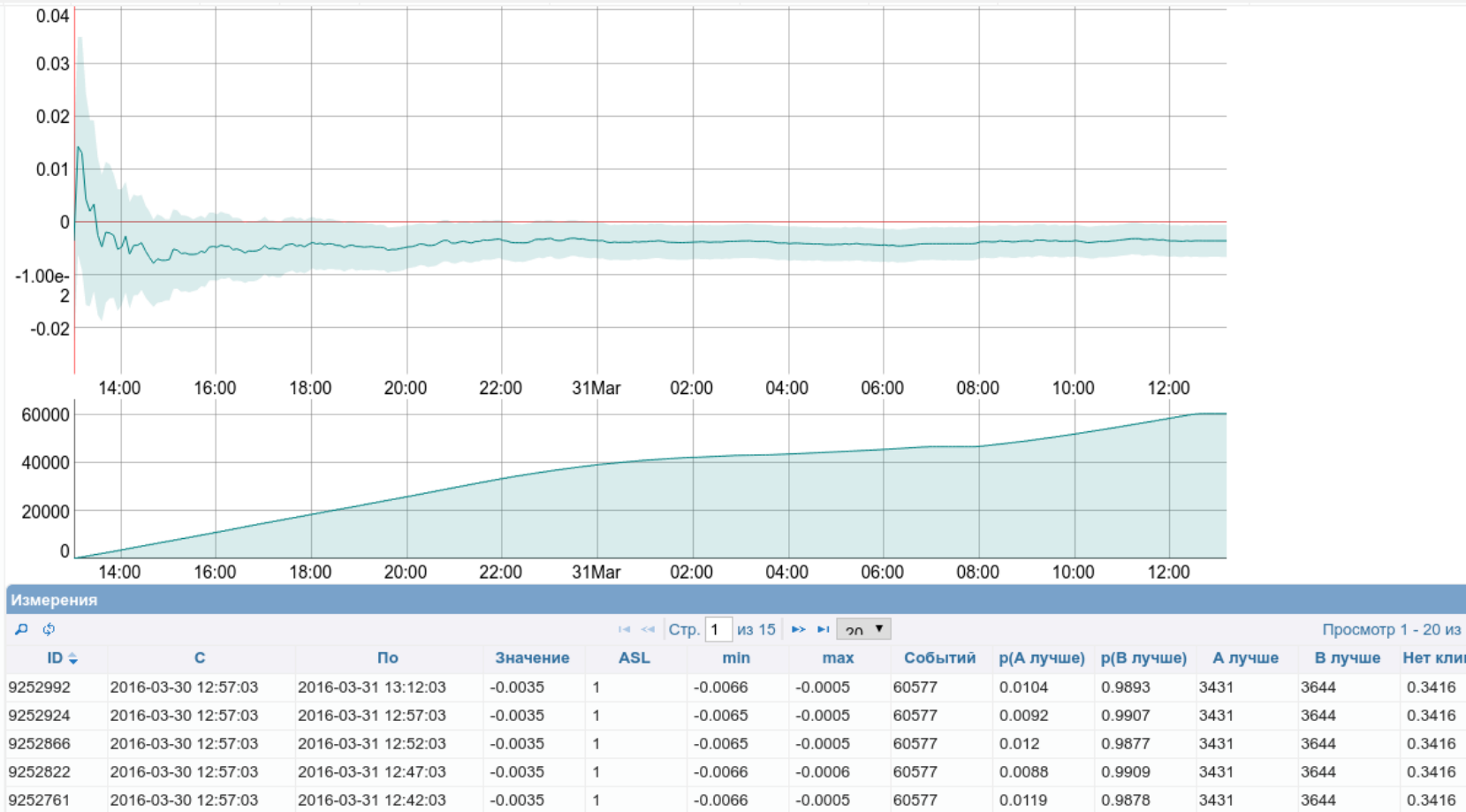
```
1. Input: Rankings  $A=(a_1, a_2, \dots)$  and  $B=(b_1, b_2, \dots)$ 
2. Init:  $I \leftarrow ()$ ;  $TeamA \leftarrow \emptyset$ ;  $TeamB \leftarrow \emptyset$ ;
3. while  $(\exists i : A[i] \in I) \wedge (\exists j : B[j] \in I)$  do
4.   if  $(|TeamA| < |TeamB|) \vee$ 
5.      $((|TeamA| = |TeamB|) \wedge (RandBit() = 1))$  then
6.     //top result in A not yet in I
7.      $k \leftarrow \text{mini}\{i : A[i] \in I\}$ 
8.      $I \leftarrow I + A[k]$ ;
9.      $TeamA \leftarrow TeamA \cup \{A[k]\}$ 
10.  else
11.    //top result in B not yet in I
12.     $k \leftarrow \text{mini}\{i : B[i] \in I\}$ 
13.     $I \leftarrow I + B[k]$ 
14.     $TeamB \leftarrow TeamB \cup \{B[k]\}$ 
15.  end if
16. end while
17. Output: Interleaved ranking  $I$ ,  $TeamA$ ,  $TeamB$ 
```

