

УДК 004.738.5
ББК 32.973.202

О РЕЙТИНГЕ ОФИЦИАЛЬНЫХ САЙТОВ НАУЧНЫХ УЧРЕЖДЕНИЙ СЕВЕРО-ЗАПАДА РОССИИ

Мазалов В. В.¹, Печников А. А.²

(Учреждение Российской академии наук
Институт прикладных математических исследований
КарНЦ РАН, Петрозаводск)

На основе новых подходов к измерениям известных вебметрических индикаторов и вычислению функций ранжирования, проведено ранжирование сайтов научных организаций РАН северо-запада России. Проанализированы результаты ранжирования по двум методикам и сформулированы предложения по продолжению исследований.

Ключевые слова: вебметрика, Интернет, веб-ресурсы, индикаторы, поисковые машины, ранжирование сайтов.

1. Введение

Термин «вебметрика» (*webometrics*), введенный более 10 лет назад, обозначает раздел информатики, в рамках которого исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к *World Wide Web* (далее – Веб) [17].

Одним из наиболее известных вебметрических проектов является проект испанской исследовательской группы

¹ Владимир Викторович Мазалов, доктор физико-математических наук, профессор (vmazalov@krc.karelia.ru).

² Андрей Анатольевич Печников, кандидат физико-математических наук, доцент (baebron@karelia.ru, моб. тел. +7 921 700 3362).

Cybermetrics Lab «Webometrics Ranking of World Universities» [23], посвященный изучению вебометрических индикаторов и ранжированию на их основе сайтов университетов и научно-исследовательских институтов. Исследования ведутся с 2004 г., накоплен большой объем информации по сайтам 16 000 учреждений высшего образования и 7 000 научных учреждений.

В январе 2009 г. был открыт новый сайт «*Ranking Web of World Research Centers*» [22], посвященный только научно-исследовательским институтам. На сайте опубликован рейтинг «*Top 2000 R&D*», в который входят 59 российских научных сайтов, занимающих достаточно скромные позиции. Например, портал Сибирского отделения РАН занимает 66-е место, портал Российской академии наук – 91-е, сайт Уральского отделения РАН – 587-е, сайт Института математики им. С. Л. Соболева Сибирского отделения РАН – 728-е.

Понятно, что на сегодняшний день большинство мировых веб-ресурсов составляют англоязычные сайты. Чтобы убедиться в том, что и наибольшее количество исследований посвящено англоязычным сегментам Веба, достаточно посмотреть публикации [19, 27] и список литературы к ним. Вместе с тем, с развитием веб-технологий появляется все большее количество исследований, ориентированных на национальные веб-ресурсы университетов и научно-исследовательских институтов, причем не только в Европе и США, но и в Австралии, Латинской Америке, Иране и Нигерии [16, 20, 21, 28].

Россия не является исключением и можно отметить работы, посвященные исследованию и ранжированию научных сайтов Сибирского отделения РАН [6, 14], а также раздел «Рейтинг сайтов научных учреждений СО РАН» официального сайта Института вычислительных технологий [11]. Ранжированию сайтов классических университетов России посвящена работа [9], а в работе [3] исследуются рейтинги университетов северо-запада России и Финляндии.

В продолжение этой тенденции было проведено обследование и ранжирование официальных сайтов научных организаций РАН северо-запада России с использованием подходов, предло-

женных в [14], и их модификаций, сделанных авторами статьи, что позволило выявить ряд интересных моментов, отраженных в данной статье.

2. Целевое множество исследования

Выбор сайтов научных организаций РАН северо-запада России в качестве объекта исследований позволяет получить достаточно обширную выборку научных сайтов (около 10% всех официальных сайтов учреждений РАН), относящихся к разным областям науки, находящихся в различных иерархических отношениях с вышестоящими инстанциями и размещенных достаточно компактно в географическом плане. При этом в целевое множество включались только официальные сайты организаций, имеющие собственные доменные имена. (Такое очевидное сужение множества имеет объяснения, следующие из дальнейшего изложения).

