

УДК [378.1:004.738.1]:37.012.4

Клименко О. А., Бабушкин А. Г., Федоров Р. Ю.

**ВЕБОМЕТРИЧЕСКИЕ ПОДХОДЫ К ИЗУЧЕНИЮ
АКАДЕМИЧЕСКОГО ВЕБ-ПРОСТРАНСТВА: СОВРЕМЕННЫЙ
ОПЫТ И ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЙ**

Ольга Анатольевна Клименко

кандидат физико-математических наук

klimenko@ict.nsc.ru

ФГБУН Институт вычислительных технологий СО РАН,

Россия, Новосибирск

Алексей Георгиевич Бабушкин

кандидат технических наук

ab@gisi.ru

ФГБУН Институт криосферы Земли Тюменского научного центра СО РАН,

Россия, Тюмень

Роман Юрьевич Федоров

кандидат философских наук

r_fedorov@mail.ru

ФГБУН Институт криосферы Земли Тюменского научного центра СО РАН,

Россия, Тюмень

**WEBOMETRICS APPROACHES TO STUDYING OF THE ACADEMIC
WEB SPACE: MODERN EXPERIENCE AND PROSPECTS OF
RESEARCHES**

Olga Anatolievna Klimenko

Institute of Computational Technologies of SB RAS, Russia, Novosibirsk

Alexey Georgievich Babushkin

Earth Cryosphere Institute, Tyumen Scientific Centre SB RAS, Russia, Tyumen

Roman Yurievich Fedorov

Earth Cryosphere Institute, Tyumen Scientific Centre SB RAS, Russia, Tyumen

***Аннотация.** В статье предпринят обзор мирового и отечественного опыта изучения академического веб-пространства посредством методологического потенциала вебометрики. Приведены результаты исследований веб-пространства Сибирского отделения РАН, включающие построение рейтинга интернет-ресурсов научных организаций и моделирование их связей посредством теории графов. Рассмотрены перспективы инфометрического изучения межличностных коммуникаций в научных сетях и наукометрических информационных системах.*

***Abstract.** The article is devoted to review of international and Russian experience of studying of the academic web-space by means of the methodological potential of webometrics. The results of researches of web-space of Siberian Branch of the Russian Academy of Sciences, including creation of rating of Internet resources of the scientific institutes and modeling of their communications by means of the theory of graphs are given. The prospects of infometrics studying of interpersonal communications in scientific networks and scientometrics information systems are considered.*

***Ключевые слова:** вебометрика, веб-пространство, инфометрия, научные коммуникации, академические веб-ресурсы*

***Keywords:** webometrics, web-space, information, scientific communications, academic web-recourses.*

На протяжении последних десятилетий развитие информационных технологий породило феномен виртуализации научных коммуникаций. Сегодня одним из инструментов их изучения является вебометрика, которая часто рассматривается в качестве составной части инфометрии — научной дисциплины изучающей измерения количественных характеристик информации. По своим методологическим принципам, родственными вебометрике являются такие дисциплины, как наукометрия, библиометрия и киберметрия. Началом развития вебометрики в качестве самостоятельного направления исследований послужили работы Т. Алминда и П. Ингверсена [10]. Первым крупным проектом,

связанным с изучением вебметрических показателей научных интернет-ресурсов стал портал Webometrics Ranking of World Universities. На сегодняшний день он представляет собой систему рейтингов сайтов более чем 12 000 университетов мира, разработанную и поддерживаемую группой Cybermetrics Lab испанского Национального исследовательского совета (CSIC) [11]. В целях и задачах, сформулированных на сайте проекта, указано, что он призван способствовать развитию открытого доступа к научной и культурной информации, а также стимулировать общедоступность и прозрачность публичного представления сведений о научно-исследовательской и образовательной деятельности университетов. Помимо этого, первоначальные задачи проекта состояли в том, чтобы «убедить академические и политические сообщества в важности веб-публикации не только для распространения академического знания, но и для измерения активности научной деятельности и ее влияния» [12]. Для этих целей в рамках проекта были созданы рейтинги сайтов университетов в масштабах отдельных государств. Основными критериями для их формирования стали показатели Presence Rank (количество страниц сайта, найденных с помощью поисковой системы Google), Impact Rank (количество внешних ссылок на сайт), Openness Rank (число документов в формате pdf, размещенных на сайте) и Excellence Rank (статьи, которые входят в 10% наиболее цитируемых публикаций по данным Scimago) [2]. Таким образом, взятые за основу данные способны отразить уровень активности представления научно-образовательной деятельности университета в Интернете, а также степень ее востребованности у аудитории сайта.

