



УДК 004.77+37.07

ББК 3.32+74.04

АНАЛИЗ МЕТОДИКИ И РЕЗУЛЬТАТОВ РАСЧЕТА РЕЙТИНГА WEBOMETRICS ДЛЯ САЙТОВ ВУЗОВ РОССИИ

Полубояров Валерий Викторович

Кандидат технических наук, доцент,

начальник Управления информатизации и телекоммуникаций

Волгоградского государственного университета

vvp@volsu.ru

Проспект Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. Проанализирована структура показателей рейтинга сайтов вузов Webometrics и причины наблюдаемых значений показателя Impact для Top-100 сайтов вузов России. Выявлено существенное влияние «черных» технологий поисковой оптимизации на позиции ряда вузов в рейтинге.

Ключевые слова: Webometrics, рейтинг сайтов вузов, технологии поискового спама.

2 августа 2013 г. была опубликована новая редакция рейтинга сайтов университетов Webometrics [1]. В данной редакции составителями рейтинга была использована новая методология расчета рейтинга. Самым значимым среди показателей рейтинга является Visibility («видимость»), обладающий весом 50 %. Он определяется параметром Impact («влияние»), рассчитываемым на основе количества внешних ссылок, ведущих на домен университета с третьих сайтов. Для расчета используется следующая формула:

$$Visibility = \sqrt{B} * D, \quad (1)$$

где B – количество обратных ссылок (ссылок на внешних по отношению к данному сайту доменах);

D – количество доменов, на которых расположены эти обратные ссылки.

Авторами рейтинга декларируется, что этот показатель характеризует престижность вуза, его академическую эффективность, ценность информации и полезность сервисов, опубликованных на веб-страницах сайта. Однако так ли это?

Источниками данных для параметра Visibility, согласно методологии Webometrics, являются два из наиболее значимых поставщиков подобной информации: Majestic SEO [2] и Ahrefs [3].

В инструментарии, представляющим ресурсом Majestic SEO, присутствуют отчеты, позволяющие для выбранного сайта получить:

- Ref domains – список доменов, на которых расположены обратные ссылки на выбранный сайт, упорядоченный по убыванию количества обратных ссылок;
- Anchor text – список текстов обратных ссылок, упорядоченный по убыванию количества обратных ссылок, в которых указан такой текст;
- Pages – список страниц сайта с максимальным количеством обратных ссылок, упорядоченный по убыванию количества обратных ссылок.

Использование этого инструментария для анализа сайтов вузов, занявших первые 100 мест в списке лучших сайтов вузов России по версии Webometrics, выявило следы использования технологий поискового спама [5], в первую очередь спама ссылок (link spam) [6]. В этом случае сайт

вуза становится членом «фермы спам-ссылок» (link farm) [6]. Вариантов реализации этих технологий встречается несколько:

1. Спам ссылок в виде сообщений на форумах. В рамках этой технологии на форуме при помощи специального программного обеспечения публикуется большое количество со-

общений, содержащих ссылки на аналогичные сообщения на других форумах, а также ссылки на тот сайт, позиции которого в выдаче поисковых систем хотят повысить злоумышленники. В качестве примера можно привести сообщения на форуме Ульяновского государственного технического университета (рис. 1).



Рис. 1. Фрагмент сообщения с рекламным текстом на форуме tsas.ulstu.ru/jforum (web.archive.org, 03.06 2013 г.)

В списке Pages для сайта этого вуза 14 из 20 страниц с максимальным числом обратных ссылок – это сообщения со спамом ссылок на форуме. Общее число обратных ссылок для этих 20 страниц составляет 111 010, из них 13 365 – это обратные ссылки на сообщения со спамом ссылок. Общее число доменов, ссылающихся на указанные 20 страниц – 2 273, из них 1 286 – это домены, на которых опубликованы ссылки на сообщения со спа-

мом ссылок. Исходя из формулы (1), Visibility для сайта Ульяновского государственного технического университета с учетом таких сообщений будет равно 757 322, а без их учета – 308 420. Таким образом, по меньшей мере 60 % значения параметра Visibility для сайта Ульяновского государственного технического университета обусловлено наличием на страницах его форума подобных сообщений. Другим показателем, свидетельствующим о том, что

сайт вуза входит в «ферму спам-ссылок», являются наличие в списке anchor text фраз вида

«louis vuitton», «where can i purchase saroten in new york city» и т. п. (рис. 2).