Основным способом выявления сайтов для целевого множества являлся анализ информации, содержащейся в Справочнике РАН [12] и Единой информационной системе РАН [2]. В случае отсутствия данных о доменном имени сайта института выполнялся поиск в поисковых системах *Яндекс* и *Google* по его названию. Каждый найденный веб-адрес проверялся на работоспособность, причем в случае отказов – несколько раз, поскольку существует вероятность отказов на уровне каналов связи и оборудования.

В случае наличия у организации нескольких сайтов в целевое множество включался сайт, удовлетворяющий одному из следующих условий (приоритет сверху вниз):

- включен в перечень информационных систем научных учреждений РАН;
- указан в соответствующем перечне на сайте регионального научного центра;
- на самом сайте сказано, что он является официальным сайтом учреждения.

Было сформировано целевое множество исследования, содержащее 3 сайта региональных научных центров (Санкт-Петербургский, Карельский и Кольский), 24 сайта институтов РАН, географически расположенных в Санкт-Петербурге и Ленинградской области (подчиняющихся Отделениям РАН по областям науки), 1 научное учреждение Санкт-Петербургского научного центра, 7 – Карельского научного центра и 5 – Кольского научного центра. Полный перечень сайтов целевого множества приводится на специализированном сайте «Вебометрика. ИПМИ КарНЦ РАН» [1] в разделе «Ранжирование сайтов РАН». Веб-ресурсы некоторых институтов не вошли в целевое множество, поскольку не имеют собственного доменного имени, а являются директориями вышестоящего домена. Кроме того, несколько доменных имен, указанных в ряде источников как имена официальных сайтов институтов, оказались неработающими (к счастью, их было очень мало).

3. Вебометрические индикаторы, методика измерений и ранжирование сайтов

Вначале определим вебометрические индикаторы S , V , R и Sc в соответствии с [23]:

– под размером сайта (S – *size*) понимается общее количество страниц, обнаруживаемых на сайте поисковыми машинами *Google*, *Yahoo*, *Live Search* и *Exalead*;

– видимость сайта (V – *visibility*) – это количество уникальных гипертекстовых ссылок с других веб-ресурсов, обнаруживаемых *Yahoo Search*, *Live Search* и *Exalead*;

– количество полнотекстовых файлов (R – «*rich files*») – это суммарное количество файлов с расширениями *PDF*, *DOC*, *PS* и *PPT*, обнаруживаемых *Google*;

– и, наконец, научность сайта (Sc – «*scholar*») – это количество ссылок на сайт, обнаруживаемых *Google Scholar*.

Одно из первых замечаний, которое возникает при чтении предыдущего абзаца, заключается в перечне используемых поисковых машин и применимости их к российскому (или

неанглоязычному) Вебу. Можно согласиться с коллегами из Новосибирска, использующими *Яндекс*, *Google* и *Yahoo*, поскольку «именно эти поисковые системы наиболее полно индексируют русскоязычную часть Интернета» [6, 14]. Первые две системы имеют необходимые сервисы для измерений индикаторов и при этом используются более 80% пользователями в России (по данным проекта *LiveInternet* [13]). *Yahoo* – далеко не самая популярная в России поисковая система, однако ее использование практически всеми исследователями основано на развитых сервисах. В частности, предоставляется возможность исключать из множества ссылок на заданный сайт ссылки, сделанные с этого же сайта, что важно при измерениях индикатора *V*. Как и в [14], мы исключим из рассмотрения файлы *PS*, поскольку их количество на российских сайтах весьма незначительно, а сервис *Яндекса* не позволяет их обнаружить.

По поводу измерения индикатора научности сайта *Sc* никаких вариантов, кроме использования системы *Google Scholar*, не остается, поскольку российские аналоги пока слишком неубедительны. Вместе с тем, можно принять дополнение, предложенное в [14], когда в качестве значения *Sc*, измеряемого средствами *Яндекса*, принимается так называемый ТИЦ, – тематический индекс цитирования, – «авторитетность» сайта с учетом качественной характеристики ссылок на них с других сайтов [4].

