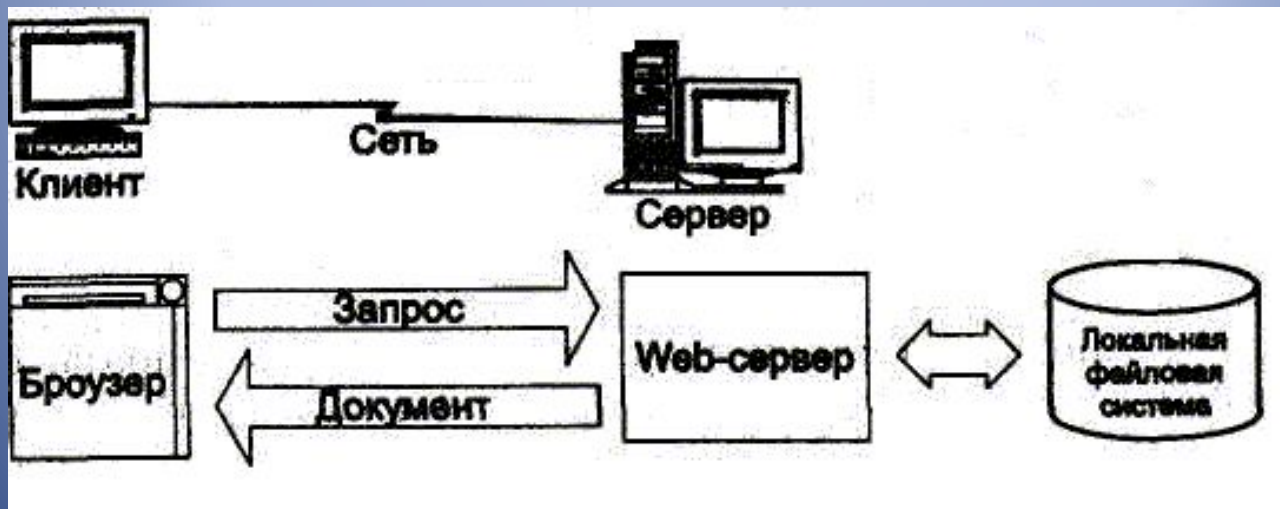


**Задачи
администрирования веб-
систем, их автоматизация.**

Веб-система

Система на основе технологии Web (web-система) — это система гипертекстовых документов, поскольку ее ресурсы связаны между собой. Базовыми элементами web-системы являются браузер клиента, сеть и web-сервер.



Задачи администрирования веб-систем

- предоставление пользователям разрешений на доступ;
- предоставление пользователям разрешений на разработку;
- определение группы пользователей, имеющих доступ к важному контенту (то есть настройка разрешений уровня веб-системы);
- определение возможностей, которые будут доступны посетителям веб-системы;
- предоставление технической поддержки пользователям веб-системы;
- назначение резервного администратора для веб-системы;
- оказание помощи при администрировании определенных возможностей, например:
 - включение и отключение доступных возможностей;
 - создание или настройка типов контента;
 - настройка региональных параметров.

Что такое лог-файлы. Анализ лог-файлов

Лог-файл (или просто лог) - это главный источник информации о поведении посетителя на вашем Web-сайте. Когда кто-то заходит на сайт, лог-файл записывает, например, какие ключевые слова он использовал, чтобы найти ваш ресурс, какие страницы посетил, насколько там задержался и т. д.

Каждое обращение к сайту фиксируется сервером и о нем остается запись в лог-файле. По этим записям можно узнать огромное количество информации о посетителях. Лог-файл сервера представляет собой текстовый файл, в котором в особом формате записаны все данные об обращениях к сайту

При обращении пользователя к сайту происходит сразу несколько событий.

Событие 1. Запрос страницы.

Браузер посетителя передает на сервер (на котором находится сайт) запрос на выдачу страницы. При этом на сервере становится известной следующая информация:

- IP-адрес посетителя;
- Запрашиваемый URL;
- Дата и время запроса;
- Географическое положение клиента;
- Используемый пользователем браузер;
- Адрес страницы с которой зашел посетитель и т.д.

Событие 2. Выдача страницы.

- Сервер возвращает клиенту запрошенный документ (страницу, картинку, файл и пр.).