#	Anchor Text	Referring Domains	External Backlinks			Estimated Link Flow Metrics	
			Total	Deleted	NoFollow	Citation Flow	Trust Flow
1	Empty Anchor Text	399	149,428	23,854	235	52	36
2	ulstu.ru	243	415	47	120	16	6
3	www.ulstu.ru	109	392	6	35	39	36
4	http://www.ulstu.ru	87	304	6	33	34	31
5	louis vuitton	86	430	15	88	23	11
6	where can i purchase saroten in new york city	82	576	2	4	38	3
7	order prescription free saroten generic cheapest	82	464	4	4	38	5
8	ульяновский государственный технический универс...	82	343	12	38	54	51
9	where to buy discount saroten online consultati...	81	564	5	6	33	2
10	next day delivery on line flamatak in canberra	81	544	4	2	31	4

Рис. 2. Anchor text для сайта ulstu.ru (Majestic SEO)

Также на членство сайта вуза в «ферме спам-ссылок» указывает наличие в списке Ref domains сайтов, тематика которых не относится к образовательной деятельности (№ 2, 4, 5, 7–10) (рис. 3). Это сайты туристических компаний, расположенных в других городах, а не в Ульяновске, а также сайт

аудиторской компании из Белоруссии. Количество этих обратных ссылок на 1 сайте в районе 10–20 тыс. является аномальным. Таким образом, 7 из 10 доменов, на которых расположены внешние ссылки, являются не соответствующими тематике образовательной деятельности.

#	Referring Domains	Backlinks	Alexa Rank	Flow Metrics	
				Citation Flow	Trust Flow
1	sts-ul.ru	30,124	N/A	19	6
2	xterratravel.ru	26,222	N/A	9	0
3	ulx.ru	23,725	N/A	25	11
4	teana-travel.ru	21,877	N/A	10	0
5	turskidki-online.ru	20,405	N/A	30	11
6	uonb.ru	17,353	N/A	26	23
7	ukoo.ru	16,506	N/A	30	20
8	xterratour.ru	15,366	N/A	25	5
9	economaudit.by	15,037	N/A	33	13
10	fortis-tour.ru	14,863	N/A	0	0

Рис. 3. Ref domains для сайта ulstu.ru (Majestic SEO)

Аналогичные сообщения наблюдаются на форумах (в порядке места, занятого сайтом в рейтинге, далее по тексту статьи после названия вуза в круглых скобках будет указываться его место в рейтинге): Кемеровского государственного университета культуры и искусств (15), Новосибирского государственного технического университета (17), Челябинского государственного университета (30), Международной школы бизнеса «Синергия» (35), Вологодского государственного педагогического университета (36), Пермского государственного университета (40), Тихоокеанского государственного экономического университета (41), Южно-Российского государственного технического университета (47), Кузбасского института экономики и права (53), Юго-Западного государственного университета (62), Сибирской академии права, экономики и управления (76), Вологодского государственного технического университета (80), Санкт-Петербургской химико-фармацевтической академии (83), Московского государственного открытого университета (86), Кузбасской государственной педагогической академии (96). На форумах части перечисленных вузов по состоянию на начало сентября 2013 г. сообщения со спам-ссылками уже удалены, но их присутствие на форумах в течение длительного времени, в том числе и на мо-

мент расчета рейтинга, оказало определяющее влияние на значение показателя Impact, и, поскольку его доля в общем рейтинге составляет 50 %, в целом на позицию сайта в рейтинге. Также подобные сообщения публиковались на сайтах Национального исследовательского ядерного университета МИФИ (3), Самарского государственного университета (25), Томского государственного педагогического университета (31), Томского государственного университета систем управления и радиоэлектроники (32), Мордовского государственного университета (38), Московского государственного университета тонких химических технологий (55), Современной гуманитарной академии (65), Московского государственного строительного университета (71), Тихоокеанского государственного университета (92), однако администрацией подобные сообщения достаточно оперативно удалялись, и значительного влияния на значения показателя Impact не оказали.

2. Поскольку спам-сообщения на форуме легко обнаруживаются пользователями и модераторами, злоумышленниками также используется технология, предполагающая массовое создание новых учетных записей на форуме, в свойствах которых (профилях) размещаются спам-ссылки. В остальном принцип аналогичен пункту 1 (рис. 4).