В данном разделе мы приведем некоторые результаты измерений индикаторов и ранжирования сайтов описанного выше целевого множества по методике Института вычислительных технологий Сибирского отделения РАН. Опишем способы измерения индикаторов, которые будем обозначать, как и в [14], латинскими буквами с индексами, указывающими поисковую систему:

$S_{\text{Яндекс}}$ – на странице <http://webmaster.yandex.ru/check.xml> в строке поиска вводится доменное имя сайта (без *http* и закрывающих \langle / \rangle);

S_{Google} – на странице <http://www.google.ru> в строке поиска вводится *site:доменное имя сайта*;

S_{Yahoo} – на странице <http://www.yahoo.com> в строке поиска вводится доменное имя сайта;

$V_{Яндекс}$ – на странице <http://www.yandex.ru> в строке поиска вводится доменное имя сайта в кавычках, из количества найденных страниц следует вычесть количество страниц, указанных в пункте «Еще с сайта» для измеряемого сайта;

V_{Google} – на странице <http://www.google.ru> в строке поиска вводится *link:доменное имя сайта*;

V_{Yahoo} – на странице <http://siteexplorer.search.yahoo.com> в строке поиска вводится доменное имя сайта, выбирается кнопка *inlinks*, выбирается опция *Except from this domain* и выбирается опция *Entire Site*;

$R_{Яндекс}$ – на странице <http://yandex.ru/advanced.html> в позиции «находятся на сайте» вводится доменное имя сайта, для одно-временного поиска по форматам *PDF*, *DOC* и *PPT* все они выбираются в окне «формат» с удерживанием клавиши *Shift*;

R_{Google} – на странице http://www.google.ru/advanced_search?hl=ru в позиции «домен» вводится доменное имя сайта, в позиции «формат файла» последовательно выбираются *PDF*, *DOC* и *PPT* и результаты суммируются;

R_{Yahoo} – на странице <http://www.yahoo.com> вводится доменное имя сайта и выбирается кнопка *Web Search*, выбираются *Options*, *Advanced Search*, затем последовательно в позиции *File Format* выбираются *PDF*, *DOC*, *PPT* и результаты суммируются;

$Sc_{Яндекс}$ – на странице <http://yaca.yandex.ru/yaca> в строке поиска вводится доменное имя сайта;

Sc_{Google} – на странице <http://scholar.google.com> в строке поиска набирается «+доменное имя сайта».

Полностью результаты замеров вебметрических индикаторов для целевого множества приведены в [1]. Учитывая динамику Веба и постоянную деятельность роботов поисковых машин, отметим, что измеренное значение индикатора – это значение, полученное данной поисковой машиной в конкретный момент времени.

Ранжирование сайтов целевого множества проводилось по методике, которую можно увидеть в Вебе по ссылке [11], но для

удобства мы вкратце опишем ее здесь. Рейтинг сайта вычислялся по формуле

$$W = V + S + 2R + 1,5Sc,$$

где

$$V = \log_{10}(V_0); V_0 = [V_{\text{Яндекс}} + V_{\text{Google}} + V_{\text{Yahoo}}]/3;$$

$$S = \log_{10}(S_0); S_0 = [S_{\text{Яндекс}} + S_{\text{Google}} + S_{\text{Yahoo}}]/3;$$

$$R = \log_{10}(R_0); R_0 = [R_{\text{Яндекс}} + R_{\text{Google}} + R_{\text{Yahoo}}]/3;$$

$$Sc = [\log_{10}(Sc_{\text{Яндекс}}) + \log_{10}(Sc_{\text{Google}})]/2.$$

Полный ранжированный список всех 40 сайтов целевого множества приведен в [1]; первая десятка выглядят так:

1. Физико-технический институт им. А. Ф. Иоффе РАН;
2. Зоологический институт РАН;
3. Институт проблем машиноведения РАН;
4. Петербургский институт ядерной физики РАН;
5. Институт русской литературы РАН (Пушкинский дом);
6. Кольский научный центр РАН;
7. Кунсткамера (Музей антропологии и этнографии РАН);
8. Пулковская астрономическая обсерватория РАН;
9. Санкт-Петербургский институт информатики и автоматизации РАН;
10. Институт лингвистических исследований РАН.

