

АНАЛИЗ РЕГИОНАЛЬНЫХ СТРУКТУР НАУЧНЫХ САЙТОВ РАН МЕТОДАМИ ВЕБОМЕТРИКИ (НА ПРИМЕРЕ АНАЛИЗА ИНТЕРНЕТ-ПРОСТРАНСТВА ДАЛЬНЕВОСТОЧНОГО ОТДЕЛЕНИЯ РАН)

К.А. Платонов

В 1997 году Tomas C. Almind и Peter Ingwersen в работе "Informetric analyses on the World Wide Web: Methodological approaches to «webometrics»"[8] ввели термин «webometrics» обозначающий раздел информатики, в котором исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к World Wide Web.

Научные сайты можно разделить на следующие группы: официальные сайты институтов, университетов и научных организаций, научных лабораторий, групп и отдельных ученых; научные электронные библиотеки, в том числе и сайты научных журналов; сайты конференций; центры хранения данных, научные информационные системы, ГИС-порталы; научные тематические ресурсы; научно-популярные ресурсы; каталоги научных ресурсов; другие сайты [2].

Доступ к научной информации глобальной сети Интернет для пользователей осуществляется, в основном, с помощью поисковых систем. Наличие высоких рейтинговых оценок для сайта очень важно с точки зрения его более высокой доступности широкой аудитории пользователей Интернет, поскольку рейтинговые оценки используют практически все поисковые системы. Наличие рейтинговых систем для сайтов, разрабатываемых и применяемых крупными поисковиками: Google, Yandex и др. при выдаче результатов поиска пользователям Интернет позволяет пользователям всего мира быстрее находить наиболее качественную и отвечающую запросу информацию. Результаты запросов пользователей сортируются поисковыми машинами и предоставляются пользователям в порядке уменьшения их рейтингов. Таким образом, наличие высоких рейтингов ставит сайт в лучшее положение по отношению к другим сайтам.

Самый известный проект в области вебометрики реализовала группа Кибернетической лаборатории Центра научной информации и документации Национального исследовательского совета Испании "Webometrics Ranking of World Universities"[9]. Используя методы вебометрики, лаборатория каждые полгода публикует мировой рейтинг сайтов университетов и научно-исследовательских институтов [10].

Российские работы в данном направлении проводились Институтом вычислительных технологий Сибирского отделения РАН и Институтом прикладных математических исследований КарНЦ РАН. За основу были приняты методы и подходы испанской группы, но с изменениями и дополнениями. Исследователями СО РАН создан ряд Интернет сервисов для вычисления рейтингов научных сайтов Сибирского отделения РАН, которые включают: выбор формулы, разбиение по направлениям исследований, выбор группы исследуемых сайтов [6].

Работы Института прикладных математических исследований КарНЦ РАН выполнены с использованием теории графов, которая применялась для анализа структуры Интернет-пространства научных организаций РАН Северо-запада России. На основе результатов авторами построена модель научного веба [1,3,4]. Предложенная модель позволяет увидеть административный каркас, тематические связи, отделить внутренние ресурсы сайта. Проведенные исследования позволили выявить особое множество веб-коммуникаторов: сайты-индукторы («мало входящих ссылок, много исходящих»), сайтов-коллекторов («много входящих, мало исходящих»), подмножество сайтов-посредников («много входящих, много исходящих»).

Работы по анализу Интернет-пространства Дальневосточного отделения РАН проводятся с 2009 года в Дальневосточном геологическом институте ДВО РАН [2,5].

Первый анализ информационного пространства Дальневосточного отделения РАН проведен В.В.Наумовой в 2009 году. Результаты анализа описаны в статье [5]. При проведении исследования за единицу анализа был принят каждый Интернет-ресурс ДВО РАН в отдельности. Хотя многие институты/учреждения поддерживают несколько различных доменных областей. Таким образом, их реальное присутствие в Сети на самом деле должно определяться множеством адресов. Кроме того, различные подразделения одной организации могут иметь собственные Интернет-ресурсы с адресами, не ассоциируемыми поисковыми машинами с адресом основного сайта. Если говорить об Интернет-ресурсах институтов РАН в целом, то они представляют собой сложный информационный комплекс. Как правило, в этом комплексе имеется официальный сайт института, а также независимые сайты лабораторий, научных журналов, электронных библиотек, страницы научных сотрудников, информационные системы, ГИС-порталы, сайты конференций, семинаров, тематические ресурсы и т.д. Следовательно, в анализе сайтов ДВО РАН 2009 г. не учтена суммарная Интернет активность институтов и организаций.