Событие 3. Запись в лог.

Сервер записывает данные о произошедшей транзакции в журнал событий (лог-файл).

Пример записей в лог-файле.

```
213.180.195.231 - - [29/Jan/2007:00:07:17 +0000] "GET /excel/lsn015.html HTTP/1.1" 200 18918 "-"  
"YaDirectBot/1.0" 74.6.71.37 - - [29/Jan/2007:00:13:03 +0000] "GET /eralash HTTP/1.0" 301 317 "-"  
"Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)«
```

```
213.186.116.166 - - [29/Jan/2007:00:26:51 +0000] "GET /eralash/soft/speed.zip HTTP/1.0" 200 217872 "-"  
"Wget/1.10.2 (Red Hat modified)"
```

```
213.180.207.25 - - [29/Jan/2007:00:27:14 +0000] "GET /robots.txt HTTP/1.1" 200 54 "-" "Yandex/1.01.001  
(compatible; Win16; I)«
```

Лог-файлы представляют собой сырые данные, которые необходимо обработать. Качество обработки и определяет качество статистики

В лог-файлах сайта содержится вся необходимая информация, которой вполне достаточно, чтобы знать все об аудитории, которая посещает сайт.

Следует иметь в виду, что абсолютно "реальную" статистику получить практически невозможно, в силу ряда технических причин. Не существует каких-то корректных способов оценить величину отклонения "реальности" от измеренных характеристик, однако принято считать, что эти отклонения в среднем не превышают 5-10%. Если сайт живет "размеренной жизнью" (не проводятся активные рекламные кампании и проч.), то можно считать ошибку измерения постоянной и успешно сравнивать между собой данные по соседним дням, неделям и месяцам.

Существуют специализированные программы для работы с лог-файлами - это, так называемые, анализаторы лог-файлов. Наиболее известные - Analog, Webalizer, Webtrends, NetPromoter, WebAnalizer.

**С помощью ЛОГНАЛИЗАТОРА можно
определить следующие
статистические данные:**

Общая статистика:

- Хосты;
- Посетители;
- Просмотры;
- Поисковики.

Обращения к страницам:

- URL страницы;
- Количество посещений;
- Заходы с поисковиков;
- Трафик (Мб);

Рефереры:

- Домены;
- URL

Поисковые запросы:

- Текст запроса;
- Количество запросов

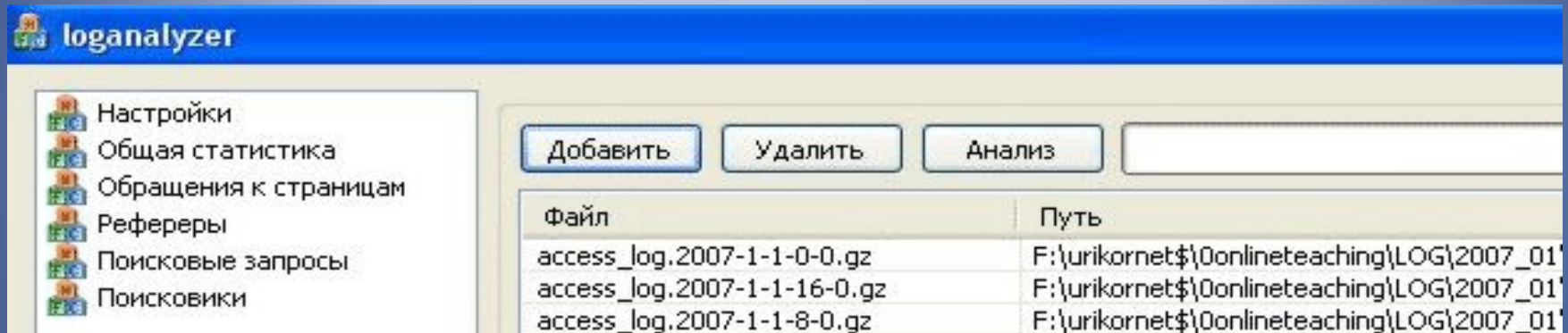
Поисковики:

- Наименование паука;
- Количество запросов;
- URL страницы;
- Код;
- IP паука;
- Дата запроса;

Анализатор лог-файлов сервера Apache

Настройки

Для начала работы необходимо в стартовом окне программы выбрать файлы (кнопка "Добавить") для анализа.