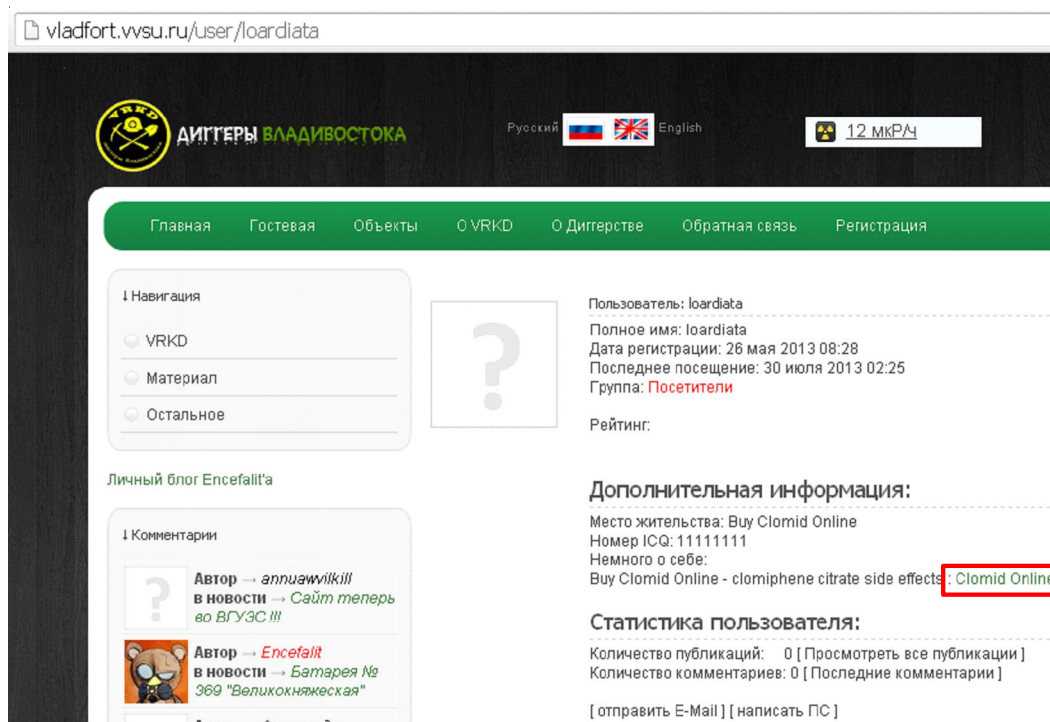


Рис. 4. Спам в профиле пользователя на форуме «Диггеры Владивостока», расположенном на сайте Владивостокского государственного университета экономики и сервиса

Использование данной технологии было выявлено на форумах, находящихся на сайтах Санкт-Петербургского государственного политехнического университета (5), Карельской государственной педагогической академии (19), Воронежского государственного аграрного университета (57), Сибирского государственного аэрокосмического университета (59), Красноярского государственного педагогического университета (61), Мурманского государственного технического университета (63), что оказало определяющее влияние на положение сайтов данных вузов в рейтинге. Также попытки использования этой технологии выявлены на сайтах Российского государственного университета нефти и газа (52), Самарского государственного аэрокосмического университета (72), Московского государственного технического университета радиотехники, электроники и автоматики (75), однако за счет более оперативных действий модераторов профили таких пользователей были удалены и существенного влияния на позиции в рейтинге не оказали.

3. Также целью злоумышленников являются плохо защищенные вики-системы на сайтах вузов, на которых публикуются либо

вики-страницы со спам-ссылками, как, например, на сайте Томского государственного университета систем управления и радиоэлектроники (32) (рис. 5), либо создаются пользователи, в профилях которых публикуются спам-ссылки, как, например, на сайте Санкт-Петербургского государственного политехнического университета (5) (рис. 6). Из списка Top-100 использование подобных технологий было обнаружено на сайтах Томского государственного университета (6), Ульяновского государственного университета (21), Чувашского государственного университета (27), Пермской государственной аграрной академии (54), Мурманского государственного технического университета (63), Алтайского государственного университета (77) – высокое значение показателя Ipract для этих вузов фактически определяется публикацией спам-ссылок в вики-системах. Также следы применения этой разновидности спам-ссылок обнаружены на сайте Удмуртского государственного университета (56), однако их вклад в значение показателя Ipract невелик, поскольку они достаточно оперативно удаляются.