В большинстве случаев можно отметить совпадение научно-образовательного авторитета университетов с позицией их сайтов в рейтинге. К примеру, по состоянию на январь 2017 г. в Великобритании три первых места в рейтинге заняли University of Oxford, University of Cambridge и University College London. В списке университетов Российской Федерации тройка ведущих мест распределилась между сайтами Московского государственного университета им. М. В. Ломоносова, Санкт-Петербургского государственного

университета и Национального исследовательского университета Высшей школы экономики. Несмотря на то, что рейтинг Webometrics Ranking of World приобрел авторитет в мировом масштабе, многие исследователи отмечают ряд недостатков его методологии. К примеру, на сайтах некоторых университетов может присутствовать большое количество информации, не имеющей прямого отношения к их научно-исследовательской деятельности. Также результаты рейтинга могут быть лишены объективности для тех учреждений, сайты внутренних подразделений которых поддерживают разные домены.

На территории России возник ряд специализированных вебометрических проектов в методиках которых были предприняты попытки учесть особенности структурной организации национального научного сообщества и отечественного Интернета. Остановимся подробнее на опыте создания Рейтинга сайтов научных учреждений Сибирского отделения РАН, разработанного в Институте вычислительных технологий СО РАН [9].

Первоначально рейтинг СО РАН строился по тем же параметрам, что и рейтинг Webometrics, затем был построен свой рейтинг, в котором с помощью разработанного программного обеспечения находились четыре параметра.

Параметр S — размер сайта (количество страниц, определяемое поисковыми системами). Для определения размера сайта S использовались поисковые системы Яндекс, Google и Yahoo, так как именно они наиболее полно индексируют русскоязычную часть Интернета.

Параметр V — видимость сайта — означает число внешних ссылок на ресурс, которое определяется поисковыми системами Яндекс, Google и Yahoo.

Параметр R — число «мощных» файлов — означает суммарное количество файлов форматов PDF, DOC и PPT, представленных на сайте.

Параметр Sc — индекс цитирования — определялся путем суммирования документов, в которых приводятся цитаты из докладов, статей и других научных материалов. Участники проекта Webometrics использовали сведения из системы Google Scholar. Авторами использовался также индекс цитирования Яндекса.

Было опробовано несколько весовых параметров. При этом выяснилось, что в результате их использования рейтинг меняется в средней части, причем несущественно. Лучшие и худшие сайты, как правило, занимали те же места. Исходя из этого, было принято правило, по которому считается среднее арифметическое всех параметров.

Рейтинг позволил выделить сайты организаций, которые лидируют по полноте представления сведений об их научно-исследовательской деятельности, а также по востребованности размещенной на них информации. Внедрение рейтинга послужило стимулом для развития ряда интернет-ресурсов научно-исследовательских институтов Сибири.

Другим направлением исследований веб-пространства Сибирского отделения РАН стало изучение связей между представленными в нем сайтами посредством теории графов. Американскими математиками в 2000 году была предложена модель описания веб-пространства в виде галстука-бабочки [11]. Узел галстука образует сильно связная компонента. Левая часть модели состоит из сайтов, гиперссылки из которых ведут в сильно связную компоненту. Правую часть модели образуют сайты, в которые ведут гиперссылки из сильно связной компоненты.

Полученные с помощью программы-краулера данные о связях между сайтами Сибирского отделения РАН показали, что для него представление в виде галстука-бабочки не подходит, потому что левая часть «бабочки» либо отсутствует, либо содержит единичные сайты. Поэтому, для визуализации связей в веб-пространстве СО РАН были опробованы другие модели (в виде звезды, шара, круга, круга с отростками или «медузы»). В модели «звезда» в центре находятся сайты с наибольшим числом входящих и исходящих гиперссылок. Они образуют ядро звезды, сайты менее всего связанные находятся на периферии (рисунок 1).

