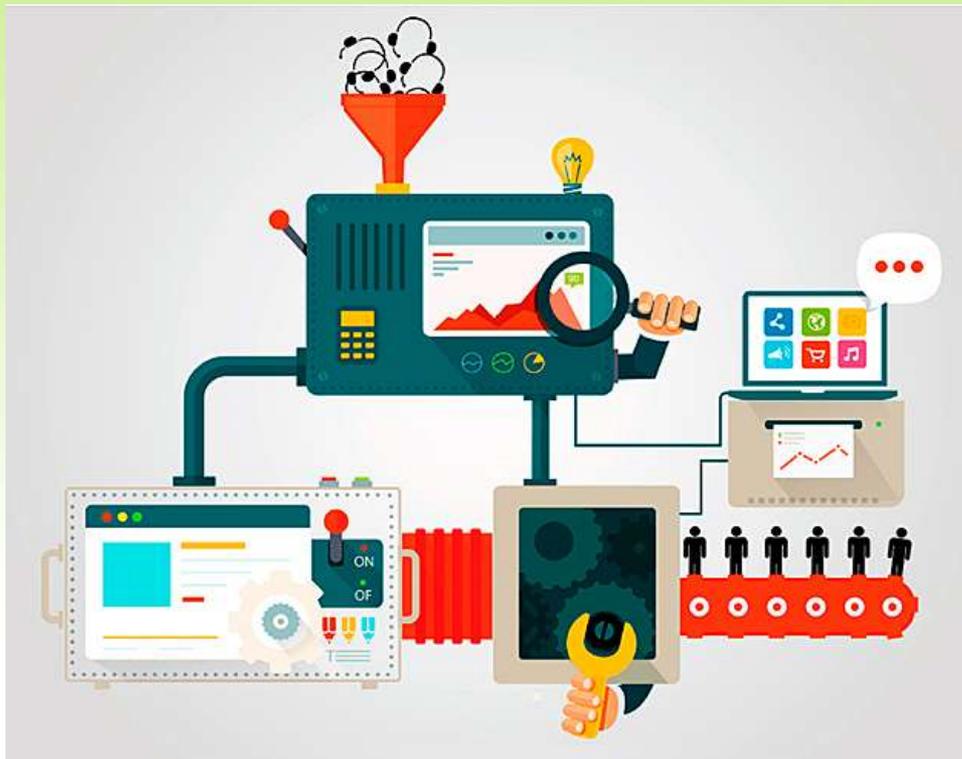


Канн Сергей Константинович  
снс ГПНТБ СО РАН  
[skann@yandex.ru](mailto:skann@yandex.ru)

# Достоверность библиотечной веб-статистики



# Значение сбора и анализа статистики сайта



- ✓ **Управление ресурсами**
  - ⚙️ мониторинг посещаемости
  - ⚙️ определение аудитории
  - ⚙️ изучение поведения и запросов пользователей
  - ⚙️ принятие решений по развитию функционала
  - ⚙️ оценка эффективности вложений
  - ⚙️ выявление проблем сайта (в структуре, навигации и контенте)
- ✓ **Выявление тенденций развития (анализ и прогноз)**
- ✓ **Отчётность**
- ✓ **Библиометрия и библиотекведение**
- ✓ **Сравнения (рейтинги)**



# Инструменты веб-статистики



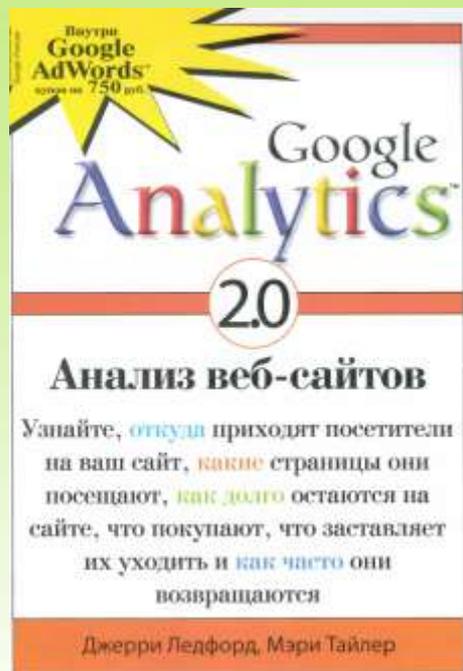
## ✓ Внешние -

- ✦ счётчики-рейтинги (hotlog, liveinternet, openstat, rambler's top 100)
- ✦ счётчики-трекеры (яндекс.метрика, google analytics)
- ✦ системы интернет-аналитики для детализации разных веб-показателей по просмотрам страниц, поведению пользователей, etc. (clicktale, iplogger, spybox, woopra)

