

# Поведенческие метрики для оценки качества поиска

Протасов Сергей

Rambler

<http://www.optimization.ru/09/>, ноябрь 2009



# Обзор доклада

- 1 **Некоторые факты**
  - О том, что пользователи ленивы
- 2 **Маркерный метод**
  - Как это было и почему это было плохо
- 3 **Поведенческие метрики**
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 **Плохие запросы**
  - Плохие и трудные запросы
- 5 **Сплит-тестирование**
  - Тестирование новых фич



# Обзор доклада

- 1 **Некоторые факты**
  - О том, что пользователи ленивы
- 2 **Маркерный метод**
  - Как это было и почему это было плохо
- 3 **Поведенческие метрики**
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 **Плохие запросы**
  - Плохие и трудные запросы
- 5 **Сплит-тестирование**
  - Тестирование новых фич



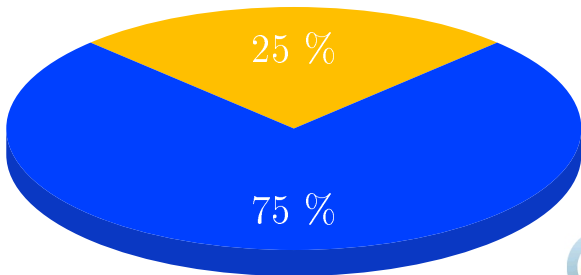
О том, что пользователи ленивы

# Факты о поведении пользователей поиска (по данным Рамблера)

Анализ выдачи до первого клика - 20 секунд

Анализ первого результата - 60 секунд

(1tm) 20 сек анализ до 1-го клика

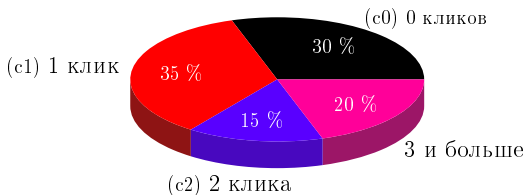


(tmr) 60 сек анализ 1-го результата



О том, что пользователи ленивы

# Распределение по кликам



метрика	число кликов	доля сессий
c0	0 кликов	30%
c1	<b>1 клик</b>	35%
c2	2 клика	15%

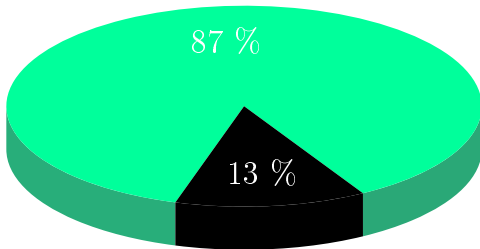


О том, что пользователи ленивы

# Факты о поведении пользователей поиска

## Посетители не ходят вглубь

первая страница



(крг) доля следующих страниц



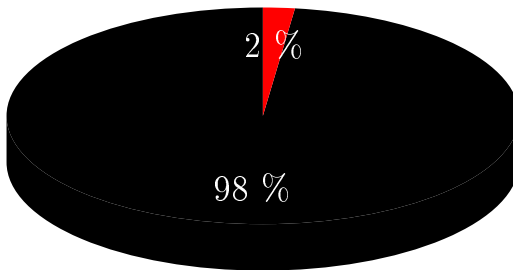
О том, что пользователи ленивы

## Факты о поведении пользователей поиска

Посетители не ходят вглубь

и не просматривают основную часть **доступных** документов

просматривается



не просматривается

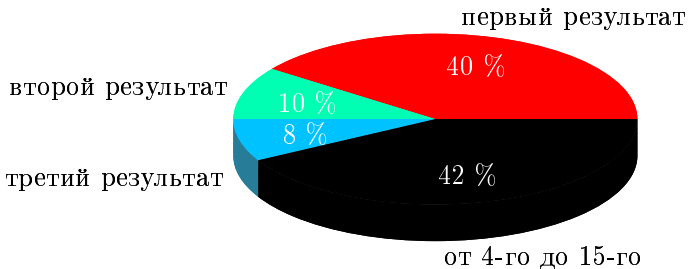


О том, что пользователи ленивы

## Факты о поведении пользователей поиска

Трафик падает очень быстро

При условии одного клика второму результату достается намного меньше, чем первому.



