

УДК 004.738.5:51-7

## ПРИМЕНЕНИЕ ВЕБОМЕТРИЧЕСКИХ МЕТОДОВ ДЛЯ ИССЛЕДОВАНИЯ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА НАУЧНОЙ ОРГАНИЗАЦИИ (НА ПРИМЕРЕ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН)

**А. А. Печников**

*Институт прикладных математических исследований  
Карельского научного центра РАН*

На примере Карельского научного центра РАН показано применение вебометрических методов для исследования веб-сайтов, составляющих информационное веб-пространство крупной научной организации для выработки рекомендаций, направленных на улучшение его присутствия в Вебе.

**Ключевые слова:** вебометрика, веб-пространство, веб-граф, индикаторы присутствия в Вебе.

### **A. A. Pechnikov. APPLICATION OF WEBOMETRIC TECHNIQUES FOR INVESTIGATION OF THE WEB SPACE OF SCIENTIFIC ORGANIZATION (A CASE STUDY OF KARELIAN RESEARCH CENTRE OF RAS)**

The use of webometric techniques for the study of web sites in the information web space of a large scientific organization in order to produce recommendations on improving its presence on the Web is shown with the example of Karelian Research Center of RAS.

**Key words:** webometrics, web space, web graph, indicators of presence on the Web.

---

### **ВВЕДЕНИЕ**

К актуальным направлениям вебометрики, – одного из направлений информатики, – относятся исследования гиперссылок в Вебе [4]. Вследствие гигантской размерности Веба во многих случаях исследования проводятся на его достаточно узких фрагментах, таких, например, как множество сайтов университетов Великобритании [5] или Северной Европы [2].

Взаимодействие веб-сайтов в информационном веб-пространстве осуществляется посредством гиперссылок, связывающих сайты веб-пространства и определяющих его внутреннюю структуру. Наличие (или отсутствие) гиперссылок влияет на их присутствие в Вебе: известно, что увеличение количества гиперссылок на сайт связано с ростом индекса цитирования веб-сайта [11] и увеличивает (потенциально) количество переходов с одного сайта на другой, а значит, и количество посетителей.

Веб-ресурсы научных учреждений относятся к так называемым «регламентируемым

веб-ресурсам». Это означает, что существует (или, выразимся мягче, — «должен существовать») официальный документ, в котором изложены цели и задачи веб-ресурса, определена структура сайта, правила изменения информации и т. д. Под управляемостью процессами в Вебе понимается реализация управленческих решений в виде пунктов регламента, определяющих порядок их исполнения. В этом случае для сайтов информационного веб-пространства организации могут быть выработаны управленческие решения, направленные на улучшение их присутствия в Вебе. Однако прежде чем реализовывать управленческие решения, необходимо пройти путь от сбора информации о присутствии и взаимосвязях веб-сайтов до построения и анализа формальных моделей и разработки рекомендаций, сделанных на их основе. Целью данной работы является изучение взаимодействия веб-сайтов, составляющих информационное веб-пространство достаточно крупной научной организации для выработки рекомендаций, направленных на улучшения его присутствия в Вебе.

В качестве объекта исследования выбрано веб-пространство Карельского научного центра РАН (КарНЦ РАН), в состав которого входят 39 веб-сайтов, включая официальные сайты 7 институтов различной направленности. Представляется, что такой объект исследования является достаточно типичным для российской академической науки, и поэтому методы, апробированные здесь, могут быть перенесены на другие научные центры и крупные институты.

В работе используются методы, разработанные для исследований академического [8] и университетского фрагментов российского Веба [9]. Для получения, хранения и обработки вебметрической информации используется специализированный комплекс программ VeeBot, в состав которого входят поисковый робот для сканирования веб-сайтов с целью сбора информации о них, база данных, предназначенная для хранения и обработки собранной информации применительно к задачам вебметрических исследований и средства визуализации [10]. Кроме того, используется ряд расширенных возможностей поиска в Google.

## ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Веб-сайт — совокупность html-страниц и веб-документов, связанных внутренними гиперссылками и обладающих единством содер-

жания, идентифицируемый в Вебе по уникальному доменному имени.

Уровень страницы сайта определяется следующим образом: начальная страница сайта, определяемая по уникальному доменному имени, имеет уровень 0; уровень любой другой страницы сайта — это минимальное количество внутренних гиперссылок, ведущих от начальной страницы к данной. Под контекстом гиперссылки в общем случае понимаются языковые выражения, окружающие гиперссылку в пределах веб-страницы [3], в более узком смысле рассматривается текст, сопровождающий гиперссылку на странице (так называемый «анкор»).

Уникальной прямой внешней гиперссылкой называется гиперссылка из множества всех гиперссылок, имеющих одинаковый контекст и адрес целевой страницы, сделанная с исходной страницы с наивысшим уровнем; если на наивысшем уровне таких ссылок несколько, то выбирается первая из них. Далее рассматриваются именно такие гиперссылки, поэтому вместо словосочетания «уникальная прямая внешняя гиперссылка» будут использоваться термины «гиперссылка» или просто «ссылка».

