

# ИНСТРУМЕНТЫ ОЦЕНКИ ИНФОРМАЦИОННОГО ОБРАЗА САЙТА

Канн Сергей Константинович

(Отделение Государственной публичной научно-технической библиотеки Сибирского отделения Российской Академии наук)

e-mail: [serge@prometeus.nsc.ru](mailto:serge@prometeus.nsc.ru)

Веб-сайты представляют виртуальные сущности (объекты), которые, несмотря на свою «виртуальность», существуют реально, обладают конкретными именами, адресами и сущностными характеристиками, позволяющими различать их и сравнивать друг с другом. Эти сравнения производятся путем сопоставления конкретных «информационных образов», в понятие которых входит совокупность отдельных свойств сайтов (функциональных, содержательных, навигационных, визуальных), формирующих представление о них у широкой аудитории пользователей. Как и «имиджи» отдельных людей или учреждений, устоявшиеся образы сайтов сочетают в себе и реальные свойства виртуальных объектов, и несуществующие, приписываемые характеристики, которые со временем образуют *репутацию*. Но, в отличие от «имиджей», образы сайтов напрочь лишены материальной основы, намного более подвижны и переменчивы и представляют значительные сложности для изучения.

С этой точки зрения, исследователи Сети испытывают те же трудности, что и естествоиспытатели XVIII века, вынужденные изучать природные объекты и явления (электричество, магнетизм, «теплород») при полном отсутствии необходимых инструментов. Постоянная изменчивость веб-узлов, лишь отдаленно напоминающих «виртуальные книги», представляет существенный фактор, отличающий «виртуальную книгу» (сайт) от традиционной. Тем не менее, нехватка инструментария не означает, что нужно отказаться от изучения виртуальных объектов в Интернете до наступления лучших времен. Любая наука будущего складывается из «кирпичей» настоящего.

Развитие информационных технологий приводит к появлению более эффективных инструментов изучения веб-образов, углубляет понимание особенностей функционирования веб-среды, вырабатывает рациональные критерии

оценки сетевой деятельности. Разрабатываются новые онлайн-инструменты веб-цитирования, статистической оценки, мониторинга и поисковой оптимизации сайтов (SEO), позволяющие проводить сравнения на основе однородной системы *внешних* показателей, независимо от *внутренней* статистики серверов, которая, как показывает опыт, не всегда достоверна. Хотя и здесь очевидно, что имеющиеся инструменты веб-анализа далеки от идеала, и чем больше будет привлечено самых разных средств, – тем лучше. Для оценки сложившихся веб-образов полезно использовать технологии ранжирования сайтов поисковыми машинами, веб-каталоги, рейтинги, блоги, форумы и пр.

*Организационное* единство сайтов, обусловленное административно-правовыми условиями и принадлежностью к определенному коллективу исполнителей, сочетается с *технологическим* единством аппаратных средств и программного обеспечения. «Материальным базисом» сайта в данном случае выступает сервер, хотя и не всегда один. Ресурсы сайта могут размещаться на нескольких серверах или «зеркалиться» на ряде узлов, имеющих разные адреса [1]. При получении статистических оценок этот фактор следует учитывать, оговаривая при публикации и сравнении показателей. В противном случае, оценки могут показаться некорректными из-за прямой зависимости между посещаемостью сайта и количеством представленных на нем документов.

Популярность Интернет-ресурсов в значительной степени зависит от технических возможностей серверов, предоставляющих хостинг. Опыт показывает, что большое количество пользователей ограничивается очень коротким временем пребывания даже на самых интересных сайтах. В 2010 г. число посетителей, находившихся на [www.prometeus.nsc.ru](http://www.prometeus.nsc.ru) не более 30 секунд, составляло 85 %, тогда как в 2005 г. этот показатель равнялся 75 %. Ускорение информационных обменов, «всеобщая мобилизация», при которой совершенствуются умения, навыки и повышается скорость обработки информации, становятся главной тенденцией Интернета. Параллельно развиваются технические средства доступа к информации (браузеры, поисковые машины, веб-навигаторы) и

пользователю уже хватает 30 секунд, чтобы определить ценность найденной информации и сохранить результаты на своем компьютере.