Интернет

Новости

Например: качество поиска

О том, что пользователи ленивы

## Факты о поведении пользователей

### Некоторые факты о поведении пользователей

- при одном клике второму результату достается **в 4 раза меньше трафика**, чем первому.
- правильный ответ с плохой аннотацией снижает трафик в десятки раз.
- посетители не прокручивают страницу браузера вниз.

Пользователи **ленивы**, это нельзя исправить, под это нужно подстраиваться



# Качество поиска в широком смысле

## Из чего складывается общее качество поиска

- Скорость выдачи, полнота индекса, скорость обновления индекса, надежность, опечатки, подсказки ...
- **Релевантность** (правильный ответ желателен на 1-2 месте + хорошая аннотация)

Маркеры оценивают релевантность результатов



О том, что пользователи ленивы

# Качество поиска в широком смысле

## Из чего складывается общее качество поиска

- Скорость выдачи, полнота индекса, скорость обновления индекса, надежность, опечатки, подсказки ...
- **Релевантность** (правильный ответ желателен на 1-2 месте + хорошая аннотация)

Маркеры оценивают **релевантность** результатов



# Обзор доклада

- 1 **Некоторые факты**
  - О том, что пользователи ленивы
- 2 **Маркерный метод**
  - Как это было и почему это было плохо
- 3 **Поведенческие метрики**
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 **Плохие запросы**
  - Плохие и трудные запросы
- 5 **Сплит-тестирование**
  - Тестирование новых фич



Как это было и почему это было плохо

## Что такое маркерный метод?

Примеры маркеров для тестирования поиска

- Запрос “xxx”, сайт ууу на 1 стр., хорошо(+1).
- Запрос “xxx”, сайт ууу **на 1 месте**, хорошо(+10).
- Запрос “xxx”, в выдаче есть фраза “zzz”, хорошо(+3).

Релевантность по маркерному методу:

Мы вводим в поиск много запросов и считаем сумму правильных ответов



Rambler

Интернет Новости

Например: качество поиска

# Маркеры: разногласия между экспертами

## Эксперты

Эксперты между собой часто имеют разногласия по поводу сайта на первой позиции

например запрос “лучший unix редактор”



**VI!**

**EMACS!**



## Посетители

А посетители могут иметь мнение, отличное от мнения экспертов.

# Сложные запросы

## Как найти людей

Как найти людей, которые способны оценивать сложные запросы?

- “snmp applet”
- “kafka svejk krusovice”
- “лучшее из оформления парилок”
- “java diagram drawing software”

и как пригласить их **поработать оценщиками** в поиске?



# Сложные запросы - это сленг

## Сложные запросы - это сленг

- Пользователи поисковика разговаривают на своём сленге
- Эксперты-оценщики их не понимают
- Сложность запросов увеличивается.

Маркерный метод не может охватить сложные запросы, это дорого и неэффективно.



# Особенности маркерного метода

## Проблемы маркерного метода

- Не обеспечивает полноты
- Не видит персонализации
- Не видит географической оптимизации
- Разногласия между экспертами
- Плохая точность

Сайты иногда умирают: требуется регулярная проверка на **актуальность**



Как это было и почему это было плохо

# Особенности маркерного метода

## Проблемы маркерного метода

Проблемы маркерного метода вынуждают нас искать новые способы измерять качество поиска.

Интересные возможности предоставляют поведенческие технологии



# Особенности маркерного метода

## Проблемы маркерного метода

Проблемы маркерного метода вынуждают нас искать новые способы измерять качество поиска.

Интересные возможности предоставляют **поведенческие технологии**



# Метрики для метрик

Прежде чем искать новые метрики качества, нам нужно решить, как мы будем их сравнивать.

## Критерии оценки метрик

- Как много нужно данных/времени/ресурсов.
- Точность (можно ли увидеть 1 процент улучшения)

далее мы увидим, что маркерный метод проигрывает поведенческим метрикам



# Обзор доклада

- 1 **Некоторые факты**
  - О том, что пользователи ленивы
- 2 **Маркерный метод**
  - Как это было и почему это было плохо
- 3 **Поведенческие метрики**
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 **Плохие запросы**
  - Плохие и трудные запросы
- 5 **Сплит-тестирование**
  - Тестирование новых фич



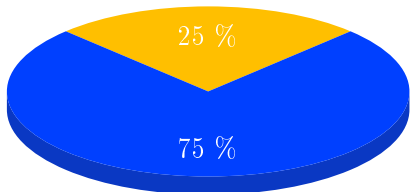
## Поведение - метрика качества

Давайте использовать поведение как метрику качества

Время на анализ поисковой выдачи и сайтов свидетельствует об удовлетворенности

- Чем меньше  $1tm$  (время до первого клика) - тем лучше.
- Если  $tmr$  меньше 15 секунд - это плохо.