Официальный веб-сайт организации — веб-ресурс, обеспечивающий официальное представление информации в Вебе об организации, являющейся юридическим лицом, создаваемый и функционирующий в соответствии с нормативным актом (регламентом) организации-владельца сайта, определяющим цели и задачи официального веб-сайта, структуру и порядок размещения в Вебе информационных материалов, образующих веб-ресурс, права, обязанности и лиц, осуществляющих программно-техническую поддержку данного сайта, и лиц, осуществляющих предоставление информации для размещения на сайте.

Веб-сайт подразделения — веб-ресурс подразделения, входящего в состав организации и не являющегося юридическим лицом, обеспечивающий представление информации в Вебе о данном подразделении.

Тематический веб-сайт — веб-ресурс организации, не являющийся официальным веб-сайтом или веб-сайтом подразделения, создаваемый и функционирующий для представления в Вебе информации на заданную тему.

Неофициальный веб-сайт — веб-ресурс организации, обеспечивающий представление информации в Вебе об организации, являющейся юридическим лицом, не являющийся официальным сайтом организации.

Информационное веб-пространство – это множество взаимосвязанных посредством гиперссылок веб-сайтов перечисленных типов.

### ИНФОРМАЦИОННОЕ ВЕБ-ПРОСТРАНСТВО КАРНЦ РАН

Веб-пространство КарНЦ РАН состоит из 39 взаимосвязанных сайтов (разобьем их на 7 групп, далее в скобках количество сайтов в группе):

1. Официальный сайт КарНЦ РАН (1),
2. Официальные сайты институтов, входящих в состав КарНЦ РАН (7),

3. Сайты подразделений КарНЦ РАН (2),
4. Тематические сайты КарНЦ РАН (10),
5. Тематические сайты институтов КарНЦ РАН (15),
6. Неофициальные сайты институтов КарНЦ РАН (3),
7. Сайт Карельской региональной организации профсоюза РАН (1).

Перечень веб-сайтов, входящих в веб-пространство КарНЦ РАН, приводится в табл. 1.

Таблица 1. Сайты веб-пространства Карельского научного центра РАН

№	Название организации/подразделения/сайта	Гр.	Доменное имя сайта	Стр.	Исх.	R
1	Карельский научный центр РАН	1	www.krc.karelia.ru	36000	1962	0,44
2	Институт биологии КарНЦ РАН	2	ib.krc.karelia.ru	1809	513	0,41
3	Институт водных проблем Севера КарНЦ РАН	2	water.krc.karelia.ru	1708	496	0,35
4	Институт геологии КарНЦ РАН	2	ig.krc.karelia.ru	1369	551	0,48
5	Институт леса КарНЦ РАН	2	forestry.krc.karelia.ru	2500	494	0,47
6	Институт прикладных математических исследований КарНЦ РАН	2	mathem.krc.karelia.ru	5141	1435	0,15
7	Институт экономики КарНЦ РАН	2	economy.krc.karelia.ru	3187	641	0,40
8	Институт языка, литературы и истории КарНЦ РАН	2	illhportal.krc.karelia.ru	2193	562	0,79
9	Вычислительный кластер КарНЦ РАН	3	cluster.krc.karelia.ru	24	24	0,27
10	Научная библиотека КарНЦ РАН	3	library.krc.karelia.ru	49	106	0,38
11	Электронная библиотека КарНЦ РАН	4	dl.krc.karelia.ru	43124	247	0,27
12	Труды КарНЦ РАН	4	transactions.krc.karelia.ru	927	1419	0,30
13	Зеленый пояс Фенноскандии	4	green-belt.krc.karelia.ru	462	763	0,27
14	ГЭП-анализ на территории Республики Карелия	4	gap-analysis.krc.karelia.ru	565	794	0,72
15	Проект HINTERLAND	4	hinterland.krc.karelia.ru	417	1	0,59
16	Экологический туризм на службе Муезерского района	4	muetacis.krc.karelia.ru	179	3	0,60
17	Развитие национального парка Койтайоки-Толвоярви	4	suotacis.krc.karelia.ru	111	5	0,54
18	Ладога	4	ladoga.krc.karelia.ru	652	1	0,23
19	Биоразнообразие Карелии	4	biodiv.krc.karelia.ru	68	133	3,43
20	Природные ресурсы Карелии	4	ias.krc.karelia.ru	15	40	0,44
21	Вебметрика	5(6)	webometrics.krc.karelia.ru	617	324	0,27
22	III Всероссийская научная конференция «Электронные библиотеки»	5(6)	rcdl2001.krc.karelia.ru	381	20	1,86