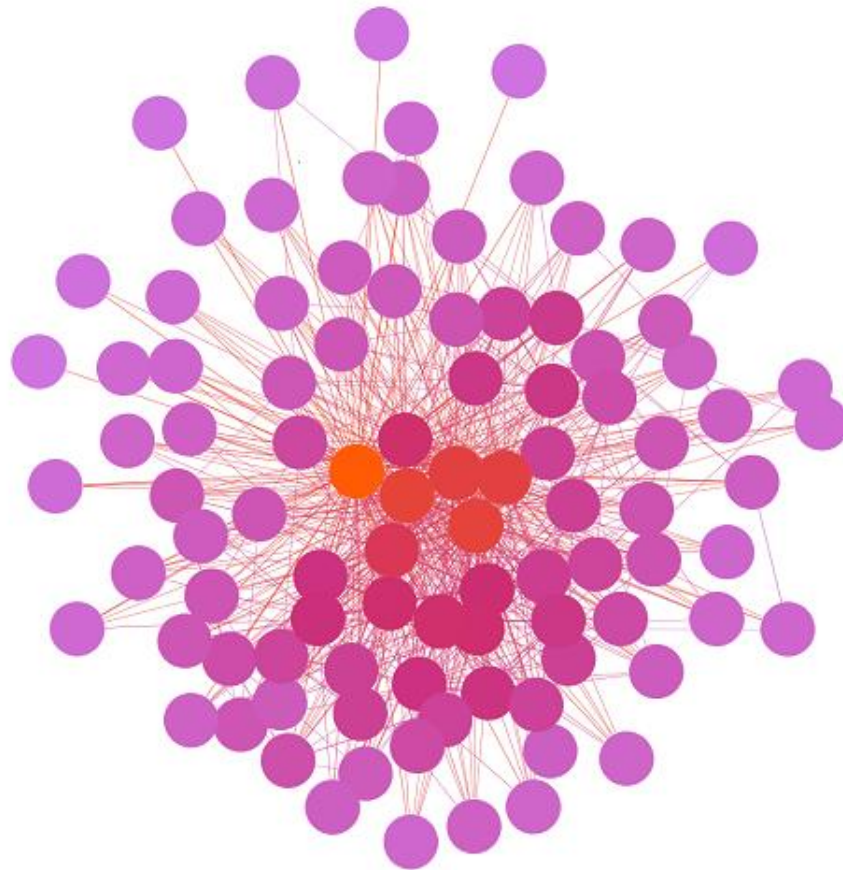


Рисунок 1 — Интернет-пространство Сибирского отделения РАН. Модель «звезда»

В ходе сравнительных исследований было установлено, что разные академические сообщества имеют схожую структуру в веб-пространстве. Ядро сообщества составляют сайты, которые образуют сильно связанную компоненту, то есть из любого сайта можно попасть в любой другой за несколько шагов. В сильно связанную компоненту может входить от 20% до 80% сайтов. В каждом сообществе есть 1–3 сайта с очень большим количеством входящих и исходящих ссылок. Большие сообщества имеют сильно связанное ядро, в котором из любого сайта можно по ссылкам попасть в любой другой сайт за 2–5 шагов. Существуют физико-математические, химико-биологические и др. сообщества, которые пересекаются. Гипотеза, что внутри больших академических сообществ есть кластеры, объединенные по географическому признаку, не подтвердилась [5]. Обычно, если у сайта большое число входящих ссылок, то он также имеет большое число исходящих ссылок. Исследования показали, что чем более развит сайт организации (чем больше на нем представлено полных текстов публикаций, проектов, материалов конференций, баз данных), тем

лучше сайт известен в научных кругах. Это объясняется тем, что на него больше ссылаются сотрудники других организаций, так как они имеют свободный доступ к результатам исследований [6].