В итоге, «наиболее характерное время ожидания», на которое соглашаются потребители веб-информации, теперь не превышает 4 секунд [2]. Если за это время документ не загружается на экран, 90% посетителей покидают сайт, не дожидаясь развязки, при этом популярность сайта неуклонно падает. Отсюда следует вывод, что технические условия быстрой загрузки страниц и оперативного веб-обслуживания становятся принципиальным фактором роста посещаемости. Хорошие сайты имеют «шустрый» веб-сервер, четкую структуру ресурсов, грамотную навигацию и оптимизированную графику.

В настоящее время в Интернете можно найти большое количество инструментов для проверки доступности веб-узлов, в том числе в режиме реального времени. Среди бесплатных служб измерения скоростей доступа назовем проверенные и внешне привлекательные сервисы SpeedTest ([www.speedtest.net](http://www.speedtest.net)) и



Рис. 1. Тестирование скорости сетевого соединения с помощью сервиса SpeedTest.net

SpeakEasy ([www.speakeasy.net](http://www.speakeasy.net)), предоставляющие данные не только о возможностях каналов связи с любой точкой земного шара (на «вход – выход»), но и практически полезные сведения об IP-адресах, провайдерах, среднем времени загрузки медиа-

файлов (графики, звука и клипов). Для измерения скорости соединения с Яндексом существует сервис [internet.yandex.ru](http://internet.yandex.ru), хотя следует иметь в виду, что партнерские программы компании с провайдерами (так называемый «локальный Яндекс») ускоряют загрузку документов и делают результаты замеров не вполне объективными.

Для изучения технических возможностей библиотечных серверов было проведено массовое сравнение скоростей загрузки сайтов с помощью инструментов онлайн-веб-анализа, таких как [pr-cy.ru](http://pr-cy.ru), [tic-pr.com](http://tic-pr.com) и [scripter.com.ua](http://scripter.com.ua).

Ниже представлены результаты теста, проводившегося утром 23.11.2010 г. (8.00 мск), одномоментно для 10 ведущих библиотечных сайтов России, ранжированных с учетом «тематического индекса цитирования» ([ТИЦ](#)) [Яндекса](#) (табл. 1). В данном случае ТИЦ, как способ определения авторитетности сайтов на основе разработанного Яндексом алгоритма «тематической близости ресурсов», выступал еще и в качестве одного из инструментов оценки информационного образа сайтов. При одновременном проведении теста преследовалась цель повысить достоверность результатов, хотя они так подвижны, что в разные моменты могут сильно отличаться даже у одного и того же сайта.

Табл. 1. Скорость загрузки ведущих сайтов научных библиотек России\*

Место	ТИЦ	Библиотека	Веб-адрес	Время загрузки сайта (секунд)	
				pr-cy.ru	tic-pr.com
1	7600	eLibrary.ru	<a href="http://elibrary.ru">elibrary.ru</a>	–	–
2	6400	РГБ	<a href="http://www.rsl.ru">www.rsl.ru</a>	0,23	0,85
3	5200	ГПНТБ России	<a href="http://www.gpntb.ru">www.gpntb.ru</a>	0,17	0,87
4	4100	«Архивы России»	<a href="http://www.rusarchives.ru">www.rusarchives.ru</a>	0,01	1,14
5	3600	РНБ	<a href="http://www.nlr.ru">www.nlr.ru</a>	0,07	–
6	2800	БЕН РАН	<a href="http://www.benran.ru">www.benran.ru</a>	0,52	–
7	2700	ГПНТБ СО РАН	<a href="http://www.spsl.nsc.ru">www.spsl.nsc.ru</a>	0,57	0,45
8	2700	ГПИБ	<a href="http://www.shpl.ru">www.shpl.ru</a>	0,29	–
9	2100	Отделение ГПНТБ СО РАН	<a href="http://www.prometeus.nsc.ru">www.prometeus.nsc.ru</a>	0,22	1,55
10	1900	ЦНМБ	<a href="http://www.scsml.rssi.ru">www.scsml.rssi.ru</a>	0,02	–

\* В случае ошибки веб-сервера или отсутствия данных поставлен прочерк.