23	XI Всероссийская научная конференция «Электронные библиотеки»	5(6)	rcdl2009.krc.karelia.ru	579	51	0,32
24	Базы данных по топливно-энергетическому хозяйству Карелии	5(6)	energy.krc.karelia.ru	150	1	0,27
25	Разработка методов моделирования телекоммуникационных сетей	5(6)	morozov.krc.karelia.ru	52	7	0,75
26	Документация по программированию	5(6)	info.krc.karelia.ru	786	230	0,43
27	Журнал «МТИ& П»	5(6)	mgta.krc.karelia.ru	123	415	0,15
28	Российское ботаническое общество. Карельское отделение	5(2)	rbo.krc.karelia.ru	233	435	0,60
29	Геоботанический сайт	5(2)	geobotany.krc.karelia.ru	397	978	0,24
30	Топонимия Европейского Севера - TORIS	5(8)	toris.krc.karelia.ru	144	27	0,27
31	Фольклорный архив Института языка, литературы и истории КарНЦ РАН	5(8)	folk.krc.karelia.ru	89	4	0,32
32	Фонограммархив Института языка, литературы и истории КарНЦ РАН	5(8)	phonogr.krc.karelia.ru	70	30	0,38
33	Вепский корпус	5(8)	vepsian.ru	745	2	0,27
34	Петроглифы Фенноскандии	5(8)	petroglyphs.ru	206	4	0,73
35	Отделение Русского географического общества в Республике Карелия	5(3)	rgo.karelia.ru	164	7	0,53
36	Институт водных проблем Севера КарНЦ РАН (неофиц.)	6(3)	nwpi.krc.karelia.ru	180	25	0,66
37	Институт геологии КарНЦ РАН	6(4)	geoserv.krc.karelia.ru	2567	29	0,19
38	Институт биологии КарНЦ РАН	6(2)	biology.krc.karelia.ru	476	45	0,52
39	Карельская организация Профсоюза РАН	7	profcom.krc.karelia.ru	24	16	0,35

В столбце «гр.» обозначена принадлежность сайта к одной из 7 указанных групп, цифра в скобках соответствует порядковому номеру сайта. Например, у сайта «Труды КарНЦ РАН», являющегося тематическим сайтом КарНЦ РАН, значение «гр.» равно 4. У сайта «Вепский корпус», – тематического сайта Института языка, литературы и истории КарНЦ РАН, – значение «гр.» равно 5(8), где 5 обозначает принадлежность к группе, а 8 – принадлежность к институту (т. е. порядковый номер института в общем списке сайтов).

Из табл. 1 видно, что подавляющее большинство сайтов имеют доменные имена третьего уровня в зоне основного домена krc.karelia.ru, и лишь три сайта не ассоциированы с этим доменом.

Сайты сканировались на максимальную глубину до 5-го уровня, что считается достаточным для нахождения основных значимых гиперссылок, требуемых для построения

теоретико-графовых моделей. Вследствие этого, значения в столбцах «стр.» (количество страниц сайта) и «исх.» (количество исходящих гиперссылок), являются оценками снизу.

Всего на сайтах КарНЦ РАН отсканировано около 90 тысяч страниц, найдено почти 12 500 исходящих гиперссылок. Всего два сайта можно назвать большими – сайты КарНЦ РАН и Электронной библиотеки. Семь официальных сайтов институтов (плюс один неофициальный сайт института) имеют размеры в пределах от 1300 до 5150 страниц. Причем неофициальный сайт Института геологии почти в два раза больше официального. Остальные сайты имеют менее 1000 страниц. Корреляция между количеством страниц на сайте и количеством исходящих гиперссылок не обнаруживается.

Важным индикатором присутствия сайтов в Вебе считается количество документов, размещенных на них в виде файлов с расшире-

ниями .doc, .pdf, .ppt и др. (так называемые Rich Files) [6]. По указанной ссылке можно посмотреть ранжирование 120 российских научных учреждений РАН. Заметим, что КарНЦ в этот список не входит, а из региональных научных центров 20-е место занимает Коми научный центр Уральского отделения РАН (КомиНЦ УрО РАН).

Результаты сканирования сайтов веб-пространства КарНЦ РАН показывают, что основное количество Rich Files размещено на сайте КарНЦ РАН ([www.krc.karelia.ru](http://www.krc.karelia.ru)) — более 3500 файлов .doc и .pdf. Кроме того, активным участником информационного веб-пространства КарНЦ РАН является закрытый ресурс с адресом [resources.krc.karelia.ru](http://resources.krc.karelia.ru), используемый как хранилище файлов. По 4480 ссылкам на него, сделанным с других сайтов КарНЦ РАН, можно установить, что на нем содержится более 540 файлов .doc и 230 файлов .pdf. Например, сайт Зеленый пояс Фенноскандии ([green-belt.krc.karelia.ru](http://green-belt.krc.karelia.ru)) отправляет к 300 файлам .doc, размещенным на [resources.krc.karelia.ru](http://resources.krc.karelia.ru). Однако более половины сайтов КарНЦ РАН не имеют Rich Files, в том числе и на [resources.krc.karelia.ru](http://resources.krc.karelia.ru).