Говоря о практической значимости изучения связей в академическом веб-пространстве, А. А. Печников отмечает, что «если рассматривать Веб как зеркало реальной жизни, то относительно научных взаимодействий можно сказать, что это зеркало достаточно хорошо отражает административные коммуникации РАН и не очень хорошо содержательные межинститутские» [8, с. 78]. Исходя из этого, на сегодняшний день, одной из перспективных прикладных задач изучения вебметрических связей может стать исследование трансформаций академического веб-пространства, связанных с реформой РАН. Наиболее заметное влияние на них может оказать присоединение к РАН учреждений, ранее входивших в ведомство Российской академии медицинских наук и Российской сельскохозяйственной академии, а также объединение отдельных профильных институтов и региональных научных центров в одно юридическое лицо. Опросы руководителей ряда реорганизованных институтов, вошедших в состав образованного в 2016 г. Федерального исследовательского центра «Тюменский научный центр СО РАН» свидетельствуют о том, что на начальном этапе объединения, большинство из них предпочло бы сохранить имевшиеся ранее интернет-ресурсы, включая их доменные имена, представив их в виде обособленных структурных подразделений нового объединенного научного центра. Во многом это связано с опасениями потери «узнаваемости» сложившихся ранее интернет-ресурсов научно-исследовательских коллективов, которая может произойти при смене их домена и изменения их устоявшейся содержательной структуры. Однако, как показывает практика развития сайтов большинства российских образовательных учреждений, прошедших в последние годы процессы объединений и укрупнений, в их деятельности возникает необходимость создания общих интернет-порталов, обеспечивающих ряд базовых функций, связанных с представлением сведений

о деятельности государственного учреждения. Подобная ситуация, в обозримом будущем может возникнуть и среди реорганизованных академических институтов. Поэтому в данный момент большую актуальность имеет отслеживание изменений активности и связности интернет-ресурсов всех научных организаций, находящихся в ведении ФАНО.

Если вебометрика интернет-ресурсов научных учреждений в первую очередь дает возможность оценить полноту представления их деятельности в Интернете, а также их ведомственные связи, то изучение информационных взаимодействий в конкретных научных сообществах оказывается более продуктивным при рассмотрении виртуальных коммуникаций отдельно взятых ученых. Задолго до массовой «виртуализации» научно-исследовательской деятельности, сети формальных и неформальных научных коммуникаций стали объектом исследования в социологии [1], [3], [4] и др. В настоящее время научные сети, а также наукометрические и библиометрические системы, действующие на веб-платформах, могут выступать в качестве эмпирической основы для изучения информационных взаимодействий в научных сообществах. Для этих целей, в подобных исследованиях, представляется наиболее эффективным комбинированное использование методологического потенциала инфометрии, включая не только вебометрику, но и библиометрию и наукометрию. Некоторые научные сети имеют встроенные инструменты анализа связей в рамках сформировавшихся в них научных сообществ. К примеру, в возможности научной сети researchgate.net входит построение графов прямых коммуникаций пользователя, коммуникаций пользователя по мощности сети у присоединенных пользователей, граф коммуникаций пользователя по странам мира, граф коммуникаций через тематические группы общих интересов и др [7]. В России большой интерес для аналитики представляют данные, отражающие публикационную активность ученых на портале elibrary.ru. На их основе возможно построение графов, позволяющих выявить неформальные научные сообщества на основе анализа соавторства и взаимного цитирования,

проследить динамику интеграции академической и вузовской науки посредством выявления двойных аффилиций (научно-исследовательского института и вуза) в публикациях авторов и т. д. Результаты подобных исследований могут иметь широкий спектр практического применения. К примеру, на основе специально разработанных алгоритмов анализа неформальных научных сообществ могут быть сформулированы практические рекомендации, направленные на выявление ведущих научных школ и формирование эффективных исследовательских коллективов. В то же время, подобные виды мониторинга могут быть применены для выявления недобросовестных научных сообществ, систематически осуществляющих фальсификацию наукометрических показателей, а также использующих некорректные заимствования и т. д.

