

ЯНДЕКС

Метрика^β

Часть 8:

**Дополнительные возможности:
регулярные выражения.**

**Метрика бета-версия, отличия и логика
работы**

Спикер: Бочарникова Валерия

Что такое регулярные выражения?

Регулярные выражения (англ. regular expressions) — формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов.

Общая задача механизма регулярных выражений - находить или не находить совпадения строки или ее части с шаблоном.

Регулярные выражения упрощают работу с целями

При создании целей в Метрике можно использовать регулярные выражения. Они обеспечивают условиям гибкость — задав шаблон, вы можете объединить все похожие целевые страницы или, наоборот, выделить среди них только некоторые нужные.

- Общие
- Код счётчика
- Вебвизор
- Цели (5)
- Фильтры

Цели

Вы можете определить 2
— какое количество про
— один или несколько

Добавить цель

№	Название цели
1.	Заявка
2.	Посетители с
3.	
4.	
5.	

Добавить цель

Название

Ретаргетинг

Условие: **Посещение страниц** JavaScript событие Составная цель

— количество страниц (в том числе переход по внешней ссылке, загрузка файла)

Добавить условие

№	Условие	URL
1.	url: содержит	<input type="text"/>

- url: содержит
- url: совпадает
- url: начинается с
- url: регулярное выражение

	url: содержит «/land-two/thankyou.html»	7744846		
	url: содержит «/land-star/thankyou.html»	7744851		

Регулярные выражения применяются как для отдельных целей типа URL страницы, так и для сложных составных целей.

Мы начнем с изучения простейших регулярных выражений.

Поскольку регулярные выражения используются для работы со строками, мы начнем с наиболее распространенной задачи — соответствия символов.

Соответствие символов

Большинство букв и символов соответствуют сами себе.

Например, регулярное выражение **test** будет в точности соответствовать строке **test** (Вы можете включить режим без учета регистра, что позволит этому регулярному выражению также соответствовать **Test** или **TEST**, но об этом позже).

Из этого правила есть исключения; некоторые символы это специальные **метасимволы**, и сами себе не соответствуют.

Вместо этого они указывают, что должна быть найдена некоторая необычная вещь, или влияют на другие части регулярного выражения, повторяя или изменяя их значение.

Большая часть нашей лекции посвящена обсуждению различных метасимволов и тому, что они делают.

Вот полный список метасимволов

. ^ \$ * + ? { [] \ | ()

Первые метасимволы, что мы рассмотрим это **[и]**. Они используются для определения класса символов, являющегося набором символов, с которыми вы ищите совпадение. Символы могут быть перечислены по отдельности, или в виде некоторого диапазона символов, обозначенного первым и последним символом, разделенных знаком '-'.
[и]

Например, **[abc]** будет соответствовать любому из символов a, b или c; это то же самое, что выражение **[a-c]**, использующее диапазон для задания того же множества символов. Если вы хотите сопоставить только строчные буквы, **[a-z]** будет иметь вид **[a-z]**.

Метасимволы не активны внутри классов. Например, **[akm\$]** будет соответствовать любому из символов 'a', 'k', 'm' или '\$'. Знак '\$' это обычно метасимвол (который обозначает конец строки), но внутри класса символов он лишается своей особой природы.

Для того, чтобы находить соответствие символам вне этого класса, в начале класса добавляется символ '^'. Например, выражение **[^5]** соответствует любому символу, кроме '5'.

Пожалуй, наиболее важным является метасимвол обратной косой черты \.

За бэкслешем могут следовать различные символы, обозначающие разные специальные последовательности.

Он также используется для экранирования метасимволов, чтобы их можно было использовать в шаблонах; например, если нужно найти соответствие [или \, для того чтобы лишить их своей особой роли метасимволов, перед ним нужно поставить обратную косую черту: \[или \\.

Метасимвол `'.'` - соответствует всем символам, кроме символа новой строки, часто используется там, где вы хотите сопоставить «любой символ».

Повторяющиеся вещи

Первый метасимвол для повторения это *****. Он указывает, что предыдущий символ может быть сопоставлен ноль и более раз, вместо одного сравнения.

Например, **ca*t** будет соответствовать
ct (0 символов a),
cat (1 символ a),
caaat (3 символа a), и так далее.