4. О применимости поисковых машин в качестве «измерительных устройств» и уточнении понятия «единица анализа»

Алгоритмы работы поисковых роботов и механизмы индексации страниц являются секретными особенностями поисковых машин, о которых можно лишь догадываться [7]. Однако полученные авторами результаты измерений не могут не вызвать вопроса о применимости коммерческих поисковых машин для измерения вебметрических индикаторов. Критические публикации на эту тему появились достаточно давно и продолжают появляться [18, 24, 25], что, однако, не останавливает исследователей, имеющих в качестве «измерительных устройств» только поисковые системы.

Продемонстрируем особенности измерений общего количества страниц на примере сайта Карельского научного центра РАН (КарНЦ РАН) www.krc.karelia.ru. Бросается в глаза существенная зависимость результатов от поисковой машины: $S_{\text{Яндекс}} = 30\,000$, $S_{\text{Google}} = 5\,670$, а $S_{\text{Yahoo}} = 18$.

Детальный анализ перечня страниц сайта www.krc.karelia.ru, проиндексированных Яндексом, показывает, что сюда же отнесены страницы самостоятельных сайтов, имеющих доменные имена 4-го уровня (*rcdl2009.krc.karelia.ru*, *tender.krc.karelia.ru*, *mathem.krc.karelia.ru* и многие другие). В то же время измерение значения $S_{\text{Яндекс}}$, к примеру, для сайта Института прикладных математических исследований КарНЦ РАН дает значение 811, т. е. этот сайт рассматривается Яндексом как самостоятельная единица анализа. Отсюда следует, что реальное значение $S_{\text{Яндекс}}$ для КарНЦ РАН существенно завышено за счет самостоятельных сайтов, имеющих доменные имена 4-го уровня, входящие в домен третьего уровня *krc.karelia.ru*. Практически та же ситуация наблюдается и при измерениях $R_{\text{Яндекс}}$.

Конечно, эти ситуации могут быть обойдены, если нам известен полный перечень всех доменных имен так называемой «доменной зоны». К сожалению, эта информация известна далеко не всегда. Более того, она ничем не поможет нам в случаях измерений посредством *Google* и *Yahoo* ($S_{\text{Google}} = 5\,670$ и $S_{\text{Yahoo}} = 18$), хотя и добавит понимания того факта, что поисковый робот обходит не весь Веб. Авторам достоверно известно, что реальное количество страниц на сайте КарНЦ РАН чуть меньше 17 000. Но вряд ли мы когда-нибудь достоверно узнаем, почему *Google* индексирует примерно треть из них, а *Yahoo* – лишь тысячную часть. (Одна из авторских гипотез зависимости $S_{\text{Яндекс}}$ от реального размера сайта и/или его структурной организации приводится в [10]).

По поводу измерений количества уникальных гипертекстовых ссылок с других веб-ресурсов, обнаруживаемых поисковыми машинами, также приведем пример для КарНЦ РАН: $V_{\text{Яндекс}} = 215$, $V_{\text{Google}} = 189$ и $V_{\text{Yahoo}} = 1\,025$. Эти результаты трудно поддаются объяснению, поэтому лишь сошлемся на критиче-

скую работу [18], в которой показано, что для конкретных случаев *Google* скрывает от 48 до 70% проиндексированных им же страниц, содержащих ссылки на заданный сайт.

Соображения по поводу применения в качестве «измерительных устройств» не поисковых машин, а другого программного обеспечения, будут изложены в следующем разделе.

Остановимся подробнее на вопросе о том, что считать единицей анализа при ранжировании сайтов. В разделе «Целевое множество исследований» было отмечено, что авторы в рамках данного исследования для каждого учреждения РАН анализируют только официальный сайт. При этом понятным кажется и подход, используемый в [14] и [22], когда в случае наличия у организации нескольких доменных имен используется их совокупность.