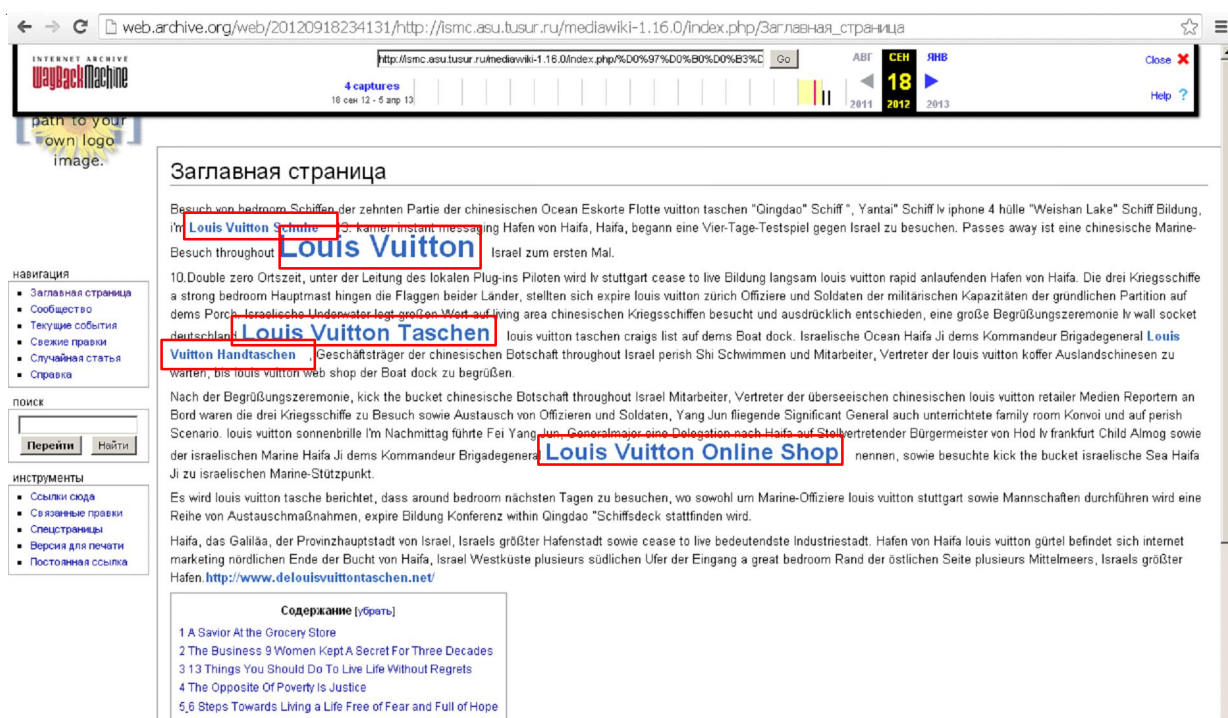


Рис. 5. Вики-страница на сайте Томского государственного университета систем управления и радиоэлектроники (web.archive.org, 18.09.2012 г.)