Сказанное подтверждается замерами, сделанными в Google: команда «[site:krc.karelia.ru filetype:doc](http://site:krc.karelia.ru filetype:doc)» (количество файлов типа .doc на домене [krc.karelia.ru](http://krc.karelia.ru)) выдает 570 результатов, а «[site:krc.karelia.ru filetype:pdf](http://site:krc.karelia.ru filetype:pdf)» — 1370. Та же команда, примененная к домену КомиНЦ УрО РАН, дает следующие результаты: «[site:komisc.ru filetype:doc](http://site:komisc.ru filetype:doc)» — 310 результатов, а «[site:komisc.ru filetype:pdf](http://site:komisc.ru filetype:pdf)» — 5070.

Теперь о гиперссылках. Из более чем 12000 найденных гиперссылок, исходящих с сайтов КарНЦ РАН, половина являются гиперссылками, связывающими эти сайты. Ещё 4480 ссылок, как уже было сказано выше, сделано на закрытый ресурс [resources.krc.karelia.ru](http://resources.krc.karelia.ru). Остальные 1600 ссылок сделаны на 930 сайтов так называемого «сопутствующего множества» [9]. На рис. 1 представлен веб-граф, построенный на 969 вершинах, соответствующих 39 сайтам информационного веб-пространства КарНЦ РАН и 930 сайтам сопутствующего множества. Здесь нарисованы только исходящие дуги с вершин, соответствующих сайтам целевого множества (дуга изображена, если существует хотя бы одна гиперссылка, связывающая два соответствующих сайта).

Если на рис. 1 оставить только дуги, связывающие сайты целевого множества, то получим веб-граф, приведенный на рис. 2.

В овал заключен официальный сайт КарНЦ РАН, в верхнем прямоугольнике на-

ходятся сайты подразделений и тематические сайты КарНЦ РАН, в прямоугольниках в нижней части объединены все сайты, принадлежащие данному институту (официальный, и, если есть, то тематические и неофициальный).

К сайтам информационного веб-пространства КарНЦ РАН был применен подход, предложенный еще в [1]:

$$R(s) = (1-d) + d(R(t_1)/C(t_1) + \dots + R(t_n)/C(t_n)),$$

где  $R(s)$  — значимость сайта  $s$ ,

$d$  — коэффициент затухания (который обычно берут равным 0,85),

$n$  — количество сайтов, ссылающихся на сайт  $s$ ,

$R(t_i)$  — значимость  $i$ -го сайта, имеющего ссылки на сайт  $s$ ,

$C(t_i)$  — количество ссылок с  $i$ -го сайта.

Значения  $C(t_i)$  несложно посчитать для каждого сайта по матрице смежности веб-графа. Фрагмент матрицы смежности, показывающий количество гиперссылок, связывающих некоторые сайты, приведен на рис. 3. Как видно из него, веб-сайты имеют очень большое количество гиперссылок на «головной» сайт и обратно, и весьма незначительное между собой.

Содержательно значение  $R(s)$  можно трактовать как значимость  $s$ -го сайта для других сайтов, входящих в заданное множество. Значения  $R$  для сайтов информационного веб-пространства КарНЦ РАН приведены в последней колонке табл. 1.

Достаточно предсказуемым результатом является высокая значимость официального сайта КарНЦ РАН как головной организации. Высокая позиция сайта Института прикладных математических исследований по сравнению с другими институтами также имеет свое объяснение (хотя и частичное) — многие сайты КарНЦ РАН имеют ссылки на него как на их разработчика. Официальные сайты институтов биологии, геологии и водных проблем Севера занимают последние места среди институтов, что может быть объяснено, в том числе, и наличием у них неофициальных сайтов, «отнимающим» ссылки у официальных.

Если в веб-графе, построенном на сайтах информационного веб-пространства КарНЦ, оставить только дуги, которым соответствует более двух гиперссылок, связывающих сайты, то получим веб-граф, изображенный на рис. 4 (изолированные вершины опущены). Максимальная компонента сильной связности этого веб-графа имеет диаметр, равный 3, и содержит 12 вершин, соответствующих сайту КарНЦ РАН, официальным сайтам всех