Расширим такой подход до уровня веб-ресурсов крупной организации в целом. Тогда следует вести речь о сложном информационном комплексе, являющемся в каком-то смысле отражением ее организационно-управленческой и научной структуры. Как правило, в этом информационном комплексе существует так называемая точка входа (основной сайт организации), а далее следуют: административный сайт, сайты лабораторий, сайты выполняемых проектов, страницы сотрудников и т. д. При этом некоторые подразделения крупных институтов (в особенности это свойственно подразделениям, профессионально связанным с информационными технологиями), имеют веб-ресурсы, зарегистрированные под именами, не содержащими доменного имени основного сайта организации. И в противовес им могут существовать сайты организаций, содержащих доменное имя основного сайта, но не имеющих отношения к его научной деятельности (сайт профсоюзной организации института – это еще куда ни шло, но могут быть и сайты туристических клубов и др.).

В этом случае только содержательный анализ ресурса может дать ответ на вопрос, является ли этот ресурс частью веб-ресурсов организации. Учитывая, что мы сталкиваемся с определенными проблемами даже на стадии выявления официаль-

ных сайтов, хотелось бы говорить о веб-ресурсах организации как о некотором официально утвержденном и опубликованном перечне. Только в этом случае под единицей анализа можно понимать «веб-ресурсы организации в целом». В противном случае очень многое зависит от субъективных знаний исследователей об институтах и трактовок о том, следует ли считать некоторый сайт веб-ресурсом организации или нет. Например, использование в качестве единицы анализа не только доменного имени официального сайта Института прикладных математических исследований КарНЦ РАН (ИПМИ КарНЦ РАН), но и всех его веб-ресурсов, известных авторам, изменяет положение в рейтинге с 34-го на 16-е.

5. Смешанный подход к измерениям индикаторов и модифицированный рейтинг

В 2008 г. в ИПМИ КарНЦ РАН стартовал проект «Вебметрические исследования научных Интернет-ресурсов Российского Интернета»¹. Опыт, накопленный в процессе исследования более 200 официальных сайтов РАН [1], позволяет сделать некоторые предложения как по поводу измерений вебметрических индикаторов, так и о соответствующей модификации формул вычисления рейтинга сайтов.

В рамках указанного проекта для сборки гиперссылок, исходящих с заданного сайта, для их последующего анализа был разработан робот *LPR* (от слов *Link*, *Page* и *Robot*). Побочным эффектом от работы *LPR* является получение реальных значений количества всех страниц отсканированных сайтов. Таким образом, мы получаем вебметрические индикаторы S и R , измеренные не поисковыми машинами, а *LPR* (обозначим их как S_{LPR} и R_{LPR}), значения которых можно использовать при ранжировании сайтов.

¹ Поддержан Российским фондом фундаментальных исследований (грант №08-07-00023а).

Основное опасение, которое могло бы сдерживать использование LPR и подобных ему роботов для измерения значений S и R , заключается в слишком больших размерах сайтов, и, следовательно, в неоправданных затратах ресурсов. Однако проведенные исследования показывают, что на сегодняшний день среди сайтов научных организаций РАН лишь Портал РАН содержит общее количество страниц большее чем 500 000. Сайты, содержащие десятки тысяч страниц (что приемлемо по затратам ресурсов для их сканирования), исчисляются десятками, а значительное количество официальных сайтов содержит до тысячи страниц. Единственный большой сайт, отсканированный LPR не до конца в нашем случае – это сайт Физико-технического института им. А. Ф. Иоффе РАН (полученных значений и так достаточно, чтобы вывести его на первое место в рейтинге).