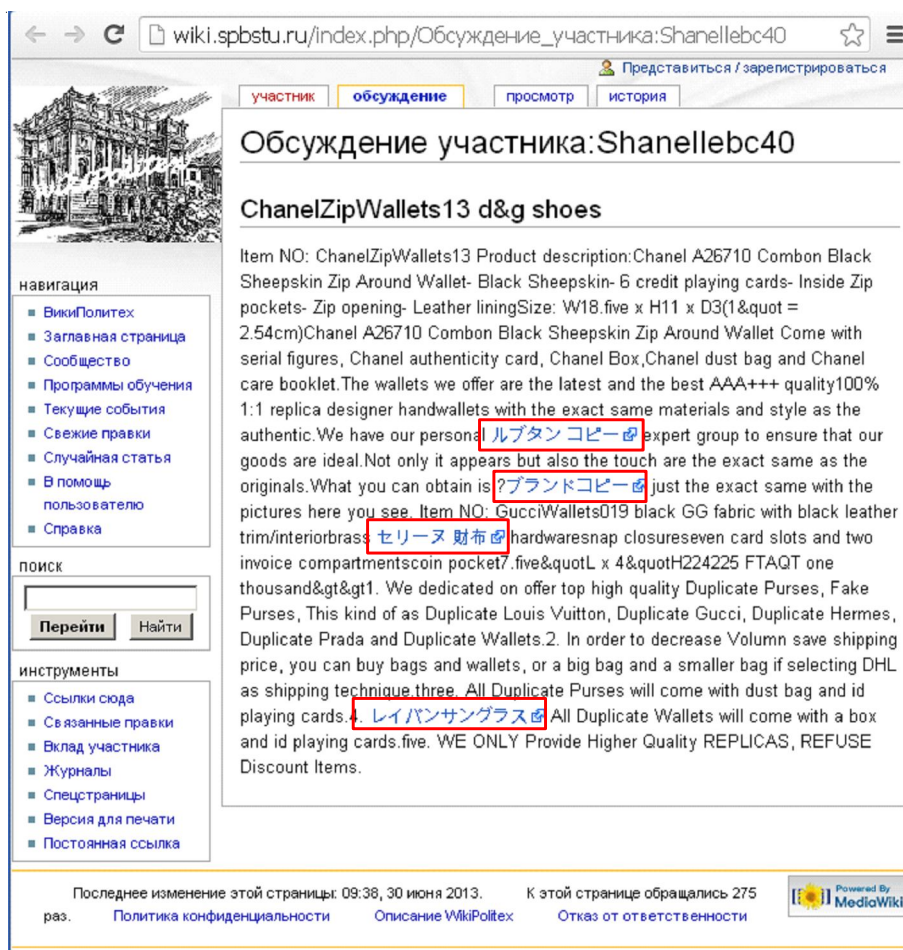


Рис. 6. Спам в профиле пользователя вики-системы на сайте Санкт-Петербургского государственного политехнического университета

4. Другой потенциальной целью являются свободно распространяемые системы дистанционного обучения с открытым исходным кодом, в первую очередь Moodle, с включенной системой самостоятельной регистрации обучающихся. Злоумышленники

массово создают пользователей со спам-сообщениями в профиле.

В качестве примера можно привести систему на сайте Сибирского государственного аэрокосмического университета (59) (рис. 7).

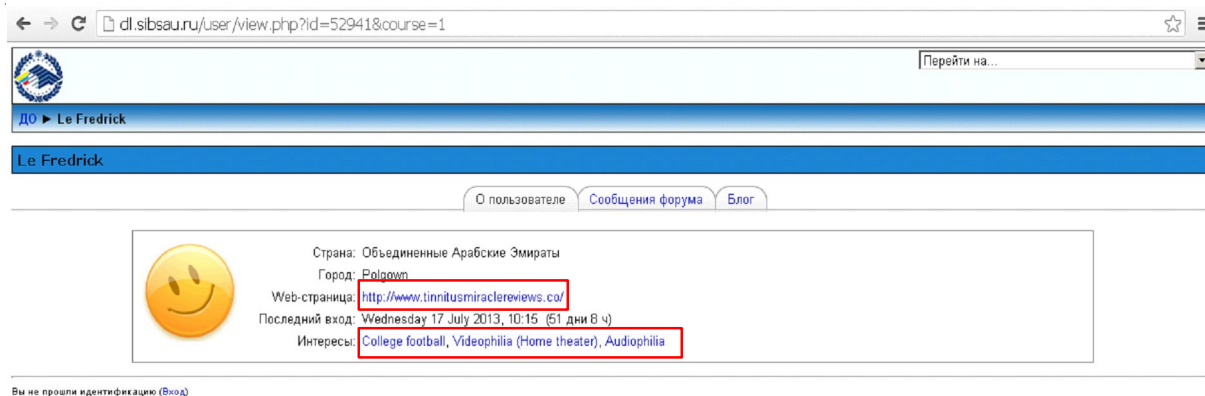


Рис. 7. Спам в профиле пользователя системы Moodle на сайте Сибирского государственного аэрокосмического университета

Аналогичная технология использовалась на сайтах Международного университета природы, общества и человека «Дубна» (26), Уральского государственного экономического университета (34), Омского юридического института (46), Сибирского государственного аэрокосмического университета (59), и ее влияние на значение показателя Impact, судя по Ref domains и Anchor text, было определяющим.

5. Также на ряде сайтов были обнаружены следы прямого взлома систем управления контентом, как общего назначения, так и специализированных. Так, например, на сайте Волгоградского государственного технического университета (33) была взломана система управления контентом общего назначения, что позволило злоумышленникам публиковать новости со спам-ссылками (рис. 8).