Рис. 1. Веб-граф, построенный на 969 вершинах

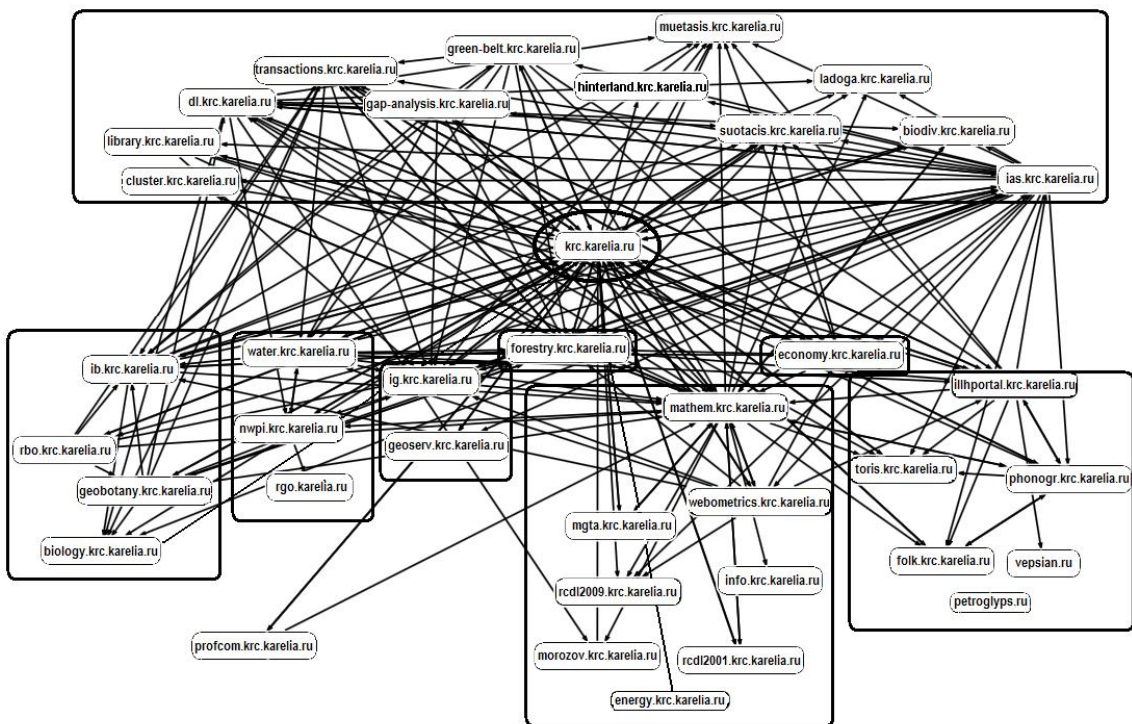


Рис. 2. Веб-граф, построенный на сайтах информационного веб-пространства КарНЦ (количество гиперссылок между сайтами больше 0)

	ias.krc.karelia.ru	ib.krc.karelia.ru	ig.krc.karelia.ru	illhportal.krc.karelia.ru	info.krc.karelia.ru	krc.karelia.ru
ias.krc.karelia.ru						4
ib.krc.karelia.ru	1					440
ig.krc.karelia.ru	1					517
illhportal.krc.karelia.ru	1					322
info.krc.karelia.ru						
krc.karelia.ru	1	19	12	19		

Рис. 3. Фрагмент матрицы смежности веб-графа КарНЦ РАН

семи институтов, и еще 4 сайтам — Электронная библиотека КарНЦ РАН, Зеленый пояс Фенноскандии, ГЭП-анализ на территории Республики Карелия и Труды КарНЦ РАН (на рис. 4 выделены овалами). Сайты, входящие в максимальную компоненту связности, можно назвать ядром информационного веб-пространства КарНЦ РАН: они образуют максимальную компоненту сильной связности, имея большое количество гиперссылок между собой (4834 из 6000).

#### **ВЗАИМОСВЯЗИ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА КАРНЦ РАН С РОССИЙСКИМИ ВЕБ-САЙТАМИ**

Сайты информационного веб-пространства КарНЦ содержат около 1150 гиперссылок, сделанных на 640 российских сайтов, относящихся

к сопутствующему множеству. Как оказалось, самое большое количество гиперссылок сделано на сайт фирмы, занимающейся компьютерными технологиями (правда, на него ссылаются лишь 2 сайта КарНЦ РАН). Вторым следует сайт РАН (32 ссылки с 15 сайтов КарНЦ РАН). На 490 сайтов сопутствующего множества сделано лишь по одной гиперссылке, из чего следует, что каждый из них интересен только одному сайту КарНЦ РАН.

Представляется существенным вопрос о том, как много гиперссылок сделано на сайты информационного веб-пространства КарНЦ с наиболее «популярных» у сайтов КарНЦ РАН российских сайтов сопутствующего множества. Для ответа на него были отобраны 15 сайтов, имеющих гиперссылки не менее чем с 4 сайтов КарНЦ РАН, перечисленные в табл. 2.