При проведении мониторинга Информационного пространства Дальневосточного отделения РАН в 2011 г. В.В.Наумовой изучена структура Информационного пространства ДВО РАН. Анализ основан на авторской классификации научных сайтов ДВО РАН, а также на визуальном анализе ссылок между сайтами.

1. В структуре Информационного пространства Дальневосточного отделения РАН, как и в любой другой структуре российского научного Интернет - пространства выделен административный каркас. В состав административного каркаса ДВО РАН входят: официальный сайт Президиума ДВО РАН <http://www.febras.ru> и официальные сайты институтов и организаций ДВО РАН. Визуальный анализ ссылок между этими сайтами отражает иерархическую соподчиненность этих организаций. При этом преобладают ссылки с сайт Президиума ДВО РАН на сайты институтов, обратных ссылок практически нет.

2. Научные подмножества сайтов ДВО РАН в основном формируются вокруг официальных сайтов институтов и представляют собой неофициальные сайты лабораторий, научных групп, личные страницы сотрудников и др.

3. Сайты научных ресурсов Дальневосточного отделения РАН также группируются около сайтов своих институтов и представляют собой:

а) Библиотечные ресурсы, в том числе сайты научных журналов и электронные библиотеки. В этой группе Интернет –ресурсов ДВО РАН необходимо отметить активно развивающийся сайт Центральной научной библиотеки ДВО РАН <http://www.cnb.dvo.ru>, который возможно в ближайшее время станет web-коммуникатором всех библиотечных ресурсов Отделения;

б) Центры хранения данных, научные информационные системы, ГИС-порталы;

в) Научные тематические ресурсы; научно-популярные ресурсы;

г) Архивы данных, в том числе и медийные архивы.

4. Множество web-коммуникаторов ДВО РАН, т.е. сайтов, имеющих входящие ссылки с и/или исходящие ссылки на множество других научных сайтов, можно разделить на группы:

а) Сайты конференций. В Дальневосточном отделении РАН сайты научных конференций пока еще в полной мере не являются web-коммуникаторами.

б) «Грантовые» коммуникаторы. В настоящее время в ДВО РАН есть один сайт из этой группы - это Информационная система «Грант ДВО РАН» <http://grant.febras.net>

в) «Сетевые» коммуникаторы. Исторически так сложилось, что одним из основных web-коммуникаторов ДВО РАН является Портал Научно-образовательной сети Владивостока и Сети ДВО РАН <http://dvo.ru>. Портал имеет наибольшее среди сайтов ДВО РАН количество входящих ссылок с и/или исходящих ссылок на множество других научных сайтов ДВО РАН.

Данная работа является логическим продолжением работ, ведущихся в этом направлении в лаборатории информационных технологий Дальневосточного геологического института ДВО РАН.

В данной работе поставлена следующая задача - разработка и создание Интернет сервиса по мониторингу научных ресурсов Сети, включающего в себя механизм автоматической интеграции сведений о научных сайтах, а также сервисов их обработки.

Автор, используя ресурсы <http://fegi.ru/institute/febras/47-febras>, учитывал в расчетах все Интернет сайты институтов.

Сбор данных осуществляется в автоматическом режиме с использованием API известных поисковых систем (Google, Bing, Yahoo и Яндекс). Данные включают:

количество уникальных гипертекстовых ссылок с других ресурсов (V – visibility, цитируемость);

Общее количество страниц сайта (S – size, размер);

Количество полнотекстовых файлов, под которыми понимаются файлы с расширениями pdf, ps, doc, xls, ppt и rtf (R – “rich files”);

Научность сайта (Sc – “shoolar”).

Далее рассчитывается ранг объекта по каждому индикатору. ($RankS$, $RankV$, $RankR$, $RankSc$), Вычисляется интегральный показатель, называемый “вебометрический ранг”(WR-Webometrics Rank), который получается в результате ранжирования суммы рангов, умноженных на коэффициенты [7]:

$Webometrics Rank(position) = 4*RankV + 2*RankS + 1*RankR + 1*RankSc$ (Таб.1)

Таб.1. Сравнительные результаты ранжирования сайтов институтов
Дальневосточного отделения РАН