- После чего нажать кнопку "Анализ".
- Процент проанализированных файлов отображается в строке состояния.
- По окончании процесса анализа будет выведено сообщение "Анализ закончен".

Общая статистика

Дата	Хосты	Посетители	Просмотры	Searches
	3046	3166	13251	1095
01/Jan/2007	227	231	665	50
02/Jan/2007	329	348	1576	95
03/Jan/2007	311	317	1181	127
04/Jan/2007	281	290	1325	118
05/Jan/2007	300	314	1352	118
06/Jan/2007	279	290	1104	111
07/Jan/2007	291	304	1189	94
08/Jan/2007	278	291	1418	84
09/Jan/2007	392	409	1897	154
10/Jan/2007	307	319	1403	141
31/Dec/2006	51	53	141	3

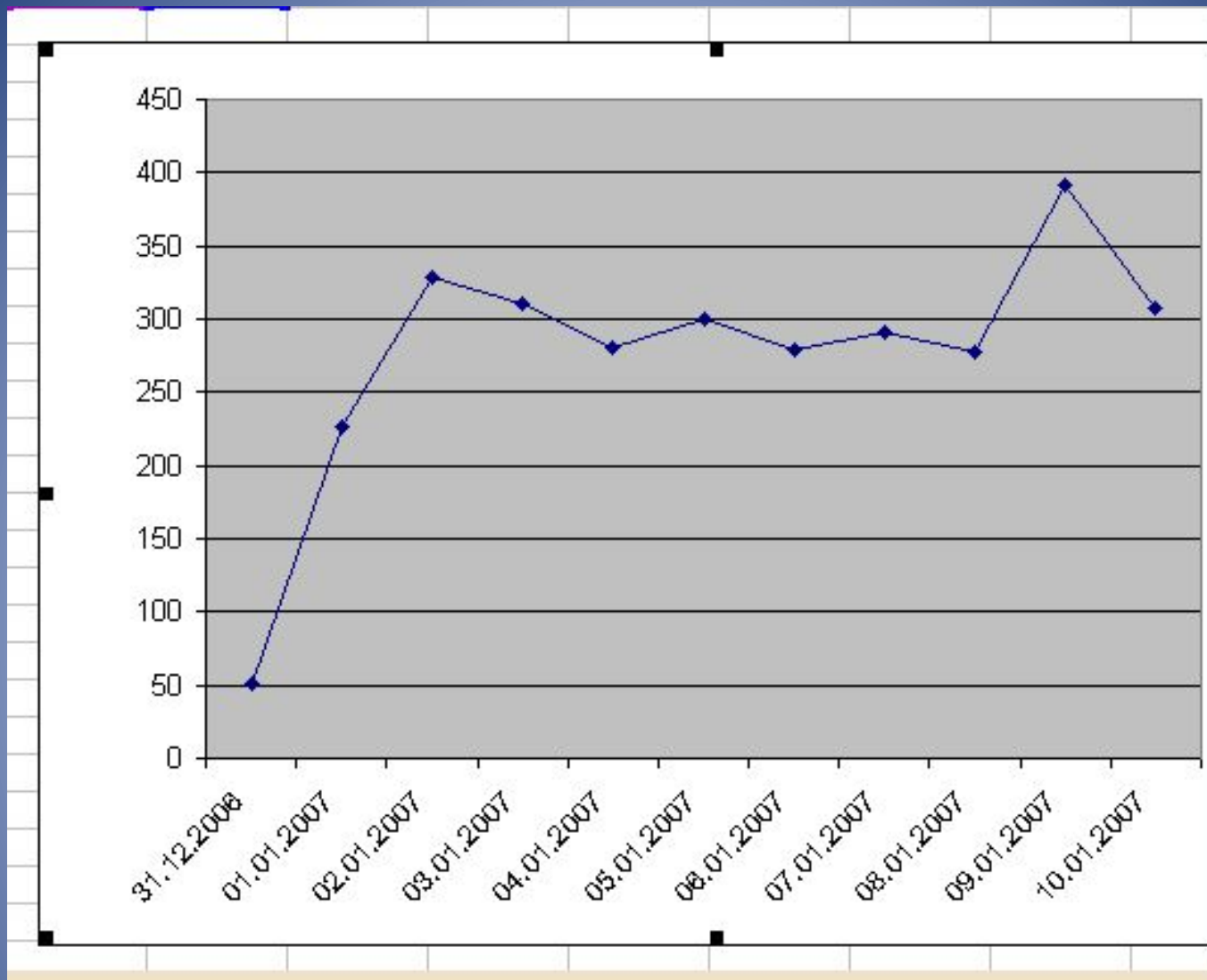
Хосты (уникальные IP-адреса) - показывает, с какого количества уникальных IP-адресов просматривались страницы сайта.