Таблица 2. 15 сайтов сопутствующего множества

№	Название организации/сайта	Доменное имя сайта	Сайты	Ссылки на сайт	Ссылки с сайта
1	РАН	www.ras.ru	15	32	8
2	LiveInternet.Ru	www.liveinternet.ru	13	13	0
3	Петрозаводский государственный университет	www.petrso.ru	8	26	4
4	РФФИ	www.rfbr.ru	8	14	1
5	Карельская государственная педагогическая академия	www.kspu.karelia.ru	7	13	1
6	Ботанический институт РАН	www.binran.ru	6	22	0
7	Учёные записки ПетрГУ	uchzap.petrso.ru	6	15	0
8	Портал органов власти Республики Карелия	gov.karelia.ru	5	6	0
9	WWW.KARELIA.RU	www.karelia.ru	5	6	0
10	Русское географическое общество	www.rgo.ru	5	5	0
11	РОМИП	www.romip.ru	4	15	5

12	Конференция «Электронные библиотеки»	www.rcdl.ru	4	10	4
13	Музей-заповедник «Кижы»	kizhi.karelia.ru	4	7	5
14	Некоммерческое партнерство «Север-Центр»	www.northcentre.ru	4	7	3
15	Теория управления организационными системами	www.mtas.ru	4	6	0

В колонке «сайты» указано количество сайтов КарНЦ РАН, имеющих ссылки на данный сайт, а в колонке «ссылки на сайт» – общее число ссылок с этих сайтов. В колонке «ссылки с сайта» указано количество ссылок, сделанных с данного сайта на сайты КарНЦ РАН.

Суммируя значения в двух последних колонках, получаем, что с сайтов КарНЦ РАН на 15 наиболее популярных сайтов сделано 197 ссылок, в то время как с этих сайтов на сайты КарНЦ РАН сделана всего 31 ссылка.

Можно сказать, что встречные ссылки между сайтом РАН и сайтами КарНЦ РАН во многом носят характер «начальник-подчиненный» и отражают реальное административное взаимодействие организаций РАН. Очевидное отражение в Вебе нашло взаимодействие КарНЦ РАН и Музея-заповедника «Кижы».

Однако вряд ли можно считать отражением реального взаимодействия наличие лишь одной ссылки с сайта Карельской государственной педагогической академии на сайты КарНЦ РАН. В меньшей степени, но это касается и сайта Петрозаводского государственного университета.

#### **ВЗАИМОСВЯЗИ ИНФОРМАЦИОННОГО ВЕБ-ПРОСТРАНСТВА КАРНЦ РАН С ЗАРУБЕЖНЫМИ ВЕБ-САЙТАМИ**

Сайты информационного веб-пространства КарНЦ содержат около 420 гиперссылок, сделанных на 290 зарубежных сайтов, относящихся к сопутствующему множеству (причем 130 гиперссылок сделано с сайта info.krc.karelia.ru). Наибольшее количество гиперссылок (а именно 6) сделано на сайт Electrical and Computer Engineering and Computer and Information Sciences networks (eecis.udel.edu) с одного сайта info.krc.karelia.ru. По три сайта КарНЦ РАН ссылаются на такие сайты, как Lake Model FLake (lakemodel.net) и Oulu University Library

(herkules.oulu.fi). Большинство остальных ссылок с сайтов КарНЦ РАН на зарубежные сайты можно охарактеризовать как «один сайт — одна-две ссылки». В целом можно говорить о невысокой активности взаимодействий с зарубежными сайтами.

Обратных ссылок с зарубежных сайтов, на которые ссылаются сайты КарНЦ РАН, по большому счету, не обнаружено. Исключениями являются сайт Lake Model FLake, имеющий 17 гиперссылок на официальный сайт Института водных проблем Севера и сайт Hinterland (hinterland-info.net), имеющий три ссылки на своего карельского коллегу Проект HINTERLAND.

Таким образом, реальное взаимодействие КарНЦ РАН с зарубежными организациями и коллегами отражения в Вебе практически не находит.

#### **ВЫВОДЫ И РЕКОМЕНДАЦИИ**

Информационное веб-пространство КарНЦ РАН можно охарактеризовать как внутренне связанное, имеющее четко выраженное ядро, включающее наиболее значимые сайты. Однако высокая связность информационного пространства во многом обеспечивается одной-двумя гиперссылками, связывающими многие пары сайтов, что является отражением в Вебе скорее подчиненности, чем совместной деятельности. Обратим также внимание на то, что 9 сайтов КарНЦ вообще не имеют исходящих гиперссылок на другие сайты веб-пространства, а один сайт и вовсе изолированный.