( $1tm$ ) 20 сек анализ до 1-го клика



( $tmr$ ) 60 сек анализ 1-го результата



# Поведенческие метрики

## Поведенческие метрики

их довольно много:

- Клики: c0, c1, ds, click-entropy,
- Время: ltm, tmr, oldtime
- Глубина поиска: c1t1, kpg
- **Лояльность**, возвращаемость: kret

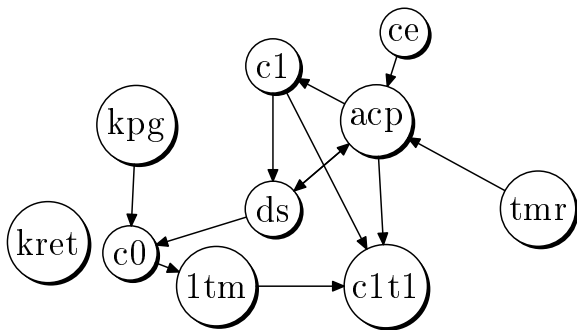
Метрики - статистические коэффициенты на основе логов действий посетителей



Какие бывают метрики?

# Многие поведенческие метрики зависимы

*и эти зависимости довольно сложные*



Rambler

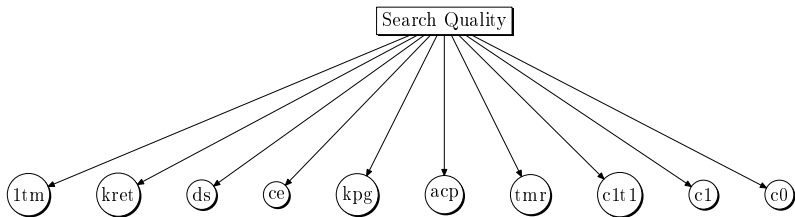
Интернет Новости

Например: качество поиска

Доказать, что работает

# Метрики: доказать, что работает

*Качество поиска - ненаблюдаемый явно параметр*



**Rambler**

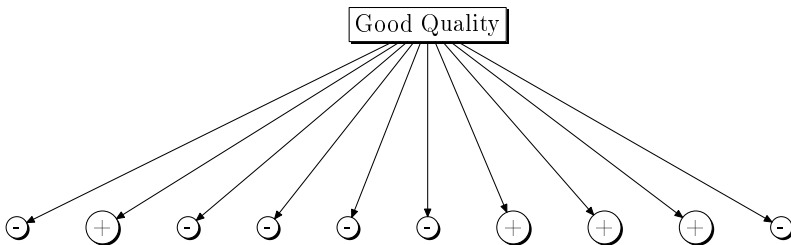
Интернет Новости

Например: качество поиска

Доказать, что работает

# Эксперимент:

*Улучшим качество, и посмотрим какие метрики изменились*



Rambler

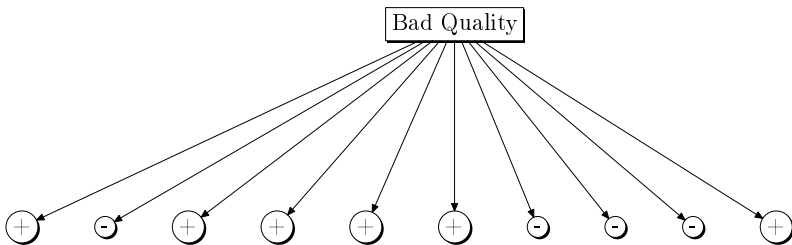
Интернет Новости

Например: качество поиска

Доказать, что работает

## Эксперимент:

Ухудшим качество ( $\frac{1}{128}$ ), какие метрики изменились в обратную сторону?