Движок регулярных выражений не позволяет проводить ему сопоставление более 2 млрд. символов 'a'.

Другой метасимвол повторения это **+**, повторяющий последовательность сравнения один или более раз.

Обратите особое внимание на разницу между ***** и **+**.

***** требует соответствия необходимой части ноль или более раз, то есть повторяемое может и не присутствовать вовсе, а **+** требует, по крайней мере одно вхождение.

Для аналогичного примера **sa+t** будет сопоставляться **cat** или, например, **саааt**, но никак не **ct**.

Утверждения нулевого размера

|
Соответствует оператору **ИЛИ**. Если A и B являются регулярными выражениями, то **A|B** будет соответствовать любая строка, которая соответствует A или B.

Метасимвол | имеет очень низкий приоритет для того, чтобы заставить его работать разумно, когда вы чередуете несколько символов строки.

Crow|Servo будет искать соответствие либо Crow, либо Servo, а не Cro('w' или 'S')ervo.

^

Ищет соответствие **только в начале строки.**

Например, если вы хотите найти только те строки, у которых в начале имеется From, то в регулярном выражении записывается **^From.**

Может использоваться и как отрицание:

Например, выражение **[^5]** соответствует любому символу, кроме '5'.

\$ - соответствует концу строки. Следует использовать при сортировке по концу строки

\b - граница слова. Слово определяется как последовательность символов чисел и/или букв, так что границы слова представляют пробелы или любые символы, не относящиеся к перечисленным.

Например, если ищем URL со словом `class`, регулярное выражение будет выглядеть так `\bclass\b`

Группировка

Группы обозначаются метасимволами в виде круглых скобок $()$. Они имеют такой же смысл, как в математических выражениях; они группируют вместе выражения, содержащиеся в них, и вы можете повторять содержание группы повторяющимися квалификаторами, такими как $*$, $+$, $?$ и $\{m, n\}$.

Например, $(ab)^*$ будет соответствовать нулю или более повторений ab .

Разберем пример

Допустим наш сайт имеет 17 страниц:

<http://www.site.com/>

<http://www.site.com/index.php>

<http://www.site.com/products/100.php>

<http://www.site.com/products/101.php>

<http://www.site.com/products/102.php>

<http://www.site.com/duals/index.html>

<http://www.site.com/ourteam/index.php>

<https://www.site.com/>

<https://www.site.com/profile>

<http://www.site.com/ua/index.php>

<http://www.site.com/ua/productu/100.php>

<http://www.site.com/ua/productu/101.php>

<http://www.site.com/ua/productu/102.php>

<http://www.site.com/search?q=widget>

<http://www.site.com/search?q=widget+thinger>

<http://www.site.com/search?q=smidges>

<http://www.site.com/index/yy.jpg>

Итак задачи:

1. Чтобы в статистике учитывать все эти страницы, в окно «регулярное выражение» пишем * (17 соответствий)
2. Все страницы с https (2 соответствия)
^https.* — все URL, начинающиеся с https
3. Все страницы на украинском языке (4 соответствия)
.*ua.* — все страницы, в URL которых содержится /ua/.

Если использовать просто ua, то в выборку попадет

<http://www.site.com/duals/index.html>, которое нам не нужно.

4. Все индексные страницы (4 соответствия)
.*index.(php|html)* — аналогично
прошлomu выражению, нельзя просто
использовать `index`.
5. Все карточки товаров (для русской и
украинской версий)
.*product(s|u).* или **.*product[su].*** — оба
варианта подходят.

Метрика β-версия:

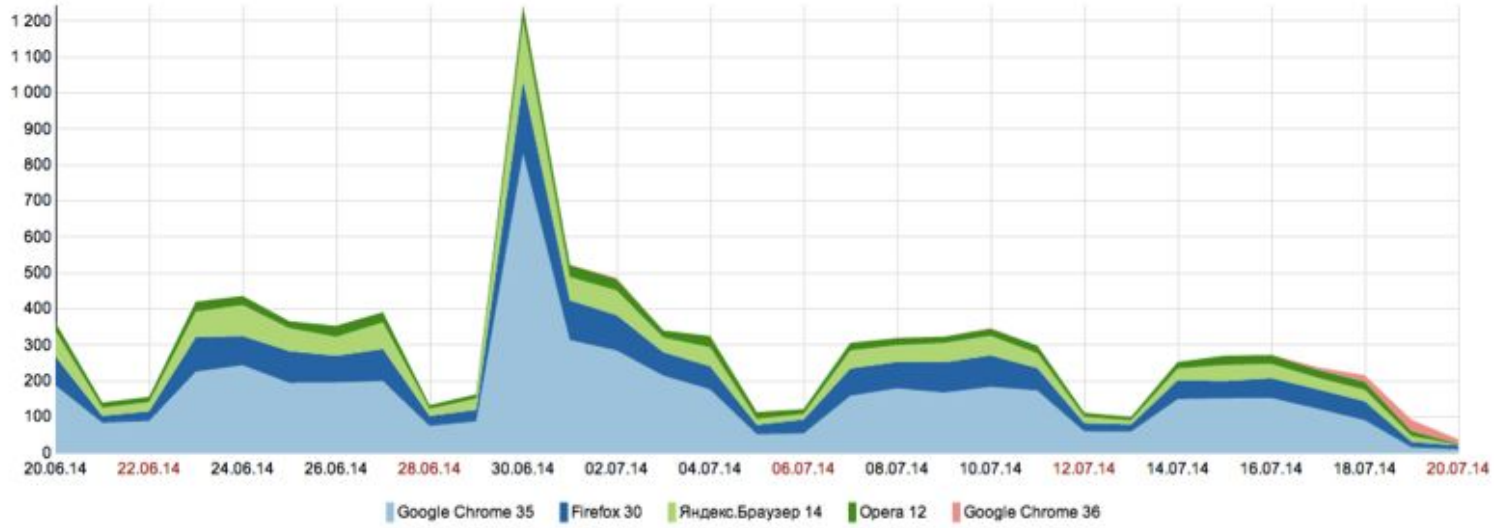
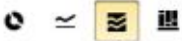
ОТЛИЧИЯ И ЛОГИКА РАБОТЫ

Ключевое отличие Метрики 2.0 — новый принцип работы с данными о посетителях сайта.

1. Теперь отчеты подстраиваются под ваши задачи

Раньше основные отчеты Метрики содержали фиксированный набор колонок. Теперь — это большой конструктор, в котором вы можете быстро дополнить любой отчет нужными данными, не уходя со страницы.

- Поведение >
- Технологии >
- Браузеры
- Разрешение дисплея
- Операционные системы
- Мобильность
- Мобильные устройства
- Версии Flash
- Версии Silverlight
- Наличие Java
- Наличие Cookies
- Наличие JavaScript
- Конструктор >
- Мониторинг >
- Острова >



Группировки ▾ Метрики ▾

- ⋮ Версия браузера × 🔒
- ⋮ Операционная система (детально) × 🔒

Визиты ▾	Просмотры ▾	Отказы ▾	Глубина просмотра ▾	Средняя длительность визита ▾
× 🔄 📊	× 🔄 📊	× 📊	× 📊	× 📊