Посетители - этот параметр показывает количество посетителей, просмотревших сайт за определенное время.

Просмотры (хиты) - количество страниц, просмотренных всеми пользователями. Этот параметр показывает, сколько информационных страниц просмотрели пользователи.

Searches - количество заходов с поисковых систем.

Верхняя строка показывает суммарные значения столбцов.



Для импорта полученных результатов в Excel, необходимо нажать кнопку "Excel".

В Excel можно обрабатывать полученную статистику. Например, строить графики.

Обращения к страницам

Excel

Страница	Посещений	Searches	Трафик(Мб)	Google	Yandex	Rambler	Mail.ru	Апорт	Webalta
	13251	1095	255,74	262	672	47	99	-	6
/	988	409	19,35	89	268	10	38	-	-
/robots.txt	745	-	0,21	-	-	-	-	-	-
/excel/index.html	647	125	9,11	13	93	2	16	-	-
/site/index.html	436	187	7,74	2	156	-	29	-	-
/html/index.html	371	18	7,19	9	6	3	-	-	-
/word/index.html	369	59	6,06	15	36	2	5	-	-
/eralash/soft/speed.zip	293	-	30,21	-	-	-	-	-	-
/excel/lsn003.html	219	8	4,89	5	1	1	-	-	-

- **Посещения** - количество обращений (посещений) к конкретной странице.
- **Searches** - суммарное количество заходов на страницу с поисковых систем.
- **Трафик (Мб)** - трафик, который сгенерировала данная страница.
- **Google** - количество заходов на страницу с данного поисковика.
- **Yandex** - количество заходов на страницу с данного поисковика.
- **Rambler** - количество заходов на страницу с данного поисковика.
- **Mail.ru** - количество заходов на страницу с данного поисковика.
- **Апорт** - количество заходов на страницу с данного поисковика.
- **Вебальта** - количество заходов на страницу с данного поисковика.

Верхняя строка показывает суммарные значения столбцов.

Рефереры

Исключить рефереры с указанного домена

on-line-teaching.com

Применить

Домены	Кол-во	URL	Кол-во
www.yandex.ru	672	http://stamina.ru/	83
stamina.ru	128	http://stamina.ru/download.htm	45
www.google.ru	121	http://e-molotok.com/ezine145.htm	34
go.mail.ru	99	http://www.yandex.ru/yandsearch?text=%EA...	31
www.google.com	66	http://www.stamina.ru/	30
www.stamina.ru	52	http://www.yandex.ru/yandsearch?stype=ww...	29
www.google.com.ua	51	http://www.yandex.ru/yandsearch?stype=ww...	26
www.rambler.ru	48	http://www.yandex.ru/yandsearch?text=Excel...	26

- **Реферер (ссылающийся ресурс)** - показывает откуда посетитель пришел на сайт.
- **Домены** - сайты, с которых был выполнен переход.
- **URL** - точный адрес страницы, с которой был выполнен переход.
- **Исключать рефереры с указанного домена** - обычно указывается домен анализируемого сайта, чтобы исключить переходы по внутренним ссылкам сайта.