№ в ДВО РАН	Результаты ранжирования ДВГИ (2012 г., май)	№ в ДВО РАН	Результаты ранжирования Кибернетической лаборатории (2012 г., февраль)
1	Дальневосточный геологический институт http://www.fegi.ru	1	Дальневосточный геологический институт http://www.fegi.ru
2	Биолого-почвенный институт http://www.biosoil.ru/	2	Биолого-почвенный институт http://www.biosoil.ru/
4	Институт автоматизации и процессов управления http://www.iacp.dvo.ru/	3	Тихоокеанский океанологический институт http://www.poi.dvo.ru/
3	Тихоокеанский океанологический институт http://www.poi.dvo.ru/	4	Институт автоматизации и процессов управления http://www.iacp.dvo.ru/
9	Институт биологии моря http://www.imb.dvo.ru/	5	Тихоокеанский институт географии http://tig.dvo.ru/
6	Институт морской геологии и геофизики http://www.imgg.ru/	6	Институт морской геологии и геофизики http://www.imgg.ru/
13	Институт космофизических исследований и распространения радиоволн http://www.ikir.ru/	7	Институт проблем морских технологий http://www.imtp.febras.ru/
8	Центральная научная библиотека http://www.lib.febras.ru/	8	Центральная научная библиотека http://www.lib.febras.ru/
5	Тихоокеанский институт географии http://tig.dvo.ru/	9	Институт биологии моря http://www.imb.dvo.ru/
12	Тихоокеанский институт биоорганической химии http://www.piboc.dvo.ru/	10	Институт прикладной математики http://www.iam.dvo.ru/
10	Институт прикладной математики http://www.iam.dvo.ru/	11	Институту истории, археологии и этнографии народов Дальнего Востока http://ihae.febras.ru/
6	Институт проблем морских технологий http://www.imtp.febras.ru/	12	Тихоокеанский институт биоорганической химии http://www.piboc.dvo.ru/
14	Горнотаежная станция им. В.Л.Комарова http://www.gts.febras.ru/	13	Институт космофизических исследований и распространения радиоволн http://www.ikir.ru/
11	Институту истории, археологии и этнографии народов Дальнего Востока http://ihae.febras.ru/	14	Горнотаежная станция им. В.Л.Комарова http://www.gts.febras.ru/

Для анализа структуры информационного пространства Дальневосточного отделения РАН использованы методы теории графов. Для построения взвешенного графа использован следующий параметр: количество исходящих и входящих ссылок с сайтов ДВО РАН на другие сайты ДВО РАН и Интернет. Выбор этого параметра обусловлен задачей поиска и анализа структуры связности этого пространства, (Рис.1)