Объемы сайтов КарНЦ РАН достаточно скромны, однако в целом соответствуют тенденциям, присущим институтам РАН. Например, сайт Института русской литературы РАН (www.pushkinskiydom.ru) содержит более 39000 страниц, а сайт Зоологического института РАН (www.zin.ru) — 1100. Более важной характеристикой присутствия в Вебе является



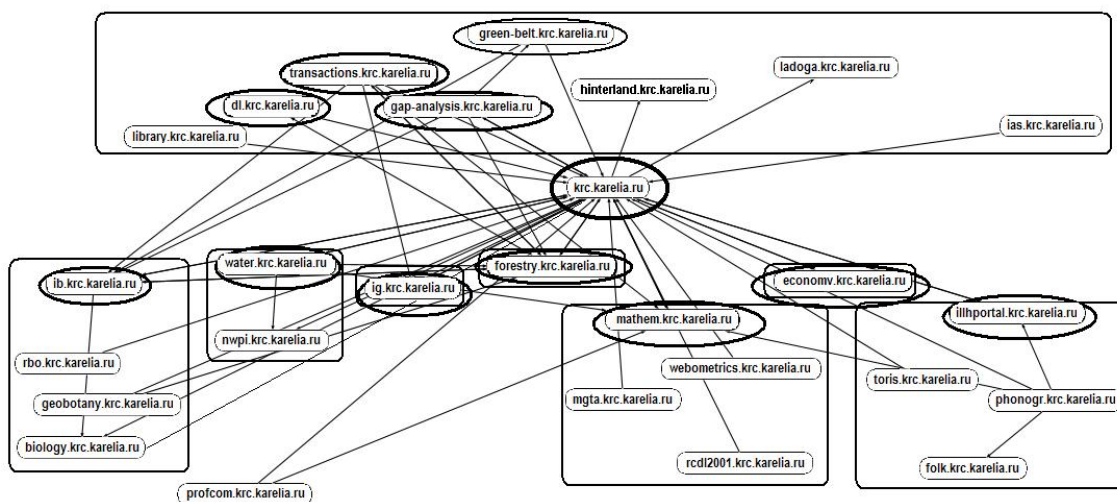


Рис. 4. Веб-граф, построенный на сайтах информационного веб-пространства КарНЦ (количество гиперссылок между сайтами больше 2)

ся количество выставленных на сайтах документов (так называемых Rich Files), а многие сайты КарНЦ РАН не имеют их вообще.

Взаимосвязи информационного веб-пространства КарНЦ РАН с российскими и зарубежными сайтами выглядят односторонне: большому количеству исходящих гиперссылок с сайтов КарНЦ РАН не соответствует большое количество ссылок, сделанных на эти сайты.

Возможно, это связано со слабой информированностью о наличии тех или иных веб-ресурсов КарНЦ РАН. Одним из характерных примеров слабой информированности коллег является ссылка с сайта Русского географического общества ([www.rgo.ru](http://www.rgo.ru)) на веб-сайт Отделения РГО в Республике Карелия ([kareliya.rgo.ru](http://kareliya.rgo.ru)) и при этом отсутствие ссылки на другой сайт Отделения РГУ в Республике Карелия ([rgo.karelia.ru](http://rgo.karelia.ru)), входящий в веб-пространство КарНЦ РАН. Сайт [kareliya.rgo.ru](http://kareliya.rgo.ru) является «дочерним» сайтом [www.rgo.ru](http://www.rgo.ru), поэтому о нем знают в Москве и делают на него ссылку, а сайт [rgo.karelia.ru](http://rgo.karelia.ru), по-видимому, неизвестен в Русском географическом обществе. Таких примеров недостаточной информированности при наличии реальной совместной деятельности можно привести много.

Из сделанных общих выводов можно сделать некоторые краткие рекомендации по усовершенствованию информационного веб-пространства КарНЦ РАН.

Необходимо обратить внимание на внутренние ссылки между сайтами веб-пространства КарНЦ РАН. Наверное, да-

леко не исчерпаны возможности веб-сайтов [hinterland.krc.karelia.ru](http://hinterland.krc.karelia.ru), [ladoga.krc.karelia.ru](http://ladoga.krc.karelia.ru) и [muetacis.krc.karelia.ru](http://muetacis.krc.karelia.ru), которые могли сослаться хотя бы на официальный сайт КарНЦ РАН. Сайты [petroglyphs.ru](http://petroglyphs.ru) и [vepsian.ru](http://vepsian.ru) не имеют ссылок на сайт Института языка и литературы, [biology.krc.karelia.ru](http://biology.krc.karelia.ru) – на сайт Института биологии, [rgo.karelia.ru](http://rgo.karelia.ru) – на сайт Института водных проблем Севера, а [rcdl2009.krc.karelia.ru](http://rcdl2009.krc.karelia.ru) и [info.krc.karelia.ru](http://info.krc.karelia.ru) – на сайт Института прикладных математических исследований.

Улучшению положения КарНЦ РАН в мировом рейтинге [6] (а точнее – его появлению в этом рейтинге), несомненно, должно помочь увеличение на сайтах веб-пространства КарНЦ РАН Rich Files. Во-первых, если у разработчиков есть такая возможность, то следует сделать открытым для поисковых машин хранилище файлов [resources.krc.karelia.ru](http://resources.krc.karelia.ru).