Тест, несмотря на ощутимый разброс скоростей, с которой откликались сервера и каналы (от самой медленной 27,81 Кб/сек у [www.prometeus.nsc.ru](http://www.prometeus.nsc.ru) – до самой быстрой 1048,34 Кб/сек у [www.rusarchives.ru](http://www.rusarchives.ru)), показал приемлемые результаты загрузки документов, которые, безусловно, должны удовлетворять пользователей. Но, в качестве очередной оговорки, укажем на «нюанс», снижающий достоверность проведенного опыта. Анализ доступа по закачке одного лишь HTML-кода не учитывает нагруженности документов графикой, ассоциированными служебными файлами и пр., весьма различной у разных сайтов.

Самым важным компонентом информационного образа сайтов является их «контент», который вместе с оформлением и средствами навигации обеспе-

чивает содержательное (*информационное*) единство сайтов. Собственно, благодаря этому единству, у пользователя и формируется целостное восприятие представленных массивов информации как единых информационных объектов. Именно это представление со временем складывается в репутацию сайтов, то есть общее мнение об их достоинствах и недостатках.

Веб-популярность в значительной степени зависит от сложившейся репутации сайтов, имеющей инерционный характер. Увеличение числа пользователей электронных библиотек – процесс постепенный и длительный, но привлеченная *контактная аудитория* настолько устойчива, что при отсутствии форс-мажорных обстоятельств ее можно считать прочно накопленным потенциалом потребителей. Объем этой аудитории является мерой успеха любого Интернет-проекта и в полной мере зависит от эффективности формирования информационного образа библиотечных сайтов.

Мониторинг результатов деятельности по развитию сайтов производится с помощью широкого спектра внутренних и внешних инструментов, различных средств веб-анализа, веб-маркетинга и веб-цитирования. Из *внутренних* инструментов сайта [Prometeus](#) назовем статистическую систему [AWStats \(версия 6.5\)](#), которая представляет удобное средство автоматической обработки лог-файлов и обеспечивает сбор различных данных на основе модулей определения стран, провайдеров и т. п. AWStats накапливает сведения об уникальных посетителях сайта, количестве визитов, просмотров страниц, о роботах, индексирующих документы, поисковых запросах и пр. Система распространяется свободно и бесплатно с сайта [SourceForge.net](#). В 2009 г. [AWStats \(вер. 6.9\)](#) была установлена на сервере ГПНТБ СО РАН и появилась возможность сравнивать данные двух сайтов [prometeus.nsc.ru](#) и [spsl.nsc.ru](#).

Сопоставление поисковых терминов (ключевых слов и фраз), с которыми пользователи обращались к документам двух сайтов, показывает особенности их образов, сложившихся в представлении потребителей информации (табл. 2). С помощью регулярно проводимых тестов можно анализировать динамику изменения пользовательских предпочтений на уровне запросов к поисковым ма-

шинам и вести управление развитием сайтов. Преимущество используемого инструментария заключается в получении и обработке большого количества показателей, позволяющего отслеживать главные тенденции.

Табл. 2. Сравнение ключевых поисковых фраз сайтов ГПНТБ СО РАН ([www.spsl.nsc.ru](http://www.spsl.nsc.ru)) и Отделения ([www.prometeus.nsc.ru](http://www.prometeus.nsc.ru)) за ноябрь 2010 г.\*