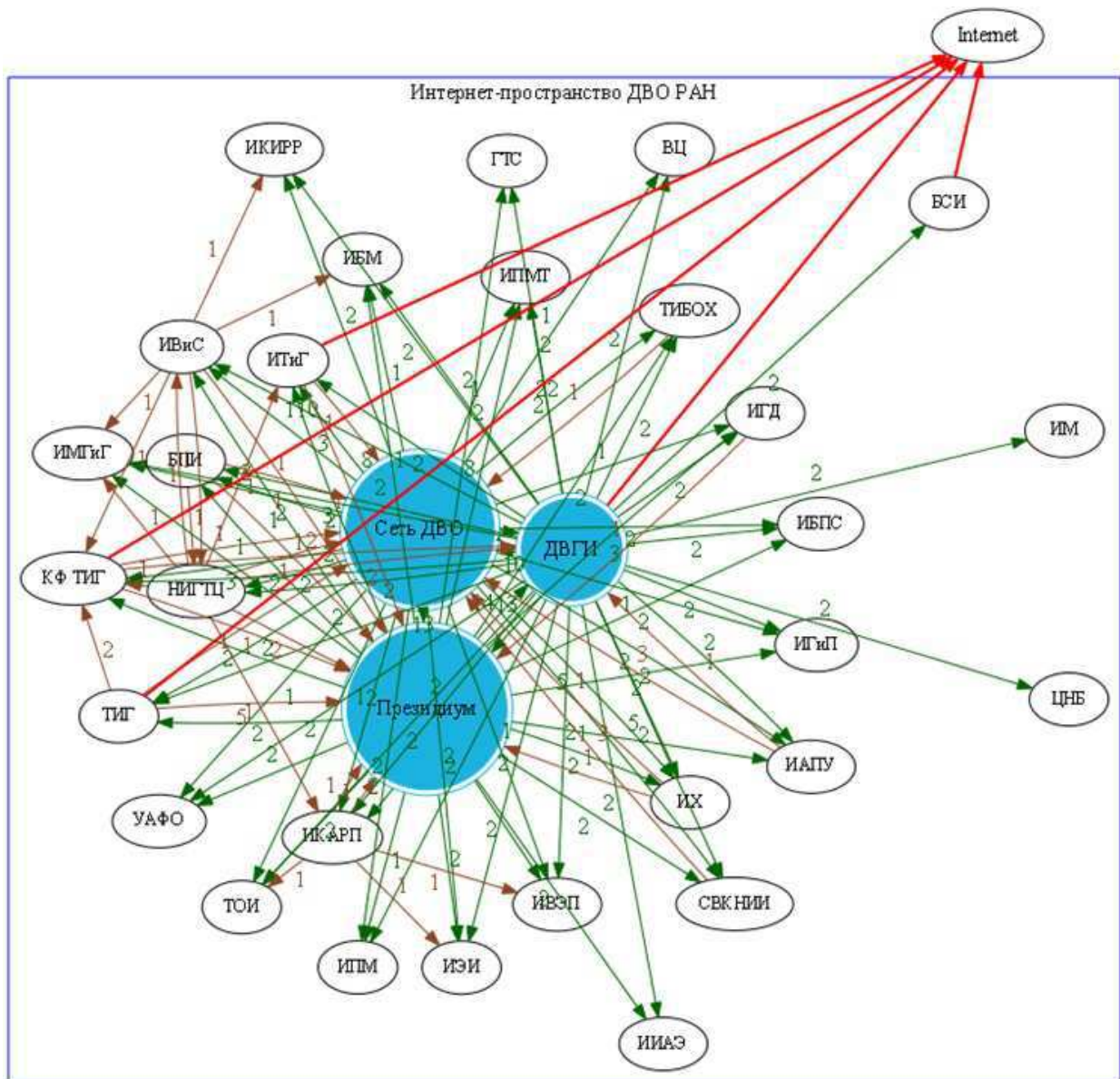


Рис 1. Информационное пространство Дальневосточного отделения РАН. При отображении структуры использованы методы теории графов. Зеленые стрелки – исходящие ссылки с веб-коммуникаторов; Красные стрелки – более 5 ссылок на внешние ресурсы; коричневые стрелки – остальные ссылки.

Примечание: Расшифровка аббревиатур в Приложении 1.

Приложение 1

Президиум – сайт Президиума ДВО РАН, <http://www.febras.ru>

Сеть ДВО – сайт Научно-образовательной сети ДВО РАН, <http://dvo.ru>

ДВГИ – Дальневосточный геологический институт ДВО РАН, <http://www.fegi.ru>

ИАПУ – Институт автоматики и процессов управления ДВО РАН, <http://www.iacp.dvo.ru>

ИБМ – Институт биологии моря ДВО РАН, <http://www.imb.dvo.ru>

ИПМ – Институт прикладной математики ДВО РАН, <http://www.iam.dvo.ru>

ИПМТ – Институт проблем морских технологий ДВО РАН, <http://www.imtp.febras.ru>

ТИГ – Тихоокеанский институт географии ДВО РАН, <http://tig.dvo.ru>

ТОИ – Тихоокеанский океанологический институт ДВО РАН, <http://www.poi.dvo.ru>

ИЭИ – Институт экономических исследований ДВО РАН, <http://www.ecrin.ru>

ИМГиГ – Институт морской геологии и геофизики ДВО РАН, <http://www.imgg.ru>

БПИ – Биолого-почвенный институт ДВО РАН, <http://www.biosoil.ru>

НИГТЦ – Научно-исследовательский геотехнологический центр ДВО РАН, <http://nigt.kscnet.ru>

ИВиС – Институт вулканологии и сейсмологии ДВО РАН, <http://www.kscnet.ru/ivs>

СВКНИИ – Северо-Восточный комплексный научно-исследовательский институт ДВО РАН, <http://www.neisri.ru>

ИБПС – Институт биологических проблем Севера ДВО РАН, <http://www.ibpn.ru>

ИГиП – Институт геологии и природопользования ДВО РАН, <http://www.ignm.ru>

ИГД – Институт горного дела ДВО РАН, <http://www.igd.khv.ru>

ИВЭП – Институт водных и экологических проблем ДВО РАН, <http://ivep.as.khb.ru>

ИМ – Институт материаловедения ДВО РАН, <http://www.im.khv.ru>

ИКАРП – Институт комплексного анализа региональных проблем ДВО РАН, <http://icarp.ru>

ИТиГ – Институт тектоники и геофизики им. Ю. А. Косыгина ДВО РАН, <http://itig.as.khb.ru>

ВЦ – Вычислительный центр г.Хабаровск ДВО РАН, <http://www.cfebras.ru>

УАФО – Уссурийская астрофизическая обсерватория ДВО РАН, <http://www.uafo.ru>

ИХ – Институт химии ДВО РАН, <http://www.ich.dvo.ru>

ИИАЭ – Институт истории, археологии и этнографии народов Дальнего Востока ДВО РАН, <http://www.ihae.org>