Поисковые запросы

Excel

Запрос	Всего	Google	Yandex	Rambler	Mail.ru	Апорт	Webalta
	1086	262	672	47	99	-	6
excel	203	5	169	-	27	-	2
как сделать свой сайт	129	-	110	-	19	-	-
word	93	-	81	-	11	-	1
как сделать свой сайт?	30	-	23	-	7	-	-
программа excel	22	-	19	-	3	-	-
Html	19	1	16	-	2	-	-
обучение HTML	15	1	12	2	-	-	-
Обучение Excel	11	4	6	-	1	-	-
текстовые редакторы	11	-	9	-	2	-	-
основы html	10	2	8	-	-	-	-
сделать свой сайт	10	-	9	-	1	-	-
Работа в Word	9	-	9	-	-	-	-

- Статистика поисковых запросов показывает по каким фразам и с каких поисковиков заходили посетители на сайт.
- Верхняя строка показывает суммарные значения столбцов.

Поисковики

Паук	Запросы	
Google	107	
Rambler	4	
Webalta	180	
Yandex	353	

Паук	Запрос	Код	IP	Дата
Google	/robots.txt	302	66.249.65.16	01/Jan/2007
Google	/	200	66.249.65.16	01/Jan/2007
Yandex	/robots.txt	302	213.180.207.25	01/Jan/2007
Yandex	/	200	213.180.207.25	01/Jan/2007
Yandex	/graphika/index.html	200	213.180.207.25	01/Jan/2007
Yandex	/map/index.html	200	213.180.207.25	01/Jan/2007
Yandex	/site/proposition.html	200	213.180.207.25	01/Jan/2007
Yandex	/site/index.html	200	213.180.207.25	01/Jan/2007

В заключительном отчете представлена информация о **визитах поисковых роботов**. Прежде чем попасть в выдачу поисковой системы, сайт должен быть проиндексирован и добавлен в базу поисковой системы. Этой работой занимаются поисковые роботы-пауки - ищут новые страницы, отсутствующие в индексе, и обновляют информацию об уже известных. В данном отчете представлена информация о посещениях поисковыми роботами сайта.

Журнал безопасности

Журнал безопасности брандмауэра подключения к Интернету (Internet Connection Firewall, ICF) позволяет опытным пользователям выбирать, какие сведения следует заносить в журнал. Журнал безопасности ICF поддерживает следующие возможности.

- **Записывать пропущенные пакеты.** Этот параметр задает запись в журнал сведений о всех потерянных пакетах, исходящих из домашней или небольшой офисной сети или из Интернета.
- **Записывать успешные подключения.** Этот параметр задает запись в журнал сведений о всех успешных соединениях, исходящих из домашней или небольшой офисной сети или из Интернета.

Журнал безопасности ICF

состоит из двух разделов

- В *заголовке* содержатся сведения о версии журнала и полях, в которые можно записывать данные. Содержимое заголовка имеет вид статического списка.
- *Тело* представляет собой откомпилированные данные, которые вводятся при обнаружении трафика, пытающегося пройти через брандмауэр. Поля журнала в теле журнала безопасности ICF заполняются слева направо, как они расположены на странице. Тело журнала безопасности представляет собой динамический список; новые данные записываются в конец журнала.

Элементы заголовка

Элемент	Описание	Пример
#Version	Номер установленной версии журнала безопасности ICF.	1.0
#Software	Имя журнала безопасности ICF.	Брандмауэр подключения к Интернету Microsoft (ICF)
#Time	Задаёт использование местного времени при записи в журнал отметок времени	Местное
#Fields	Отображает статический список полей, доступных для записей журнала безопасности при наличии данных. Поля — date, time, action, protocol, src-ip, dst-ip, src-port, dst-port, size, tcpflags, tcpsyn, tcsack, tcpwin, icmptype, icmpcode и info.	Src-port, dst-port, size

Тело журнала

Имя столбца	Описание	Пример
Date	Указывает год, месяц и день регистрации транзакций. Дата представляется в следующем формате: ГГ-ММ-ДД где ГГ обозначает год, ММ — месяц, а ДД — число.	2001-01-27
Time	Указывает время регистрации транзакции с точностью до секунды. Время записывается в следующем формате: ЧЧ:ММ:СС где ЧЧ обозначает часы в 24-часовом цикле, ММ — минуты, а СС — секунды.	21:36:59
Action	Указывает операцию, зарегистрированную брандмауэром. Доступные параметры: OPEN (открытие), CLOSE (закрытие), DROP (отклонение) и INFO-EVENTS-LOST (потерянные события). Для действия INFO-EVENTS-LOST указывается число событий, которые имели место, но не были занесены в журнал.	DROP
Protocol	Указывает протокол, использовавшийся для соединения. Возможная запись протокола — TCP, UDP, ICMP или число, если использовавшийся протокол был отличен от TCP, UDP или ICMP.	ICMP
Src-ip	Указывает исходный IP-адрес (адрес компьютера, пытавшегося установить связь). Исходный IP-адрес записывается в точечно-десятичном формате: (число).(число).(число).(число)	192.168.0.0
Dst-ip	Указывает IP-адрес назначения (адрес пункта назначения, с которым исходный компьютер пытался установить связь). IP-адрес назначения записывается в точечно-десятичном формате:	192.168.0.1