Таким образом, вместо значений S и R , измеренных с помощью поисковых машин, предлагается взять значения, измеренные LPR . (К сожалению, поступить подобным образом с индикаторами V и Sc мы не можем по причине слишком большой размерности Веба). Тогда модифицированная формула для вычисления рейтинга сайта будет выглядеть следующим образом:

$$WM = V + S_M + 2R_M + 1,5Sc,$$

где

$$V = \log_{10}(V_0), \quad V_0 = [V_{\text{Яндекс}} + V_{\text{Google}} + V_{\text{Yahoo}}]/3;$$

$$S_M = \log_{10}(S_{LPR}), \quad R_M = \log_{10}(R_{LPR});$$

$$Sc = [\log_{10}(Sc_{\text{Яндекс}}) + \log_{10}(Sc_{\text{Google}})]/2.$$

Полный ранжированный список всех 40 сайтов целевого множества по модифицированной формуле для WM приведен в [1], а первая десятка выглядят так:

1. Физико-технический институт им. А. Ф. Иоффе РАН;
2. Зоологический институт РАН;
3. Институт проблем машиноведения РАН;
4. Карельский научный центр РАН;
5. Кольский научный центр РАН;
6. Кунсткамера (Музей антропологии и этнографии РАН);

7. Институт лингвистических исследований РАН;
8. Институт прикладной астрономии РАН;
9. Петербургский институт ядерной физики РАН;
10. Ботанический институт им. В. Л. Комарова РАН.

Сравнивая места сайтов первой десятки по критериям W и WM , нетрудно убедиться в том, что списки совпадают на 70%. (Коэффициент корреляции между значениями W и WM для всего целевого множества равен 0,81, см. [1]).

Поясним на двух примерах причины, по которым, как нам кажется, в первую десятку по критерию WM не вошли сайты с очевидно высоким рейтингом. Сайт Института русской литературы РАН (Пушкинский дом) имеет доменное имя второго уровня *pushkinskijdom.ru*, но при этом имеются его разделы, адресуемые как доменные имена третьего уровня (например, раздел «Электронные публикации Института русской литературы» *lib.pushkinskijdom.ru*, содержащий более 6 000 страниц) и другие разделы. В данном случае, как и в случае с сайтом КарНЦ РАН, $S_{Яндекс}$ имеет гораздо большее значение, чем S_{LPR} .

Сайты некоторых институтов робот LPR не смог отсканировать по другой причине. Известно, что роботы, сканирующие Веб, могут доставлять владельцам сайтов ряд проблем, таких как рост трафика, вторжение в закрытую информацию, нарушение авторских прав и поведение, похожее на деятельность программ-вирусов [26]. Поэтому владельцы сайтов и/или веб-серверов могут вводить ограничения доступа к их содержимому. Из Кодекса этического поведения АСМ [15], в частности, следует, что не должны прилагаться усилия по преодолению введенных ограничений доступа, и разработчики робота LPR следуют эти правилам. В рамках данной статьи невозможно дать развернутого ответа на вопрос о том, нарушается ли этика поведения роботами поисковых машин. Короткий ответ может выглядеть так: возможно, не нарушается, а по ответам на запросы выдаются данные из архивов, сформированных раньше, чем были установлены ограничения доступа к сайтам.

6. Заключение и предложения

Проделанная работа по ранжированию сайтов научных организаций РАН северо-запада России позволяет сделать ряд выводов и рекомендаций для дальнейших исследований.

Соглашаясь с определением основных вебометрических индикаторов, отметим, что выбор тех или иных «измерительных устройств» является дискуссионным. В ряде работ, включая и данную, показаны примеры, заставляющие задуматься о надежности применимости поисковых машин. В то же время, самостоятельные разработки программного обеспечения также могут быть подвергнуты сомнению до тех пор, пока не докажут свою надежность. Это вопрос, который должен быть тщательно исследован и обсужден специалистами.

В то же время показано, что выбор «измерительных устройств» и методик ранжирования не слишком влияет на положение явных лидеров. Ранжирование по двум различным методикам на замерах, сделанных для одних и тех же единиц анализа (в нашем случае это доменные имена официальных сайтов), как продемонстрировано в статье, дает результаты по первой десятке, совпадающие на 70 процентов. (Интересно, что результаты ранжирования по методике *Cybermetrics Lab* [23], которые не приводятся в данной статье, дают еще лучшее совпадение!) Причем сайты теряют высокие места именно в связи с тем, что происходит сужение совокупности веб-ресурсов института, а не из-за изменения функции вычисления значения рейтинга или методики измерения индикаторов.