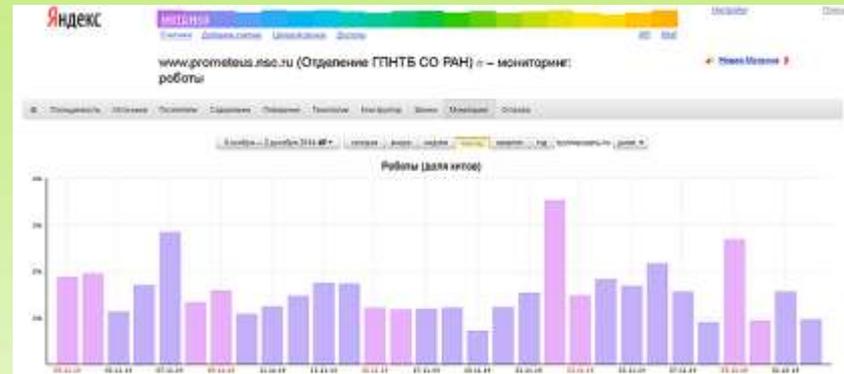
## ✓ Внутренние -

- ✦ лог-анализаторы (analog, awstats, piwik, semonitor, webalizer, webtrends)
- ✦ специализированные модули статистики (lc-bitrix-web-analytics, drupal statistics, joomlastats, wordpress wp-statistics)

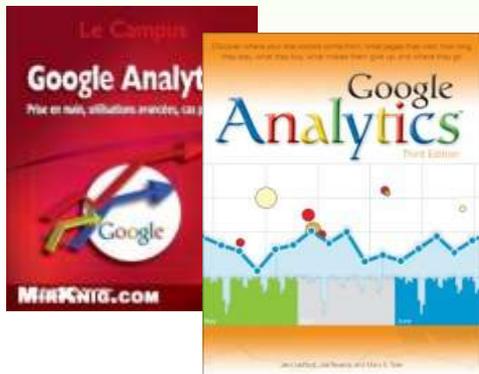




# Недочёты внешней статистики



- ✓ Предъявляя единые требования к мониторингу разных сайтов, внешние системы слежения **слишком зависимы от каналов связи** (пропускной способности, устойчивости, особенностей конфигурации сети и пр.); обрабатываемые узлы ставятся в заведомо неравное положение
- ✓ В профессиональной печати всё чаще ставятся вопросы о точности и корректности данных **Google Analytics**
- ✓ В **Яндекс.Метрике** сами авторы проекта пишут: «Большинство роботов не исполняют код Метрики и могут быть обнаружены только в логах веб-сервера»



# Точность данных в Google Analytics

**CMS** magazine

аналитический портал рынка веб-разработок

06.11.2013 | Автор: **Алексей Макаров** (ведущий специалист по аналитике Russian Promo)