Src-port	<p>Указывает номер порта на компьютере-отправителе. Номер порта компьютера-отправителя задается целым числом в диапазоне от 1 до 65 535. Действительное значение порта возвращают только протоколы TCP и UDP. Все остальные протоколы не поддерживают параметр src-port, для них в этом поле записывается - (дефис).</p>	4039
Dst-port	<p>Указывает порт компьютера назначения. Номер порта компьютера назначения задается целым числом в диапазоне от 1 до 65 535. Действительное значение порта возвращают только протоколы TCP и UDP. Все остальные протоколы не поддерживают параметр dst-port, для них в этом поле записывается - (дефис).</p>	53
Size	<p>Указывает размер пакета в байтах.</p>	60
Tcpflags	<p>Указывает флаги управления TCP, содержащиеся в заголовке TCP пакета IP:</p> <ul style="list-style-type: none"> • Ack Acknowledgment field significant (Включение поля подтверждения) • Fin No more data from sender (Конец массива данных отправителя) • Psh Push Function (Функция принудительной доставки) • Rst Reset the connection (Сброс подключения) • Syn Synchronize sequence numbers (Синхронизация порядковых номеров) • Urg Urgent Pointer field significant (Включение поля указателя срочных данных) <p>Флаги записываются прописными буквами. Содержимое полей tcpflags предназначено для пользователей, обладающих углубленными знаниями протокола TCP (Transmission Control Protocol). Дополнительные сведения о</p>	AFP

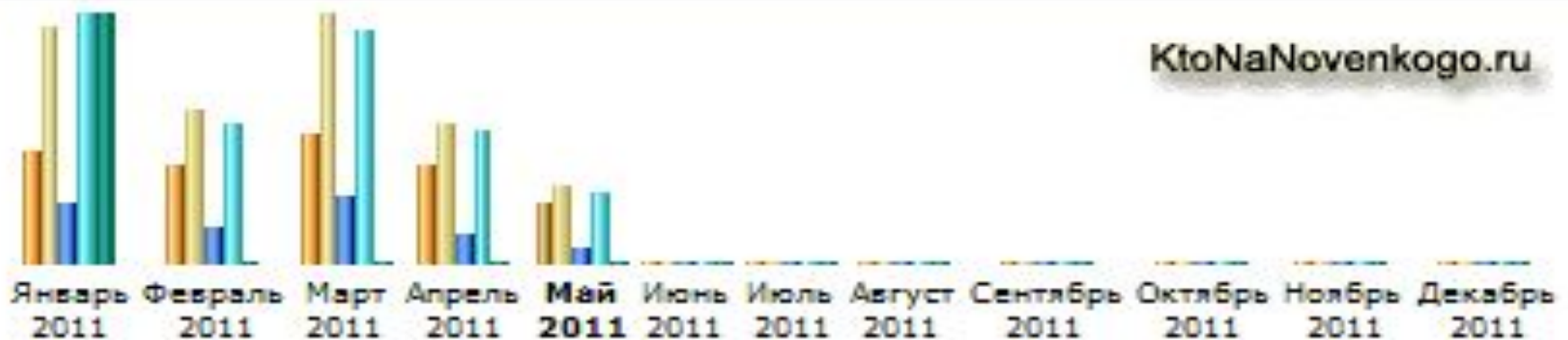
Tcpsyn	Указывает порядковый номер TCP в пакете. Содержимое поля tcpsyn предназначено для пользователей, обладающих углубленными знаниями протокола TCP.	1315819770
Tcrack	Указывает номер подтверждения TCP в пакете. Содержимое поля tcrack предназначено для пользователей, обладающих углубленными знаниями протокола TCP.	0
Tcrwin	Указывает размер окна TCP в пакете. Размер указывается в байтах. Содержимое поля tcrwin предназначено для пользователей, обладающих углубленными знаниями протокола TCP.	64240
Icmptype	Число, представляющее собой значение поля type (тип) в сообщении ICMP.	8
Icmrcode	Число, представляющее значение поля code (код) в сообщении ICMP.	0
Info	Указывает сведения, зависящие от типа действия. Например, действие INFO-EVENTS-LOST отображает число событий, которые произошли с момента последней регистрации события этого типа, но не были занесены в журнал.	23