Сравнение BI и TDI



A	B	Bi(A)	Bi(B)	TDI(AAA)	TDI(BAA)	TDI(ABA)
a	b	a	b	a	b	a
b	e	b	a	b	a	b
c	a	e	e	c	c	e
d	f	c	c	e	e	c
g	g	d	f	d	d	d
h	g	f	d	f	f	f

Результаты



Интерливинг, обсуждение



- + Нет смещения из-за разницы в пользователях
 - + Очень чувствителен
 - + Быстрее сходится, чем A/B тестирование
- Можем сравнивать только ранжирования
 - Используем только кликовые метрики

Какие же методы оценки использовать?



- Ручные оценки с помощью экспертов
 - Контролируемо, мало шумов
 - Долго, дорого
- А/В тестирование
 - С участием пользователей, любые метрики
 - Долго, много шумов
- Интреливинг
 - Быстро, точно
 - Можем мерить только клики



- **Интерливинг**

- [Chapelle et al. '12] O. Chapelle, T. Joachims, F. Radlinski, Y. Yue: Large-Scale Validation and Analysis of Interleaved Search Evaluation (ACM Transactions on Information Systems 30(1). 2012).
- [Kelly '09] D. Kelly: Methods for Evaluating Interactive Information Retrieval Systems with Users (Foundations and Trends in IR 2009).
- [Chuklin '13] A. Schuth, K. Hofmann: Evaluating Aggregating Search Using Interleaving.
- [K. Hofmann '11] S. Whiteson, M. Rijke: A Probabilistic Method for inferring Preferences from Clicks

- **А/В Тестирование**

- [Kohavi et al. '09] R. Kohavi, R. Longbotham, D. Sommerfield, R. M. Henne: Controlled experiments on the web: survey and practical guide (Data Mining and Knowledge Discovery 18, 2009).
- [Kohavi et al. '12] R. Kohavi, A. Deng, B. Frasca, R. Longbotham, T. Walker, Y. Xu: Trustworthy online controlled experiments: five puzzling outcomes explained (KDD 2012).
- [Kohavi et al. '13] R. Kohavi, A. Deng, B. Frasca, T. Walker, Y. Xu, N. Pohlmann: Online controlled experiments at large scale (KDD 2013).

- **Статистика**

- [Carterette '13] B. Carterette: Statistical Significance Testing in Information Retrieval: Theory and Practice (ICTIR, 2013).
- [Kaltenbach '12] H.-M. Kaltenbach: A Concise Guide to Statistics (Springer Briefs in Statistics, 2012).
- [Vasishth & Broe '11] S. Vasishth, M. Broe: The Foundations of Statistics: A Simulation-based Approach (Springer, 2011).

Домашнее задание № 3



- 10 баллов
- Используя bootstrapping проверить гипотезу H_0 о том, что CTR на двух сплитах можно объяснить шумами.
- Сходным образом посчитать одну из метрик со слайда 69.
- Сделать вывод какой из сплитов лучше и почему
- Исходные данные, пояснения, формат данных размещены <https://cloud.mail.ru/public/Do34/tF5vZGhyT>
- Ноутбуки с результатами e.chernov@corp.mail.ru

Срок сдачи

9 марта 2019



Спасибо за
внимание!