## На стороне сервера

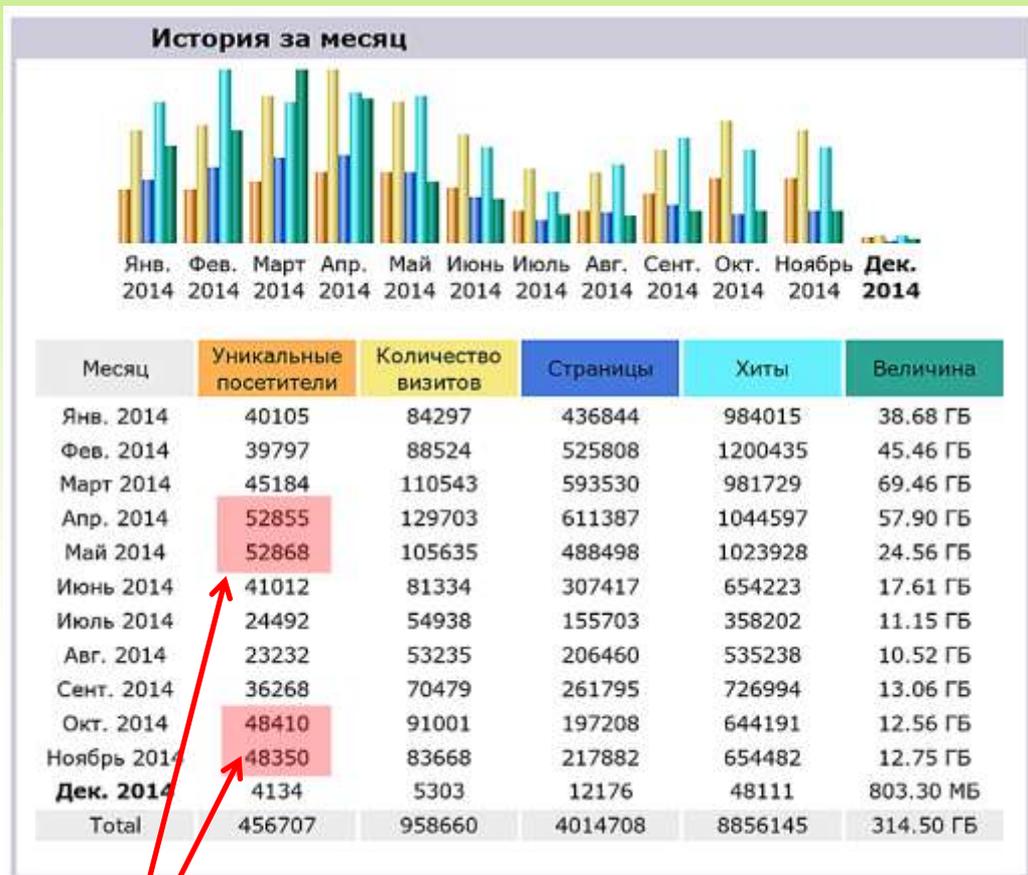
- ✓ Не на всех страницах установлен код отслеживания (Google Analytics Tracking Code)
- ✓ Код отслеживания неверно расположен (не успевает загрузиться)
- ✓ Путаница кодов отслеживания на одном ресурсе
- ✓ Ошибки выполнения JavaScript
- ✓ Ошибки в синтаксисе кода отслеживания
- ✓ Подсчет собственных посещений
- ✓ Некорректная настройка utm-меток

## На стороне клиента

- ✓ В браузере не включен JavaScript
- ✓ Отключены или удалены cookie-файлы
- ✓ Истёк срок действия cookie-файлов
- ✓ Одно устройство – несколько пользователей; один пользователь – разные устройства
- ✓ Данные обрабатываются не сразу

Google

# Инструменты внутренней веб-статистики

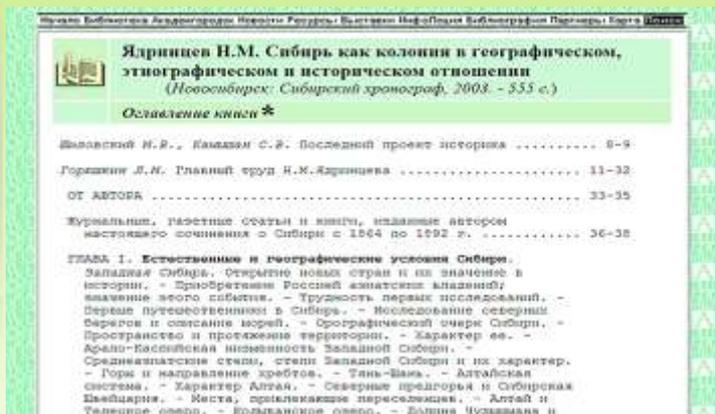


- ✓ На сайте Отделения ГПНТБ СО РАН используется статистическая система AWStats 6.5 (устаревшая)
- ✓ Система не знает браузера Chrome, ОС Windows 7, многих мобильных приложений
- ✓ Общие недостатки систем лог-анализа – плохое отслеживание роботов (ботов)
- ✓ Статистика искажается при использовании разных средств «анонимизации» посещений