Посещаемость сайта

Самый точный способ проверки статистики посещений заключается в использовании так называемой серверной статистики посещаемости (обычно анализируются логи сервера). Большинство виртуальных хостингов предоставляют такую услугу, ну, а если ваш проект размещен на реальном или виртуальном сервере, то вы и сами можете установить и активировать соответствующие скрипты, позволяющие узнать, проверить и провести подробный анализ посещаемости вашего сайта.

AWStats — анализатор логов сервера

История за месяц



Месяц	Уникальные посетители	Количество визитов	Страницы	Хиты	Объем
Январь 2011	674	1,413	2,345	9,675	22.88 ГБ
Февраль 2011	593	922	1,394	5,446	24.31 МБ
Март 2011	771	1,492	2,583	9,112	44.82 МБ
Апрель 2011	590	831	1,174	5,095	47.23 МБ
Май 2011	359	461	634	2,764	19.32 МБ
Июнь 2011	0	0	0	0	0
Июль 2011	0	0	0	0	0
Август 2011	0	0	0	0	0
Сентябрь 2011	0	0	0	0	0
Октябрь 2011	0	0	0	0	0
Ноябрь 2011	0	0	0	0	0
Декабрь 2011	0	0	0	0	0
Total	2,967	5,119	8,130	32,092	23.01 ГБ

Во многих [CMS](#) ([CMS Joomla](#), [SMF](#) или [Вордпресс](#)) как платных, так и бесплатных, существует возможность отслеживать статистику посещений сайта. Если такой возможности не заложено в самой CMS, то наверняка найдутся подходящие сторонние расширения проверки и анализа статистики посещаемости.

Но, как правило, такие системы сбора статистики в CMS далеко не совершенны (хотя, в Битриксе она очень даже ничего), да и опять же встает вопрос предоставления доступа к статистике для потенциальных рекламодателей, а они ни на грош не поверят тому, что насчитает статистика посещений, имеющаяся в вашей CMS.

Счетчики посещаемости

- [Счетчик liveinternet](#) — самый популярный в рунете способ узнать и проверить посещаемость сайта. Говорить тут особо нечего, тем более, что по ссылке вы найдете подробную статью по получению и установке кода счетчика liveinternet. Вообще, для большинства не коммерческих проектов (где не нужно отслеживать конверсию) [статистики liveinternet](#) будет более, чем достаточно.
- [Яндекс Метрика](#) — позволяет не только узнать и проверить посещаемость, но и позволят вам формировать желаемые цели и отслеживать их достижение. Вообще, это уже, конечно же, не обычные счетчики посещений, а полноценные системы аналитики. Хотя, Яндекс Метрика не так давно совершила шаг в сторону того, чтобы ее можно было назвать [счетчиком Яндекса](#).
- [Google Analytics](#) — самая продвинутая из всех доступных бесплатно систем аналитики посещаемости. Называть Google Analytics счетчиком уже язык как-то не поворачивается. Вообще, довольно давно Google Analytics назывался немного иначе и был платной системой аналитики посещаемости сайта (несколько сотен долларов в месяц за то, чтобы узнать посещаемость сайта), но потом его выкупил Google и сделал Google Analytics доступным всем желающим и абсолютно бесплатным. К тому же, система статистики посещаемости Google Analytics с тех пор усиленно развивается и совершенствуется, достигнув к текущему моменту небывалой мощи. Одно расстраивает — в Google Analytics очень трудно разобраться, особенно сходу.

Спасибо за внимание!