Rambler

Интернет Новости

Например: качество поиска

Доказать, что работает

# Поведенческие метрики: найти и отфильтровать

## Поведенческие метрики: найти и отфильтровать

- Мозговой штурм - придумываем метрики
- Выбираем 1/128 аудитории для опытов
- Повторяем эксперимент несколько раз
- Оставляем только те метрики, которые **зависят** от качества

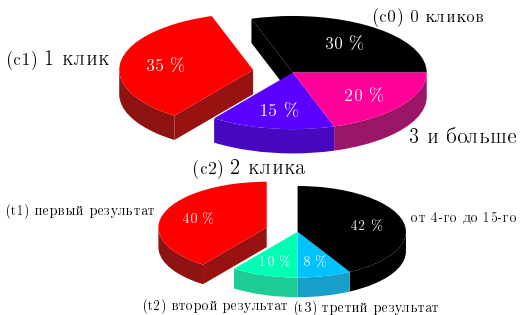
Остаются только статистически **значимые** метрики



# Удовлетворенность c1t1 - одна из хороших метрик

c1t1: доля 1-го результата при одном клике.

Среднее значение c1t1 -  $0.4 \times 0.35 = 14$  процентов.



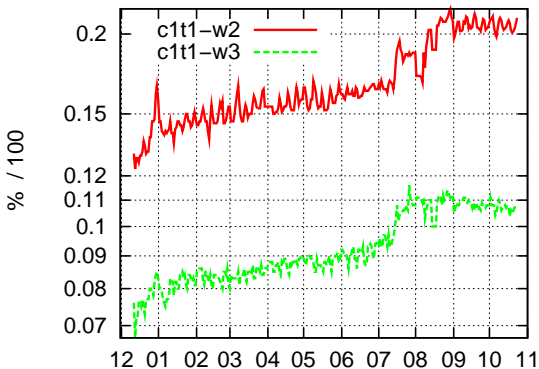
Интернет

Новости

Например: качество поиска

# Удовлетворенность c1t1 для двухсловных и трехсловных запросов - тенденция

*Тренд: удовлетворенность поиском увеличивалась в течении всего 2009 года*



Rambler

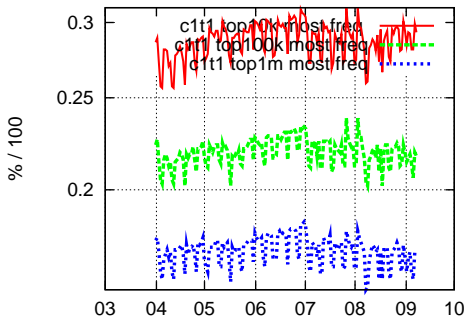
Интернет

Новости

Например: качество поиска

# clt1 - корзины запросов

- Чем более редкие запросы, тем выше сложность запросов
- Чем выше сложность запросов, тем хуже удовлетворенность поиском.



Rambler

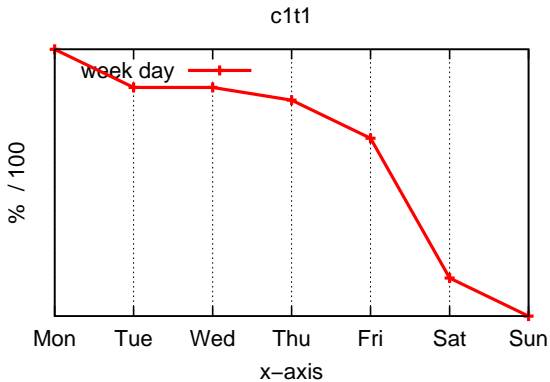
Интернет

Новости

Например: качество поиска

# Удовлетворенность c1t1 - день недели

Максимум удовлетворенности наблюдается в понедельник.



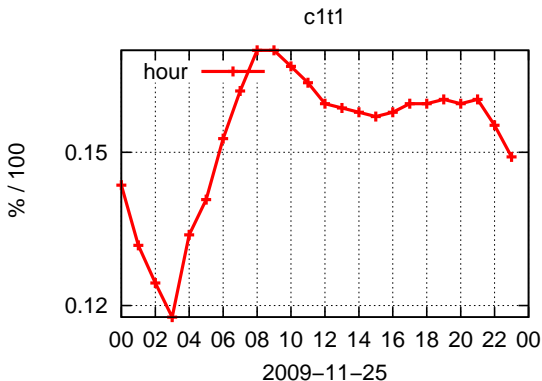
Rambler

Интернет Новости

Например: качество поиска

# Удовлетворенность c1t1 - время суток

Максимум в 8-9 часов утра, минимум в 3 ночи



Интернет

Новости

Например: качество поиска

# Качество, сложность и удовлетворенность

Метрики показывают **удовлетворенность** поиском, а не качество поиска. Мы вводим модель, которая их связывает.