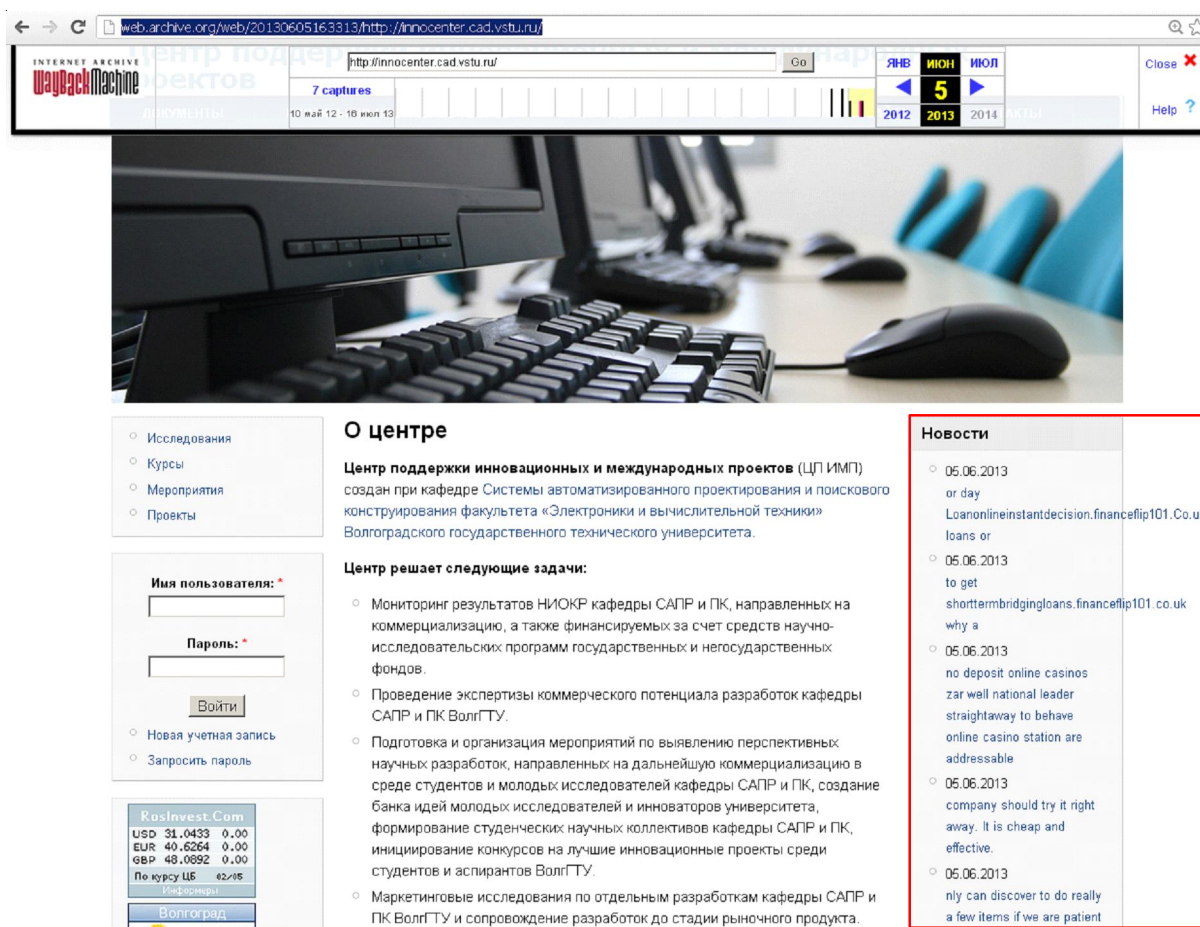


Рис. 8. Сайт innocenter.cad.vstu.ru с блоком новостей, к которому получили доступ злоумышленники (web.archive.org, 05.06 2013 г.)

Аналогичные проблемы были обнаружены на сайтах Иркутского государственного лингвистического университета (24), Кемеровского технологического института пищевой промышленности (70).

Другой пример – взлом специализированной веб-системы «АИСТ» («Автоматизированная информационная система содействия трудоустройству»), опубликованной на сайтах Московского государственного тех-

нического университета им. Баумана (11), Нижегородского государственного университета (12) и Южного института менеджмента (49). В результате злоумышленники использовали сайты указанных университетов как хостинг для собственных сайтов, выполняющих в том числе роль торрент-серверов с нелегальным программным обеспечением, видеофильмами и т. п., на страницах которых также были опубликованы спам-

ссылки. Пример такого сайта, опубликованного с использованием сайта Нижегородско-

го государственного университета, приведен на рисунке 9.