По-видимому, более существенным на сегодня вопросом является вопрос о том, что следует считать единицей анализа при ранжировании сайтов. Как было отмечено, даже само выявление сайтов некоторых научных учреждений сегодня представляет определенные трудности, тем более трудновыполнимой представляется задача обнаружения всей совокупности веб-ресурсов одного учреждения. Возможно, следовало бы на официальном уровне (например, в отделениях и научных центрах РАН) разработать рекомендации по содержательному наполне-

нию первых страниц официальных сайтов институтов. Одной из рекомендаций могло бы быть наличие раздела под названием «Веб-ресурсы института (организации, учреждения)», так, как это сделано на первых страницах сайтов ИПМИ КарНЦ РАН и Петрозаводского госуниверситета [5, 8].

Вебометрические исследования рейтингов сайтов являются хорошей основой для сравнительного анализа некоторых характеристик уже созданных веб-ресурсов, а значит, позволяют определить направления их дальнейшего развития.

Литература

1. *Вебометрика. Институт прикладных математических исследований КарНЦ РАН* [Электронный ресурс] – Режим доступа: <http://webometrics.krc.karelia.ru>.
2. *Единая информационная система РАН* [Электронный ресурс] – Режим доступа: <http://www.ras.ru/scientificactivity/eis.aspx>.
3. ИЛЮКЕВИЧ О. Г., ПЕЧНИКОВ А. А. *Рейтинг официальных web-сайтов университетов России и Финляндии: сравнительный анализ* // Информационные ресурсы России. – 2008. – №3(103) – С.25-28.
4. *Индекс цитирования* [Электронный ресурс] – Режим доступа: <http://help.yandex.ru/catalogue/?id=873431>.
5. *Институт прикладных математических исследований КарНЦ РАН* [Электронный ресурс] – Режим доступа: <http://mathem.krc.karelia.ru>.
6. КЛИМЕНКО О. А., ШОКИН Ю. И. *О рейтинге сайтов научных организаций СО РАН* // Наука в Сибири. – 2008 г. – №44 (2679). – URL: <http://www-sbras.nsc.ru/HBC/hbc.phtml?5+479+1> (дата обращения: 18.03.2009).
7. НЕКРЕСТЬЯНОВ И., ПАНТЕЛЕЕВА Н. *Системы текстового поиска для Веб* [Электронный ресурс] // Группа исследования методов организации информации. – URL: <http://meta.math.spbu.ru/~nadejda/papers/web-ir/web-ir.html> (дата обращения: 18.03.2009).

8. *Петрозаводский государственный университет* [Электронный ресурс] – Режим доступа: <http://petsu.karelia.ru>.
9. ПЕЧНИКОВ А. А. *Вебометрические исследования Web-сайтов университетов России* // Информационные технологии. – 2008. – №11 – С. 74-78.
10. ПЕЧНИКОВ А. А., ЛУГОВАЯ Н. Б. *Измерения научных сайтов* // Телематика'2008. Труды XV Всероссийской научно-методической конференции, С-Петербург, 2008 г. – Т. 1. – С. 166.
11. *Рейтинг сайтов научных учреждений СО РАН* [Электронный ресурс]. – Режим доступа: <http://www.ict.nsc.ru/ranking>.
12. *Российская академия наук. Справочник 2007 года. Часть II* [Электронный ресурс] – Режим доступа: <http://www.ras.ru/fstorage/download.aspx?id=209736f0-13f2-4509-882f-d90161a9b120> (дата обращения: 18.03.2009).
13. *Статистика сайта «Сайты Рунета»* [Электронный ресурс] – Режим доступа: <http://www.liveinternet.ru/stat/ru/searches.html>.
14. ШОКИН Ю. И., КЛИМЕНКО О. А., РЫЧКОВА Е. В., ШАБАЛЬНИКОВ И. В. *Рейтинг сайтов научных организаций СО РАН* // Вычислительные технологии. – 2008. – Т. 13, №3. – С. 128-135.
15. *ACM code of ethics and professional conduct* // Association for Computing Machinery : сайт. – URL: <http://www.acm.org/constitution/code.html> (дата обращения: 18.03.2009).
16. AGUILLO I. F. *Indicadores de contenidos para la web académica iberoamericana [Content indicators for Iberoamerican academic webs]* // BiD: textos universitaris de biblioteconomia i documentació. – 2005. – №15: сайт. – URL: http://www2.ub.edu/bid/consulta_articulos.php?fichero=15aguil2.htm (дата обращения: 18.03.2009).
17. ALMIND T., INGWERSEN P. *Informetric analyses on the World Wide Web: Methodological approaches to «webometrics»* // Journal of Documentation. – 1997. – №53(4). – P. 404-426.