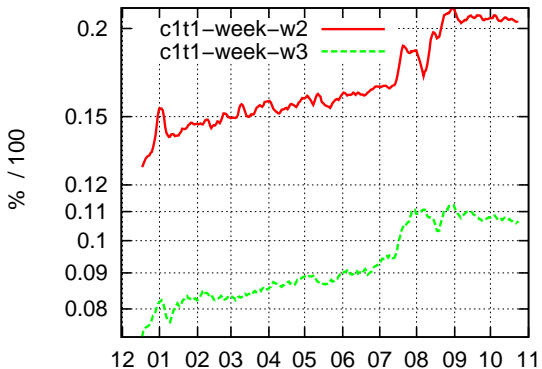
- Качество поиска - свойство алгоритмов, оборудования
- Сложность запросов - свойство посетителей
- Удовлетворенность посетителей =  $\frac{\text{Качество поиска}}{\text{Сложность запросов}}$

Наша простая модель не учитывает усталось посетителей, воздействие рекламы и другие эффекты.



# Графики: Удовлетворенность - усредняем по неделям

*Сложность запросов меняется в течении недели, поэтому усредняем по неделям.*



Рambler

Интернет Новости

Например: качество поиска

# сложность запросов растёт

## Сложность запросов растёт

Чтобы удовлетворять посетителей, поиск вынужден непрерывно мониторить поведенческие метрики.

Ежедневно находить новые сложные (плохие) запросы и улучшать их



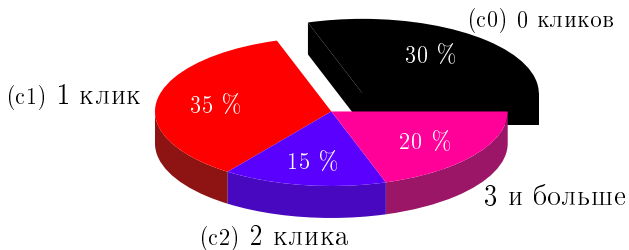
# Обзор доклада

- 1 **Некоторые факты**
  - О том, что пользователи ленивы
- 2 **Маркерный метод**
  - Как это было и почему это было плохо
- 3 **Поведенческие метрики**
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 **Плохие запросы**
  - Плохие и трудные запросы
- 5 **Сплит-тестирование**
  - Тестирование новых фич



# Практика: плохие запросы с0 (доля нулевых кликов)

Среднее значение для с0 - 30 процентов



Интернет

Новости

Например: качество поиска

# Практика: плохие запросы с0

с0 - доля нулевых кликов

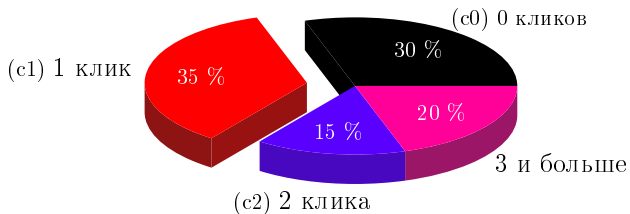
запрос	частотность	с0 %/100
как отпраздновать хеллоуин	43309	0.76
поиск в интернете	8659	0.93
фото	4386	0.62
скачать бесплатно	2663	0.76
интернет магазин	2056	0.70
скачать	1676	0.81

Таблица: Плохие запросы с0



# Практика: плохие запросы c1

Среднее значение для c1 - 35 процентов



# Практика: плохие запросы c1

c1 - доля одного клика

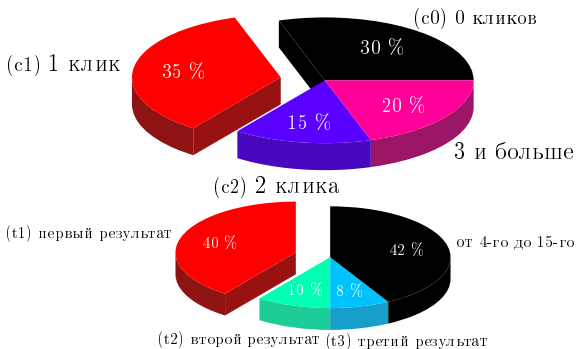
запрос	частотность	%/100
поиск в интернете	8659	0.089
официальный сайт	643	0.090
быть	282	0.000
руководство по эксплуатации	200	0.08
исаев	183	0.065
все для вас	174	0.097