<a href="http://www.spsl.nsc.ru">www.spsl.nsc.ru</a> ☀ AWStats 6.9		<a href="http://www.prometeus.nsc.ru">www.prometeus.nsc.ru</a> ☀ AWStats 6.5		
	использовалась <b>36691</b> фраза	поисков	использовалось <b>58166</b> фраз	поисков
1	<b>герб россии</b>	797	<b>глобальные проблемы современности</b>	566
2	<b>золотая орда</b>	713	<b>политические партии россии в начале 20 в.</b>	341
3	<b>новгородское княжество</b>	589	<b>abbreviation list</b>	183
4	<b>киевская русь</b>	555	<b>гпнтб новосибирск</b>	162
5	<b>гпнтб со ран</b>	524	<b>web of science</b>	155
6	<b>курильские острова история</b>	503	<b>list of abbreviation</b>	151
7	<b>гпнтб новосибирск</b>	485	<b>техногенные катастрофы</b>	126
8	<b>курильские острова</b>	471	<b>глобальные проблемы</b>	125
9	<b>рерих картины</b>	347	<b>демографическая ситуация в россии</b>	117
10	<b>ледовое побоище</b>	328	<b>политические партии россии</b>	95
	остальные фразы	<b>68371</b> (92.7 %)	остальные фразы	<b>79352</b> (97.5 %)

\* По данным статистических систем Advanced Web Statistics библиотечных серверов.

Тесты подтверждают, что наиболее спрашиваемые термины и фразы лидируют не только в рамках ежемесячной статистики, но и по итогам года. Так, десятку Top10 самых популярных запросов сайта [Prometeus](http://Prometeus) на 1.12.2010 г. возглавляют «глобальные проблемы современности» (в разной формулировке) – 5174 поиска, далее идут «справочник англоязычных сокращений» (спрашивался 3170 раз), «политические партии XX века» (1262) и т. д. За неполный год общее количество поисковых фраз превышает 564 тысячи, а количество запросов с их использованием – 910 тыс. (!). Количество обращений к одному лишь библиографическому списку по глобалистике превышает 25 тыс. раз (5-е место по числу просмотров среди 17 тыс. документов сайта). Столь большие размеры веб-обслуживания свидетельствуют об устоявшемся информационном облике, набранном «весе» и репутации сайта Отделения ГПНТБ СО РАН.

Анализ ключевых слов, использованных для доступа к сайту [Prometeus](http://Prometeus), дает еще более показательную картину. Из 270 тыс. поисковых терминов в пер-

вую пятерку, кроме союзов и предлогов, попали слова «Россия» (94 тыс. поисков за 11 мес. 2010 г.) и «проблемы» (39 тыс.). Вряд ли этот результат нуждается в каких-то комментариях, так как четко рисует сложившийся образ сайта. Симптоматично, что среди всех поисковых терминов на 10-м месте присутствует слово «скачать», которое за 11 мес. 2010 г. спрашивалось 21773 раза, но при этом оно практически совсем не употребляется на сайте Отделения ([поисковая машина](#) выдает только 3 второстепенных документа). Как бы не оценивать приведенные факты, но совсем их игнорировать нельзя.

На наш взгляд, использование любых инструментов веб-статистики требует крайней осторожности в оценках, что подтверждают достаточно убедительные примеры. Так, на сайте [Prometeus](#) представ-

лен сравнительно небольшой файл [Od.ssi](#) (около 7,5 Кб), являющийся индексом 9 книжных оглавлений на букву «Д». Многолетний мониторинг обращений к этому файлу показывает, что он давно стал одним из «лидеров спроса». С 9 декабря 2004 г. его посещали не менее 588 тыс. раз, и в этом смысле он конкурирует даже со входной («домашней») страницей сайта (рис. 2).

Первое время делались попытки найти рациональное объяснение такому высокому спросу, анализировалась содержательная сторона документа [Od.ssi](#). Наконец, мы обратились к сплошной выборке из лог-файла и оказалось, что 83% обращений содержат в служебном поле «user-agent» запись «Indy Library», которая позволяет утверждать, что магистральный трафик к таинственной странице [Od.ssi](#) создают спам-боты [3]. Для чего им нужен этот файл – для входа на сайт, сбора e-mail'ов или чего-то еще, – не столь уж и важно. Главное, что эти «роботизированные» посещения значительно искажают статистику реального доступа к ресурсам со стороны «живых» посетителей. Появляется повод