[режимы приватности браузеров («инкогнито»), очистка журнала, удаление cookie, подавление скриптов с помощью плагинов (disallow script button, ghostery, noscript), блокирование сценариев, изображений и встроенных объектов (пиксельных тегов, веб-маяков)]

устойчивость показателей

# Проблема различения «ботов» и людей



- ✓ Автоматические средства работы с сайтом являются главным фактором искажения веб-статистики
- ✓ При отслеживании ботов необходимо использовать несколько инструментов

Статистика за:  
www.prometheus.nsc.ru

- Общие
- Когда:
- История за месяц
- День месяца
- День недели
- Часы
- Кто:
- Страны
- Полный список
- Хосты
- Полный список
- Последний визит
- Неразрешенный IP адрес
- Роботы/Пауки посетители
- Полный список
- Последний визит
- Навигация:
- Продолжительность визитов
- Тип файла
- Просмотров
- Полный список
- Вхождение
- Выход
- Операционные системы
- Версии
- Неизвестный

Последнее обновление: 23 Сент. 2014 - 11:50  
Отчетный период: Июнь 2014

Вернуться на главную страницу

Фильтр: Фильтр исключений:

Адрес страницы				
Total: 22726 Различные url	Просмотров	Средний размер	Вхождение	Выход
/	73221	63.23 КБ	35198	35149
/elibrary/siberia/yadr1882.pdf	61670	87.36 КБ	10	36
/baseforcalen.txt	4394	18.18 КБ	21	636
/woda/woda/journals.cgi	2772	3.46 КБ		
/www.scian.com/Go	2772	3.48 КБ		
/woda/woda/journals.cgi	2772	3.48 КБ		
/www.elsevier.nl/Go	2772	3.48 КБ		
/woda/woda/journals.cgi	2772	3.48 КБ		
/www.elsevier.com/Go	2772	3.48 КБ		

```
89.223.104.101 -- [31/May/2014:23:25:11+0700] "GET /elibrary/siberia/yadr1882.pdf HTTP/1.1" 200 65536 "-" "AndroidDownloadManager/4.1.1 (Linux; U; Android 4.1.1; Build/JRO03C)"
89.223.104.101 -- [31/May/2014:23:25:12+0700] "GET /elibrary/siberia/yadr1882.pdf HTTP/1.1" 200 65536 "-" "AndroidDownloadManager/4.1.1 (Linux; U; Android 4.1.1; Build/JRO03C)"
81.176.235.101 -- [31/May/2014:23:25:13+0700] "GET /acquisitions/12-03-27/cont05f.ssi HTTP/1.0" 200 21480 "-" "Flexum/2.0"
89.223.104.101 -- [31/May/2014:23:25:14+0700] "GET /elibrary/siberia/yadr1882.pdf HTTP/1.1" 200 65536 "-" "AndroidDownloadManager/4.1.1 (Linux; U; Android 4.1.1; Build/JRO03C)"
```



# Основные блоки веб-статистики сайта



- ✓ Достоверность веб-статистики сайта должна определяться по каждому из блоков
- ✓ для ключевых показателей необходимо определять доверительный интервал
- ✓ в конечном итоге веб-статистика сайта сводится к выявлению фактов и трендов



# Заключение

- ✓ Значимость библиотечной веб-статистики в первую очередь определяется потребностями рационального управления ресурсами (анализ и прогноз развития сайтов)
- ✓ Для совершенствования статистических отчётов по библиотечным сайтам необходимо совершенствовать инструменты и методики сбора и анализа веб-статистики
- ✓ Наиболее важной проблемой в области получения достоверной статистики представляется проблема идентификации и отделения веб-статистики ботов от статистики реальных пользователей
- ✓ В любом случае, представляется, что на данный момент веб-анализ библиотечных сайтов является больше искусством, чем сложившейся наукой



Канн Сергей Константинович

старший научный сотрудник ГПНТБ СО РАН

skann@yandex.ru | serge@prometeus.nsc.ru

**СПАСИБО ЗА ВНИМАНИЕ !**