Таблица: Плохие запросы c1



# Практика: плохие запросы c1t1

Среднее значение c1t1 -  $0.4 \times 0.35 = 14$  процентов



Интернет

Новости

Например: качество поиска

# Практика: плохие запросы c1t1

запрос	частотность	c1t1 %/100
поиск в интернете	8659	0.022
хеллоуин	3851	0.038
скачать бесплатно	2663	0.039
интернет магазин	2056	0.049
скачать	1676	0.040
василий лыкшин	1441	0.049

Таблица: Плохие запросы c1t1



# Практика: плохие запросы аср

Среднее место кликов по результатам (аср) - 3.3

запрос	частотность	аср
эротика	10027	4.78
игры	8998	3.83
поздравления с днем рождения	8952	5.77
фильмы онлайн	5393	4.4
скачать фильмы бесплатно	3812	4.56
музыка скачать бесплатно	3251	3.41

Таблица: Плохие запросы аср



# Практика: плохие запросы ts

ts - доля, где среднее время просмотра результатов меньше 20 секунд

запрос	частотность	ts %/100
карта европы	303	0.39
поделки из спичек	85	0.41
миит официальный сайт	65	0.44
стихи тютчева	48	0.41
кристиен стюарт фото	41	0.36
симметрия в природе	38	0.36

Таблица: Плохие запросы ts



# Практика: списки трудных запросов

Не для всех запросов удастся повысить удовлетворенность.

## Трудные запросы

запрос	частотность
фото	2725
рассказы	2278
скачать бесплатно	2034
девушки	976
тесты	764



Rambler

Интернет Новости

Например: качество поиска

# Обзор доклада

- 1 Некоторые факты
  - О том, что пользователи ленивы
- 2 Маркерный метод
  - Как это было и почему это было плохо
- 3 Поведенческие метрики
  - Какие бывают метрики?
  - Доказать, что работает
  - Графики поведения
- 4 Плохие запросы
  - Плохие и трудные запросы
- 5 Сплит-тестирование
  - Тестирование новых фич



# Внедрение подсказок и опечаток в июле 2009

## Вопрос

*Как изменилось качество поиска, удовлетворенность посетителей и сложность запросов?*

The screenshot shows the Rambler search engine interface. At the top, there are navigation tabs: "Интернет" (selected), "Новости", and "Картинки". The search bar contains the text "энтр|". Below the search bar, a dropdown menu displays several search suggestions:

- энтропия
- энтророс
- »» энтропия это
- энтропа
- энтеробиоз
- энтропия системы
- энтропия и информация
- энтропийный подход

On the left side of the interface, there is a registration form with the text "ty @rambler.ru" and a "помнить меня" (remember me) checkbox.

The screenshot shows the Rambler search engine interface with a globe icon on the left and the "Rambler" logo on the right. Below the logo, there are navigation tabs: "Интернет" (selected) and "Новости". The search bar is empty, and there is a search button to the right of the search bar. Below the search bar, there is a small text prompt: "Например: качество поиска".

# Внедрение подсказок и опечаток в июле 2009

Вспомним, что качество можно оценить через сложность и удовлетворенность

При постоянном качестве поиска и увеличении сложности запросов, удовлетворенность падает.