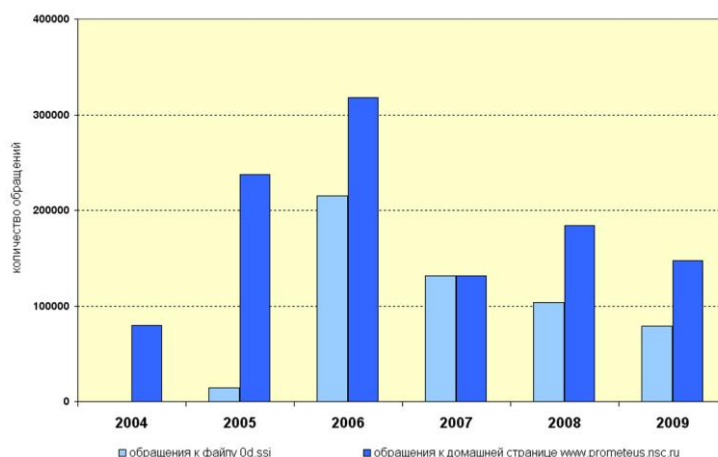


Рис. 2. Сравнение количества обращений к файлу [Od.ssi](#) (голубой) и «домашней» странице сайта [prometeus.nsc.ru](#) (синий) за 2004-2009 гг.

сомневаться в объективности любых цифр, полученных внутренней статистикой сервера. Масса роботов, не отмечающихся в файле [robots.txt](#) и «маскирующихся» под реальных посетителей, легко может ввести в заблуждение доверчивых веб-статистиков, посчитавших популярными совсем не те документы, ко-



Рис. 3. Использование сервиса [cy-pr.com](#) для анализа характеристик информационного образа сайта [prometeus.nsc.ru](#)

(рис. 3): проверить ранги сайта в разных системах, наличие в каталогах, получить список внешних ссылок, проиндексированных страниц и изображений, ссылок с форумов и блогов, анализ ключевых слов и фраз. С помощью разного рода операций можно увидеть какие вопросы пользователей приводят их к документам (например, с [сервиса ответов на Mail.Ru](#) [4]), проследить [историю развития ресурсов сайта](#), если таковая имеется, или даже найти [возможный плагиат](#).

К сожалению, замечено, что многие средства веб-анализа и цитирования нацелены на Интернет-коммерцию и бизнес в сети, что не всегда является подспорьем в работе Интернет-библиотекарей (можно сказать, «компютекарей») и некоммерческих библиотечных сайтов. Системы размещения платных статей ([Miralinks](#) и др.), оценка рекламных возможностей сайта и поисковых терминов в долларах не всегда соответствует библиотечным целям и задачам. Кроме того, поначалу бесплатные, инструменты оценки сайтов затем, с ростом их популярности, имеют тенденцию «внезапно» становиться закрытыми или платными, как это произошло с проверкой популярности ссылок «[Link Popularity Check](#)»,

которые есть на самом деле.

Эффективным способом избежать «обмана» и «самообмана» становится комплексное использование *внешних* инструментов статистики и «многоуровневый» подход к оценке посещаемости. Так, различные средства поисковой оптимизации ([cy-pr.com/analysis](#), [pr-cy.ru/analysis](#),



на которую уже невозможно попасть из-за 403-й ошибки (HTTP Error 403.6 – Forbidden).

Не следует забывать и о других инструментах оценки, предоставляемых крупными поисковыми и тематическими порталами – рейтингах, блогах, форумах, виртуальных справках и пр. Так, в обновленном рейтинге [Топ100 Рэмблера](#) имеются широкие возможности для сравнительного анализа посещаемости и содержания сайтов, активности и географии посетителей, оценки популярности контента с группировкой по страницам и директориям (ресурсам). Поиск по [блогам Яндекса](#) предоставляет веб-ссылки и рекомендации, а также независимые оценки сайтов и комментарии их ресурсов потребителями информации, в том числе и библиотечными работниками [5]. Специализированные библиотечные сервисы, такие как служба [«Виртуальной справки»](#) РНБ или информационно-справочный портал [Library.Ru](#), используют материалы сайта Отделения ГПНТБ СО РАН для обслуживания своих читателей и тоже могут служить инструментом оценки его потребительских свойств [6].