2. Трафик можно поделить на сегменты

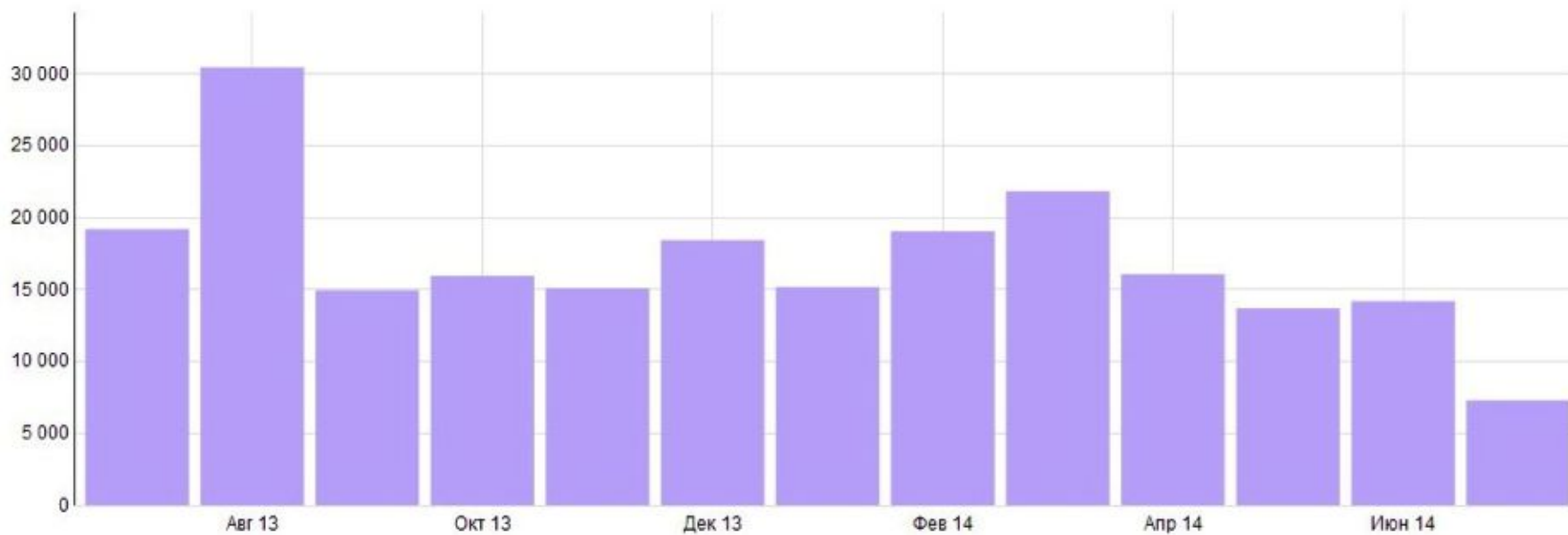
В Метрике 2.0 вы можете посмотреть все отчеты по определенному срезу аудитории. Скажем, выбрать для анализа только постоянных посетителей, которые уже были на сайте несколько раз и возвращаются с определенной периодичностью.

Новый сегмент









Сохранить

Удалить всё

-  **Посетители** ≈317К визитов
История > Общее кол-во визитов > Больше 5 
-  **Посетители** ≈793К визитов
История > Регулярность (кол-во дней между визитами) > Меньше 10, дни 



3. Сегменты легко сравнивать между собой У вас есть возможность вывести на экран наглядное сравнение двух любых сегментов трафика.

 Источники 42K визитов Первый источник > Тип источника > Переходы из поисковых систем		 Источники 58K визитов Первый источник > Тип источника > Переходы по рекламе	
 Посетители 47K визитов История > Новизна (новый/не новый) > Вернувшиеся		 Посетители 47K визитов История > Новизна (новый/не новый) > Вернувшиеся	



4. Вы никогда не потеряете источник первого визита

Если пользователь приходил на сайт несколько раз, вы легко разберетесь, что привело его к вам впервые, а какие источники он использует для возвращений.

Сегментировать ▾

Сравнить сегменты ▾

Цель: Вход в корзину ▾

Всего 7К визитов



Прямые заходы | Переходы из поисковых систем | Переходы по рекламе | Переходы с почтовых рассылок | Внутренние переходы | Остальные



Группировки ▾

Метрики ▾

Атрибуция: Первый переход ▾

Так же стало очень удобно искать необходимые отчеты.

Просто в разделе «Отчеты» в окне поиск набираем необходимый нам отчет и из выпадающего списка выбираем нужный!

The screenshot displays the Yandex.Metrica dashboard. At the top, the navigation bar includes the Yandex logo, a 'Метрика^β' badge, and menu items: 'Счётчики', 'Целевой звонок', 'Представители', 'API', and 'Клуб'. On the right side, there is a user profile icon and the text 'a1seo.ru'. A dark sidebar on the left contains navigation options: 'Сводка', 'Отчеты', 'Карты', 'Вебвизор', and 'Настройка'. The 'Отчеты' section is active, showing a search bar with a red arrow pointing to it. Below the search bar is a list of 'Стандартные отчёты' (Standard reports) including: 'Посещаемость', 'Конверсии', 'Источники', 'Посетители', 'Содержание', 'Технологии', 'Мониторинг', and 'Целевой звонок'. To the right of this list is a detailed view of the 'Посещаемость' (Visits) report. It features a line chart with a green line showing data from 27.04.15 to 17.05.15. The chart has a 'Детализация: по дням' (Detail: by day) dropdown menu. Below the chart, there are sections for 'Исходящая поисковая фраза' (Outgoing search phrase) and 'Отказы' (Bounces).

Спасибо за внимание!