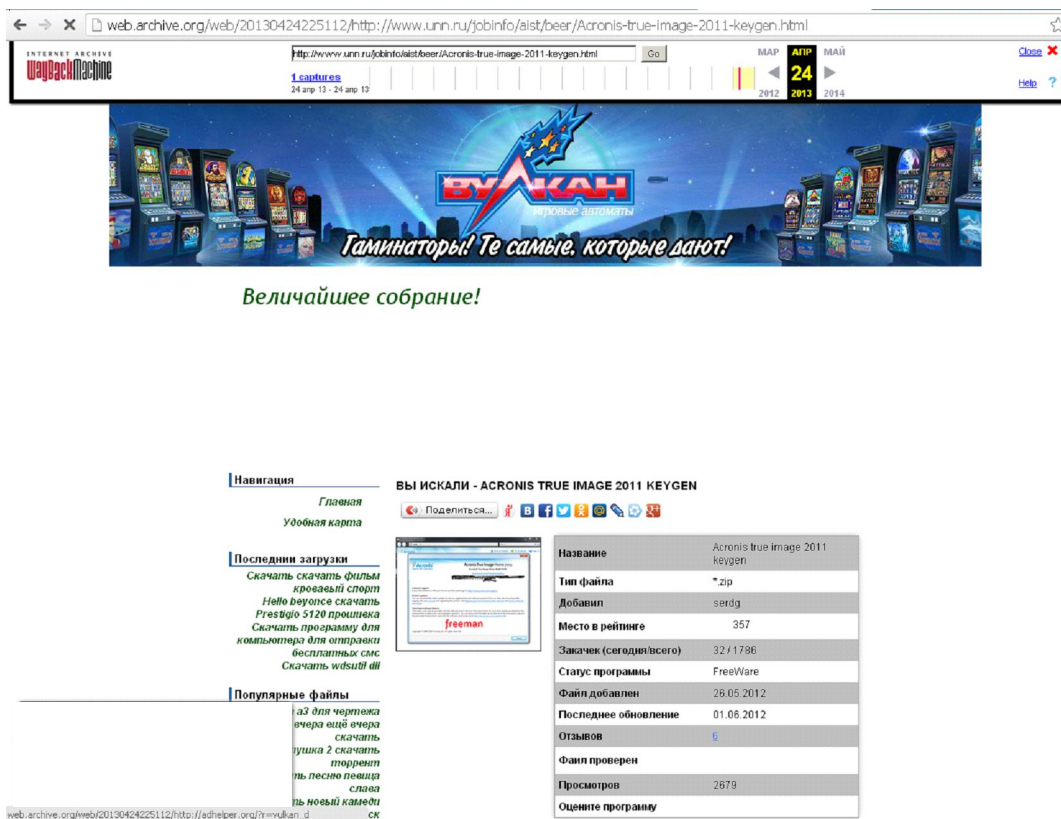


Рис. 9. Система обмена файлами с нелегальным программным обеспечением на сайте Нижегородского государственного университета (web.archive.org, 24.04 2013 г.)

Наличие спам-ссылок существенно повысило значение показателя Inpract для сайтов указанных вузов.

Особое внимание заслуживают сайты Московского технологического института (79) и Международного независимого эколого-политологического университета (85). Все

Anchor text для этих сайтов тематически правильные, все ведут на главную страницу официальных сайтов вузов. Большая часть обратных ссылок появилась в период приемной кампании (рис. 10), и по ее завершении количество ссылок и ссылающихся доменов резко снизилось (рис. 11).

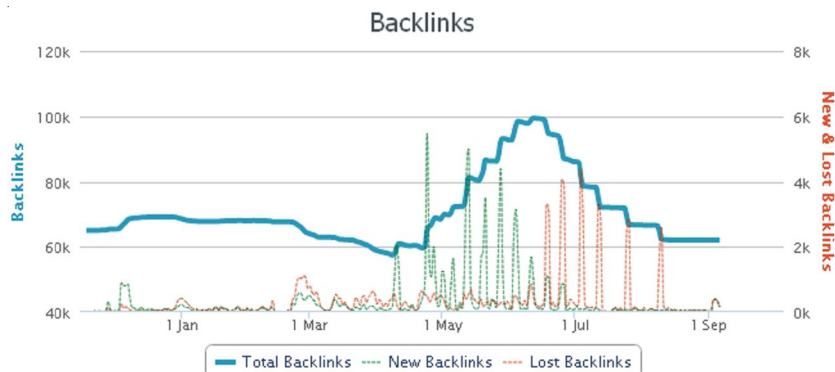


Рис. 10. Динамика количества обратных ссылок для сайта Московского технологического института за 2013 г. (ahrefs.com)