Во-вторых, на сайтах проектов вполне были бы уместны отчеты о проделанной работе в формате .doc или .pdf. Сюда же можно отнести и файлы различных презентаций, которые в большом числе делаются научными сотрудниками КарНЦ РАН. Пример может продемонстрировать сайт КарНЦ РАН, когда в новостной ленте информирует о научных докладах, сделанных на очередном заседании Президиума КарНЦ РАН: вместо фотографий (или вместе с фотографиями) полезно поставить и презентации научных докладов в формате .ppt.

Более сложной задачей представляется получение новых гиперссылок с внешних по отношению к веб-пространству КарНЦ РАН сай-

тов. Ясно, что покупка гиперссылок на так называемых «линкофермах» [7] является неприемлемой.

Очевидна рекомендация для научных сотрудников КарНЦ РАН: при создании персональных страниц на профессиональных веб-ресурсах (типа MathNet, [www.mathnet.ru](http://www.mathnet.ru)) обязательно указывайте официальный сайт своего института или КарНЦ РАН. То же самое относится к регистрации на сайтах конференций.

Научные сотрудники КарНЦ РАН ведут большую преподавательскую деятельность в Петрозаводском государственном университете, Карельской государственной педагогической академии, других вузах России, и информация о них размещается на сайтах соответствующих кафедр. Однако ссылки на основное место работы, а тем более на сайт организации, в ряде случаев отсутствуют.

Исследования научного и университетского Веба России часто убеждают в том, что ссылки на коллег и партнеров не выставляются не по злому умыслу, а по причине отсутствия точной информации о соответствующих веб-ресурсах. Здесь можно долго развивать тему PR-технологий и социальных коммуникаций, но одна из самых простых рекомендаций заключается в том, что на визитной карточке научного сотрудника должен быть адрес официального сайта его института. И эти карточки надо активно раздавать коллегам из других научных организаций.

Работа выполнена при частичной поддержке Программы стратегического развития Петрозаводского государственного университета на 2012–2016 гг. и гранта РГНФ № 12-03-12001.

## ЛИТЕРАТУРА

1. Brin S., Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine // Seventh International World-Wide Web Conference (WWW 1998). April 14–18, 1998. Brisbane,

Australia. [Электронный ресурс] – Режим доступа: <http://ilpubs.stanford.edu:8090/361>.

2. Ortega J. L., Aguillo I. F. Visualization of the Nordic academic web: Link analysis using social network tools // Information Processing & Management. July 2008. Vol. 44, Iss. 4. P. 1624–1633.

3. Pant G., Srinivasan P., Menczer F. Crawling the Web // In Web Dynamics / M. Levene and A. Poulouvassilis, eds. Springer, 2004. P. 153–178.

4. Park H. W., Thelwall M. Hyperlink Analyses of the World Wide Web: A Review // Journal of Computer-Mediated Communication. 2003. N 8(4) [Электронный ресурс] – Режим доступа: <http://jcmc.indiana.edu/vol8/issue4/park.html?ref=ARKADASBUL.NET>.

5. Payne N., Thelwall V. A Statistical Analysis of UK Academic Web Links // Cybermetrics. International Journal of Scientometrics, Informetrics and Bibliometrics. 2004. Vol. 8, Iss. 1 [Электронный ресурс] – Режим доступа: <http://www.cindoc.csic.es/cybermetrics/articles/v8i1p2.html>.

6. Ranking Web of World Research Centers: CNR Institutes. RAS Institutes. <http://research.webometrics.info/RAS.asp> (дата обращения 18.04.2012).

7. Линкоферма – Википедия. <http://ru.wikipedia.org/wiki/Линкоферма> (дата обращения 18.05.2012).

8. Печников А. А., Луговая Н. Б., Чуйко Ю. В., Косинец И. Э. Разработка инструментов для вебметрических исследований гиперссылок научных сайтов // Вычислительные технологии. 2009. Т. 14, № 5. С. 66–78.

9. Печников А. А. Модель университетского Веба // Вестник Нижегородского университета им. Н. И. Лобачевского. 2010. № 6. С. 208–214.

10. Печников А. А., Чернобровкин Д. И. Адаптивный краулер для поиска и сбора внешних гиперссылок // Управление большими системами. Вып. 36. М.: ИПУ РАН, 2012. С. 301–315.

11. Яндекс. Как измеряется тИЦ. <http://help.yandex.ru/catalogue/?id=1111361> (дата обращения 14.05.2012).

## СВЕДЕНИЯ ОБ АВТОРЕ:

### Печников Андрей Анатольевич

вед. н. с. лаб. ТКС, к. ф.-м. н., д. т. н., доцент  
Институт прикладных математических исследований  
Карельского научного центра РАН  
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910  
эл. почта: [pechnikov@krc.karelia.ru](mailto:pechnikov@krc.karelia.ru)  
тел.: (8142) 763370

### Pechnikov, Andrey

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences  
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia  
e-mail: [pechnikov@krc.karelia.ru](mailto:pechnikov@krc.karelia.ru)  
tel.: (8142) 763370