БСИ – Ботанический сад-институт ДВО РАН, <http://botsad.ru>

КФ ТИГ – Камчатский филиал Тихоокеанского института географии ДВО РАН, <http://www.terrakamchatka.org>

ГТС – Горнотаежная станция им. В.Л.Комарова ДВО РАН, <http://www.gts.febras.ru>

ЦТБ – Центральная научная библиотека ДВО РАН, <http://www.lib.febras.ru>

ИКИРР – Институт космофизических исследований и распространения радиоволн ДВО РАН, <http://www.ikir.ru>

ТИБОХ – Тихоокеанский институт биоорганической химии ДВО РАН, <http://www.piboc.dvo.ru>

Визуальный анализ построенного графа позволяет сделать вывод об относительной изолированности Интернет пространства Дальневосточного отделения РАН от ресурсов Интернет, включая другие ресурсы РАН. Подобная ситуация уже не раз встречалась при анализе научных интернет множеств. Количество исходящих ссылок из Интернет пространства ДВО РАН в Интернет ничтожно мало. Плохо прослеживается как административный каркас, так и тематические связи. Структура фактически замкнутая. Количество перекрестных ссылок внутри пространства ДВО РАН незначительно. Однако можно выделить три сайта: сайт Президиума ДВО РАН, сайт Научно-образовательной сети ДВО РАН и сайт Дальневосточного геологического института, которые несколько исправляют эту ситуацию. Они содержат значительное количество ссылок на ресурсы ДВО РАН. Их можно назвать сайтами-коммуникаторами Дальневосточного отделения РАН, следуя терминологии Печникова [4].

В результате выполненных автором работ:

1. Реализованы программные инструменты для сбора информации о научных сайтах в Интернет;
2. Реализован алгоритм ранжирования научных сайтов Дальневосточного отделения РАН в Интернет;
3. Начаты работы по анализу структур научных сайтов с использованием методов прикладной математики;
4. Реализована тестовая версия сайта <http://webometrics.fegi.ru>, содержащего ряд сервисов для сбора информации о научных сайтах, ранжирования научных сайтов и анализа структур научных сайтов.

При дальнейшем анализе научных сайтов с использованием современных методов анализа данных возникают следующие задачи:

- а) поиск признаков, которые наиболее влияют на рейтинг научного сайта в Интернет;
- б) анализ структуры научных сайтов тематических отделений РАН на примере сайтов Отделения наук о Земле РАН;

в) изучение взаимоотношения структур научных сайтов на примере взаимоотношений сайтов региональных отделений и научных центров РАН;

г) изучение взаимоотношений структур научных сайтов РАН и структур мировых научных сайтов в области наук о Земле;

д) и др.

Анализ региональных структур научных сайтов с использованием ссылок, а также других возможных параметров, позволит построить разумные стратегии продвижения научных сайтов и структур в Интернет.

ЛИТЕРАТУРА:

1. Мазалов В.В, Печников А.А. О рейтинге официальных сайтов научных учреждений северо-запада России //Управление большими системами. Выпуск 24. М.:ИПУ РАН, 2009. С.130-146.
2. Наумова В.В. Мониторинг Информационного пространства Дальневосточного отделения РАН//Вестник ДВО РАН,2012, №2, стр. 117-122
3. Печников А.А. Вебометрические исследования Web-сайтов университетов России//Информационные технологии,2008,№11
4. Печников А.А. Модель университетского Веба / Вестник Нижегородского университета им. Н.И. Лобачевского. №6. 2010. С. 208-214.
5. Ханчук А.И., Наумова В.В. Информационное пространство Дальневосточного отделения РАН//Вестник ДВО РАН,2009, №4, стр.122-130
6. 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
7. Aguillo, I.F.; Granadino, B.; Ortega, J.L.; Prieto, J.A. Scientific research activity and communication measure with cybermetrics indicators. Journal of the American Society for the Information Science and Technology,2006, 57(10):1296-1302.
8. Tomas C. Almind and Peter Ingwersen "Informetric analyses on the World Wide Web: Methodological approaches to «webometrics»". Journal of Documentation 53 (4): 404–426.
9. Ranking Web of World Research Centers [Электронный ресурс] – 2012. – Режим доступа: <http://research.webometrics.info>
10. Ranking Web of World Universities [Электронный ресурс] – 2012. – Режим доступа: <http://www.webometrics.info>