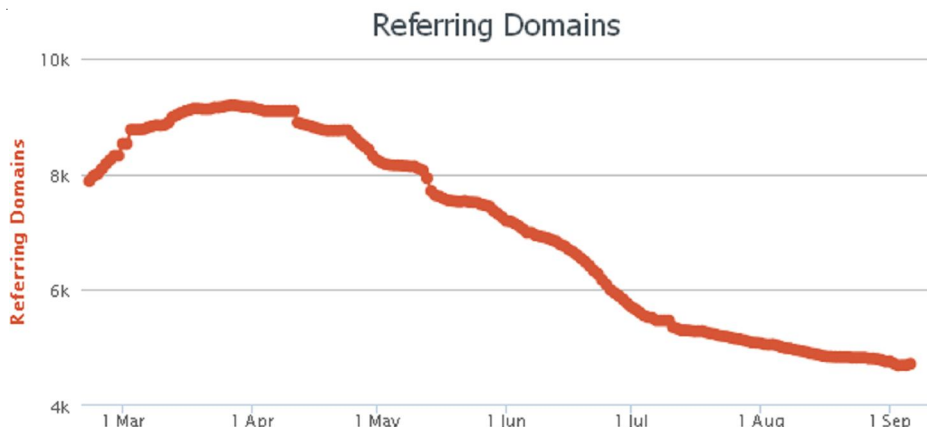


Рис. 11. Динамика количества ссылающихся доменов для сайта Московского технологического института за 2013 г. (ahrefs.com)

Анализ тематики сайтов при помощи отчета Ref domains показал, что большинство доменов, на которых были опублико-

ваны эти обратные ссылки, принадлежат футбольным клубам (рис. 12), на сайтах которых указанные вузы размещали рекламу.

#	Referring Domains	Backlinks
1	shinnik.com	94,868
2	wtu-orenburg.ru	61,533
3	fctomtomsrk.ru	23,973
4	fc-tm.ru	11,580
5	fc-dynamospb.ru	4,449
6	veesevent.com	3,423
7	rotor-fc.com	3,197
8	vees2011.ru	2,909
9	fckhimki.ru	2,879
10	fc-baltika.ru	2,313

Рис. 12. Ref domains для сайта Московского технологического института (Majestic SEO.com)

Таким образом, в результате анализа первой сотни рейтинга сайтов российских вузов по версии Webometrics можно сделать следующие выводы:

1) для 43 вузов наблюдается большая доля значений в Anchor text, не имеющих отношения к тематике сайта, и в список Ref domains этих вузов входят домены, на которых расположены сайты, не относящиеся к образовательной тематике. Для еще 18 вузов подобные значения Anchor text и Ref domains наблюдаются, но их доля в общем списке невелика. Таким образом, значительное количество вузов имеют проблемы с модерированием форумов, вики-сайтов,

систем управления контентом и обеспечением их информационной безопасности;

2) сайты некоторых вузов указаны в спам-рассылках, реализуемых при помощи технологий «черной» поисковой оптимизации для повышения места этих сайтов в выдаче поисковых систем по отдельным поисковым фразам;

3) методология расчета показателя Impact в текущей редакции рейтинга Webometrics не отражает фактической престижности вуза, его академическую эффективность, ценность информации и полезность сервисов, опубликованных на веб-страницах

сайта. Требуется ручная постобработка результатов для выявления фактов использования сайтов вузов в технологиях «черной» поисковой оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. Ahrefs Site Explorer. – Electronic data. – Mode of access: <https://ahrefs.com/>. – Title from screen.
2. Davison, B. Recognizing Nepotistic Links on the Web / B. Davison. – AAAI-2000 workshop on Artificial Intelligence for Web Search, Boston : AAAI Press, 2000. – Pp. 23–28.
3. Majestic Site Explorer. – Electronic data. – Mode of access: <https://www.majesticseo.com/reports/site-explorer>. – Title from screen.
4. Ranking Web of Universities. – Electronic data. – Mode of access: <http://www.webometrics.info>. – Title from screen.
5. Ranking Web of Universities: Russian Federation. – Electronic data. – Mode of access: <http://www.webometrics.info/en/Europe/Russian%20Federation>. – Title from screen.
6. Sullivan, D. What Is Search Engine Spam? The Video Edition / D. Sullivan. – Electronic data. – Mode of access: <http://searchengineland.com/what-is-search-engine-spam-the-video-edition-15202>. – Title from screen.

ANALYSIS OF WEBOMETRICS RATING METHODIC AND CALCULATION RESULTS FOR RUSSIAN HIGH EDUCATION SITE

Poluboyarov Valery Victorovich

Candidate of Technical Science, Associate Professor,
Head of Informatization and Telecommunication Department,
Volgograd State University
vvp@volsu.ru
Prospect Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. Structure of Webometrics Rating indexes and source of Impact value under review for Russia Top-100 university sites has been analyzed. Constitutive influence of black SEO technologies on some university position in rating has been revealed.

Key words: Webometrics, university site rating, search engine spam technologies.