- Удовлетворенность посетителей =  $\frac{\text{Качество поиска}}{\text{Сложность запросов}}$
- $\text{Качество поиска} = \text{Сложность запросов} \times \text{Удовлетворенность посетителей}$



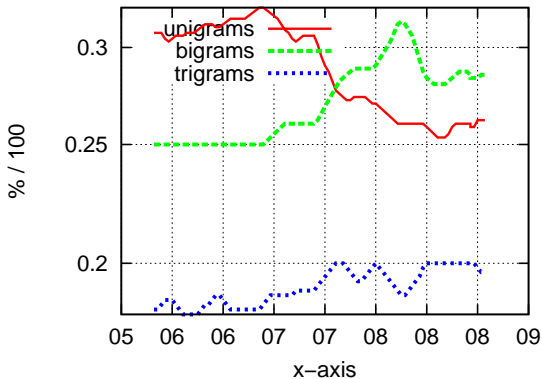
Rambler

Интернет Новости

Например: качество поиска

# Внедрение подсказок и опечаток в июле 2009

*Двухсловных запросов стало больше однословных, запросы стали **более сложные***



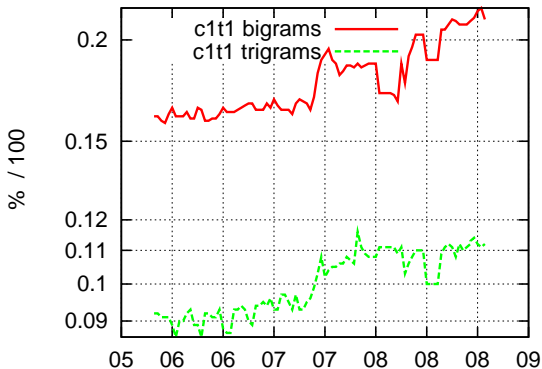
**Rambler**

Интернет Новости

Например: качество поиска

# Внедрение подсказок и опечаток в июле 2009

Несмотря на увеличение сложности запросов,  
*удовлетворенность* также увеличилась



Рambler

Интернет

Новости

Например: качество поиска

# Внедрение подсказок и опечаток в июле 2009

## Обычный “графический” способ

Удовлетворенность примерно увеличилась на 20 - 40 процентов

нам мешает множество возмущающих факторов

## Сплит-система

Удовлетворенность увеличилась на 27.7 процентов

точность улучшается в сотни раз



# Сплит-система: как внедрять новые фишки

## Последовательное увеличение аудитории новой фишки

- Выбрать критерий качества  $f(c1t1, \text{аср}, c0, c1, \text{кгг}, ..)$
- Запустить новую функциональность на  $\frac{1}{128}$  аудитории
- Проверить, что наша новая **поведенческая релевантность**  $f(c1t1, \text{аср}, c0, c1, \text{кгг}, ..)$  улучшилась относительно оставшихся  $\frac{127}{128}$
- Увеличить аудиторию до  $\frac{1}{64}, \frac{1}{32}, \dots, \frac{1}{2}, \dots, \frac{127}{128}$



Rambler

Интернет Новости

Например: качество поиска

# Сплит-система: как внедрять новые фишки

- Эффект 1: у разных посетителей - **разная** выдача поиска
- Эффект 2: мы можем увидеть улучшения в **1 процент**, потратив всего несколько дней.
- Эффект 3: мы можем **одновременно** тестировать десятки фишек.
- Эффект 4: это бесплатно!



Итоги года: что сделано

# Итоги на ноябрь 2009 года

## Итоги: декабрь 2008 - ноябрь 2009 года

- Запущены опечатки и подсказки в поисковой строке
- Новое оборудование, скорость поиска
- Новый дата-центр для повышения надежности
- На части аудитории тестируются новые фишки и алгоритмы
- Удовлетворенность увеличилась на 60 процентов



Rambler

Интернет Новости

Например: качество поиска

Итоги года: что сделано

# Итоги на ноябрь 2009 года

## Итоги: декабрь 2008 - ноябрь 2009 года

- Запущены опечатки и подсказки в поисковой строке
- Новое оборудование, скорость поиска
- Новый дата-центр для повышения надежности
- На части аудитории тестируются новые фишки и алгоритмы
- **Удовлетворенность увеличилась на 60 процентов**



Rambler

Интернет Новости

Например: качество поиска

# Советы оптимизаторам

## Советы поисковым оптимизаторам

- оптимизируйте только редкие и многословные запросы, где пользователи **неудовлетворены** первыми результатами.
- бойтесь нецелевой аудитории.



Rambler

Интернет Новости

Например: качество поиска

Thank you

Всем спасибо!

;P



*Задайте мне ваш поисковый вопрос...*

**Сергей Протасов.**

Пишите мне: [s.protasov@rambler-co.ru](mailto:s.protasov@rambler-co.ru)