18. Bar-Ilan J. *How much information do search engines disclose on the links to a web page? A longitudinal case study of the 'cybermetrics' home page* // Journal of Information Science. – 2002. – Vol. 28, No. 6. – P. 455-466.
19. BJÖRNEBORN L, INGWERSEN P. *Toward a basic framework for webometrics* // Journal of the American Society for Information Science. – 2004. – Vol. 55, Iss. 14. – P. 1216-1227.
20. BOELL S., WILSON C., COLE F. *A Webometric Analysis of Australian Universities using Staff and Size dependent Web Impact Factors (WIF)* // Fourth International Conference on Webometrics, Informetrics, and Scientometrics & Ninth COLLNET Meeting, Berlin, 2008. – URL: <http://www.collnet.de/Berlin-2008/BoellWIS2008waa.pdf> (дата обращения: 18.03.2009).
21. NORUZI A. *Web Impact Factors for Iranian Universities* // Webology. – 2005. – Vol. 2, Num. 1 : сайт. – URL: <http://www.webology.ir/2005/v2n1/a11.html> (дата обращения: 18.03.2009).
22. *Ranking Web of World Research Centers* [Электронный ресурс] – Режим доступа: <http://research.webometrics.info>.
23. *Ranking Web of World Universities* [Электронный ресурс]. – Режим доступа: <http://www.webometrics.info>.
24. SNYDER H.; ROSENBAUM H. *Can search engines be used as tools for web-link analysis? A critical view* // Journal of documentation. – 1999. – Vol. 55(4). – P. 375-384.
25. THELWALL M. *Web impact factors and search engine coverage* // Journal of Documentation. – 2000. – Vol. 56(2). – P. 185-189.
26. THELWALL M. *Web Crawling Ethics Revisited: Cost, Privacy and Denial of Service* // Journal of the American Society for Information Science and Technology. – 2006. – №57. – P. 1771-1779.
27. THELWALL M., VAUGHAN L., BJÖRNEBORN L. *Webometrics* // Annual Review of Information Science and Technology. – 2005. – Vol. 39, Iss. 1. – P. 81-135.

28. UTULU S. C. *Webometric Ranking and Nigerian Private Universities: A Case Study of Bells University of Technology, Ota* // E-prints in Library and Information Science. – 2007. – URL: <http://eprints.rclis.org/9007/1/F0927C9D.pdf> (дата обращения: 18.03.2009).

ON THE RATING OF OFFICIAL SITES OF SCIENTIFIC INSTITUTIONS OF THE RUSSIAN NORTHWEST REGION

Vladimir Mazalov, Institute of Applied Mathematical Research of Karelian Research Centre of RAS, Petrozavodsk, Doctor of Science, professor (vmazalov@krc.karelia.ru).

Andrey Pechnikov, Institute of Applied Mathematical Research of Karelian Research Centre of RAS, Petrozavodsk, Cand.Sc., assistant professor (baebron@karelia.ru, моб. тел. +7 921 700 3362).

Abstract: On the basis of new approaches to measurements known webometrics indicators and to computation of ranging functions, ranging of sites of the scientific organizations of the Russian Academy of Sciences of the northwest of Russia is spent. Results of ranking by two techniques are analyzed and offers on continuation of researches are formulated.

Keywords: webometrics, Internet, web-resources, indicators, search engines, ranking of sites.

Статья представлена к публикации членом редакционной коллегии Д.А. Новиковым