В заключение укажем на основные принципы, необходимые для объективной оценки информационного образа библиотечного сайта. Необходимо:

- использовать максимально возможное количество инструментов анализа и применять комплексный подход;
- проводить постоянный мониторинг изменений;
- учитывать интегральный характер показателей;
- собирать архив данных и изучать динамику их развития;
- проявлять крайнюю осторожность в оценках.

Наконец, следует признать, что вся деятельность по оцениванию и сопоставлению сайтов находится еще в самом зачаточном состоянии, о чем свидетельствует вся практика сайтостроения. Поэтому наиболее актуальной становится задача найти ответы на первоочередные вопросы: о целях оценки сайтов (- для чего?), о критериях для сравнения, особенно в случае значительных различий в контенте (- как?) и о погрешностях оценивания (- насколько?).

## ПРИМЕЧАНИЯ

1. В словарях компьютерной лексики понятие сайта неточно связывают с одним компьютером, подключенным к сети Интернет (Дорот В. Л., Новиков Ф. А. Толковый словарь современной компьютерной лексики. – 3-е изд. – СПб.: БХВ-Петербург, 2004. – С. 423). Такие же неточные определения широко распространились в Интернете, где веб-сайтом называют «совокупность веб-страниц, навигационно и физически находящихся на одном веб-сервере» (см., например: Веб-сайты [Электронный ресурс] // Глоссарий.ru. – URL: [http://www.glossary.ru/cgi-bin/gl\\_sch2.cgi?RCIh-xgpy](http://www.glossary.ru/cgi-bin/gl_sch2.cgi?RCIh-xgpy): (дата обращения: 12.12.2010) ; Глоссарий [Электронный ресурс] // Обучение и бизнес в Интернет. – URL: [http://tva.jino.ru/glos\\_1.htm](http://tva.jino.ru/glos_1.htm) (дата обращения: 12.12.2010) ; Создание, поддержка и раскрутка сайтов [Электронный ресурс] // WSDS. Термины и определения. – URL: [http://www.wds.biz/pub\\_terms.html](http://www.wds.biz/pub_terms.html) (дата обращения: 12.12.2010) ; и др.
2. Мациевский Н. С. Реактивные веб-сайты: клиентская оптимизация в алгоритмах и примерах / Н. Мациевский, Е. Степанищев, Г. Кондратенко. – М.: Интернет-университет информационных технологий, 2010. – С. 6.
3. Любопытно, что количество «входов» на сайт через файл 0d.ssi в 3-5 раз превышает количество «уходов» через него (см. «текущую статистику» сайта. – URL: <http://www.prometeus.nsc.ru/cgi-bin/awstats/awstats.pl>). Одно время проблема спам-ботов «нашего вида» широко обсуждалась на форумах компьютерщиков: URL: <http://xpoint.ru/forums/internet/sitemngmnt/thread/36362.xhtml>, <http://www.phpbbguru.net/community/topic5760.html> (дата обсуждений: конец 2005 – начало 2006 г.; дата обращения: 12.12.2010 г.) и др.
4. Здесь нередко встречаются просто «душераздирающие» послания читателей: «люди спасите срочно, надо чего-нибудь по истории права и побольше» или: «помогите найти материалы по основам исследовательской деятельности студентов». Анализ подобных просьб вносит в деятельность «компютекарей» вполне «живое», естественное начало.
5. На 12.12.2010 г. поиск по блогам Яндекса показал около 900 результатов со ссылками на [«Прометей библиографический»](#), [«Новую Россию»](#) и другие ресурсы сайта [www.prometeus.nsc.ru](http://www.prometeus.nsc.ru).
6. На 12.12.2010 г. [«Виртуальная справка»](#) показала 136 справок (из общего количества в 13,7 тыс.), а [«Library.Ru»](#) – 201 справку из 68,5 тыс., с упоминаниями библиографических списков и других ресурсов сайта Отделения ГПНТБ СО РАН по самой различной тематике. Правда, библиотечный портал [«Library.Ru»](#) предоставляет возможность полноценного поиска по всему массиву справок и документов только зарегистрированным библиотекарям.