К сожалению, на сегодняшний день, возможности инфометрических исследований коммуникаций в научных сетях и системах цитирования остаются затруднительными по нескольким причинам. Первая из них состоит в ограниченности возможностей стандартных аналитических сервисов, доступных пользователям этих ресурсов. Во многих случаях они могут предоставить статистику о деятельности конкретного пользователя, однако не дают представления о глобальных коммуникативных взаимодействиях всех пользователей сети. Вторая проблема состоит в том, что в отличие от интернет-ресурсов, количественные показатели, взаимодействия и связи в научных сетях и системах цитирования, в большинстве случаев невозможно или затруднительно получить с помощью внешних инструментов (например, краулеров), из-за приватности их данных. Еще одной проблемой является тот факт, что при выявлении связанных структур и компонентов как правило используются инструменты, работающие в основном со статическим контентом и значительная часть динамического контента, который постоянно растет, остается не охваченной анализом. Решение данных проблем возможно путем реализации новых подходов к изучению межличностных научных коммуникаций в Интернете. Первый из них состоит в разработке специализированных аналитических платформ на базе крупнейших отечественных наукометрических систем типа elibrary.ru в

первую очередь предназначенных для аналитического обеспечения работы специалистов, осуществляющих организационное обеспечение научных исследований. Второй подход заключается в разработке специализированных средств для внешнего мониторинга определенных видов открытой информации, размещаемой в научных сетях, аналогичных по принципам работы инструментальным средствам SMM (Social Network Monitoring), ранее доказавшим свою эффективность преимущественно в сфере маркетинговых исследований социальных сетей.

Список литературы

1. Анализ современной научной коммуникации: лингвистические и образовательные технологии / под ред. Н. Л. Шубиной. – Санкт-Петербург : Книжный дом, 2010. – 43 с.
2. Бойков, П. А. Анализ факторов, влияющих на продвижение ТПУ в рейтинге Webometrics / П. А. Бойков, И. А. Середов, Ю. В. Бабушкин // Информационные технологии в науке, управлении, социальной сфере и медицине : сборник научных трудов III Международной научной конференции, 23–26 мая 2016 г., Томск. – Томск : Издательство ТПУ, 2016. – Ч. 1. – С. 447–450.
3. Дюментон, Г. Г. Исследование сети научных контактов и его практическое значение / Г. Г. Дюментон // Социологические проблемы науки. – Москва : Наука, 1974. – С. 348–368
4. Дюментон, Г. Г. Сети научных коммуникаций и организация фундаментальных исследований / Г. Г. Дюментон. – Москва : Наука, 1987. – 103 с.
5. Клименко, О. А. Алгоритмы и методы нахождения сообществ на примере ВЕБ-пространства Сибирского Отделения РАН и общества Фраунгофера в Германии / О. А. Клименко, А. Н. Медведев // Математическое моделирование и вычислительно-информационные технологии в меж-дисциплинарных научных исследованиях : материалы IV Всероссийской конференции. – Иркутск : Институт динамики и теории управления СО РАН, 2014. – С. 34–34.

6. Клименко, О. А. Модели представления академического веб-пространства / О. А. Клименко // Информационные и математические технологии в науке и управлении. – 2016. – № 2. – С. 103–110.
7. Мартышенко, Н. С. Использование коммуникаций в сети Интернет в научных исследованиях / Н. С. Мартышенко // Теория и практика сервиса: экономика, социальная сфера, технологии. – 2010. – № 4 (6). – С. 108–119.
8. Разработка инструментов для вебметрических исследований гиперссылок научных сайтов / А. А. Печников, Н. Б. Луговая, Ю. В. Чуйко, И. Э. Косинец // Вычислительные технологии. – 2009. – Т. 14. – № 5. – С. 697–698.
9. Рейтинг сайтов научных учреждений СО РАН [Электронный ресурс]. – Режим доступа: <http://w.ict.nsc.ru/ranking/> (дата обращения: 8.12.2018).
10. Almind, T. Infometric analyses on the World Wide Web: methodological approaches to ‘webometrics’ / T. Almind, P. Ingwersen // Journal of Documentation. – 1997. – V. 53. – № 4. – P. 404–426.
11. Graph structure in the Web / A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener // Computer Networks. – 2000. – Vol. 33. – № 1–6. – P. 309–320.
12. Ranking Web of Universities [Electronic resource]. – Access mode: <http://www.webometrics.info>.
13. Ranking Web of Universities. Methodology [Electronic resource]. – Access mode: <http://www.webometrics.info/en/Methodology>.