

На правах рукописи

Печников Андрей Анатольевич

РАЗРАБОТКА МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ, МЕТОДОВ И
ПРОГРАММНЫХ СРЕДСТВ ДЛЯ ИССЛЕДОВАНИЯ ВЗАИМОСВЯЗЕЙ
РЕГЛАМЕНТИРУЕМЫХ ВЕБ-САЙТОВ

Специальность 05.13.18 – Математическое моделирование,
численные методы и комплексы программ

Автореферат
диссертации на соискание ученой степени
доктора технических наук

Петрозаводск – 2011

Работа выполнена в Институте прикладных математических исследований Карельского научного центра РАН

Официальные оппоненты:

доктор физико-математических наук,
профессор Елизаров Александр Михайлович

доктор технических наук,
профессор Кулагин Владимир Петрович

доктор технических наук,
профессор Путилов Владимир Александрович

Ведущая организация:

Санкт-Петербургский институт информатики и автоматизации Российской академии наук

Защита состоится «17» июня 2011 г. в 13:00 часов на заседании диссертационного совета Д 212.190.03 при Петрозаводском государственном университете по адресу: 185910, г. Петрозаводск, пр. Ленина, д.33.

С диссертацией можно ознакомиться в научной библиотеке Петрозаводского государственного университета.

Автореферат разослан «___» _____ 2011 г.

Ученый секретарь диссертационного совета

к.т.н., доцент

Р.В. Воронов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. *World Wide Web, WWW, W3, Web, Сеть, Всемирная паутина, Веб* – все эти термины используются для обозначения одного явления глобального масштаба. Определим *Веб* как распределенную систему взаимосвязанных гипертекстовых документов, содержащихся на различных компьютерах, подключенных к Интернету. Термин «*вебометрика*» (*webometrics*) обозначает научное направление, в рамках которого исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к Вебу (*T. Almind, P. Ingwersen Informetric analyses on the World Wide Web: Methodological approaches to «webometrics» // Journal of Documentation. – 1997. – № 53 (4). – P. 404-426.*

Сегодня официальные веб-сайты являются средствами для публикации открытой и доступной пользователям Веба информации и выполняют коммуникационные функции между организациями-владельцами и людьми, что полностью соответствует Берлинской декларации об открытом доступе к научному и гуманитарному знанию (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>). Сказанное относится и к веб-ресурсам органов государственной власти, создание которых направлено на повышение открытости их деятельности и оказание государственных услуг в электронном виде (*План реализации Стратегии развития информационного общества в Российской Федерации до 2011*, <http://www.infosovet.ru/actionplan>). Такой взгляд на Веб требует разработки новых подходов, направленных на создание благоприятной информационной среды, способствующей доступу пользователей Веба к открытой информации.

Одним из основных вопросов вебометрики является вопрос об оценке присутствия веб-ресурсов в Вебе. Для единичного веб-сайта речь идет о таких индикаторах, как, например, количество страниц («*объем*») и количество гиперссылок, сделанных на него с других сайтов («*узнаваемость*»). Для сообщества сайтов одной тематики можно говорить о таких характеристиках, как возможность перехода по гиперссылкам с одного сайта на другой («*связность*») и количество таких переходов («*расстояние*»). Рост присутствия веб-сайтов и их сообществ влияет на их поисковые характеристики, облегчая пользователям нахождение требуемой информации. Исследования, оценивающие присутствие веб-ресурсов университетов и научных организаций, показывают существенное отставание российских сайтов от зарубежных аналогов. По данным испанской группы *Cybermetrics Lab* на январь 2010 года, сайт Московского госуниверситета занимал 226 место в рейтинге университетов, а портал

РАН – 114-е в рейтинге научных организаций в мире (www.webometrics.info). Скромные позиции российских сайтов объясняются такими причинами, как малое количество англоязычных веб-страниц, низкая активность ученых и преподавателей в формировании контента веб-ресурсов и недостаток внешних гиперссылок, подчеркивающий невысокую коммуникабельность в Вебе.

Веб-ресурсы органов государственной власти, научных учреждений и вузов относятся к так называемым «*регламентируемым веб-ресурсам*». Это означает, что существует официальный документ, в котором изложены цели и задачи веб-ресурса, основные структурные составляющие, правила изменения информации и т.д. Под управляемостью процессами в Вебе понимается реализация управленческих решений в виде пунктов регламента, определяющих их исполнение. В этом случае для единичных сайтов и их сообществ могут быть выработаны управленческие решения, направленные на улучшение их присутствия в Вебе. Однако прежде чем реализовывать управленческие решения через регламенты, необходимо пройти путь от сбора информации о присутствии и взаимосвязи веб-сайтов в Вебе до построения формальных моделей, формулировки и решения математических задач и разработки рекомендаций, сделанных на их основе.

Учитывая размерность Веба, в качестве объектов исследований принимаются фрагменты Веба, задаваемые перечислением веб-сайтов заданной тематики (т.н. «*тематические целевые множества*» или «*тематические сообщества*»). Такой подход имеет ряд недостатков, в частности, из поля зрения выпадают веб-ресурсы, не вошедшие в целевое множество, но оказывающие существенное влияние на присутствие сайтов в Вебе. Для реализации целенаправленного процесса построения расширенного объекта исследований необходимо располагать достаточно большим множеством веб-ресурсов, сформированным независимо от предпочтений исследователя, которые потенциально могут воздействовать на целевое множество посредством гиперссылок, и иметь автоматизированные процедуры отбора из него допустимых (в смысле некоторых формальных критериев) веб-ресурсов с отсеком заведомо неприемлемых. Однако на сегодняшний день не разработаны подходы к формализации понятия «фрагмент Веба», не выявлено различий между важными и второстепенными веб-ресурсами. Это делает актуальной разработку концептуальной модели, отражающей общие характеристики фрагмента Веба, описывающей его разбиение на составляющие подмножества и их свойства, и определяющей связи между этими подмножествами.

Анализ концептуальной модели позволяет перейти к использованию теоретико-графовых подходов для исследования наиболее важных

компонент, причем размерность такой задачи заведомо меньше, чем размерность задачи исследования фрагмента Веба в целом. Исследования зарубежного Веба показывают, что, несмотря на видимое отсутствие непосредственных управляющих воздействий на процессы взаимосвязи сайтов посредством гиперссылок, в тематических сообществах возникают компоненты сильной связности большой мощности и малого диаметра. Такие исследования для российских фрагментов Веба на сегодня практически не проводились. Отсутствуют исследования, посвященные влиянию на связность тематических сообществ веб-сайтов, выполняющих функции посредников, которые не входят в заданное тематическое сообщество, но имеют большое количество гиперссылок, связывающих их с сайтами сообщества. Исследования связности сообществ и анализ влияния сайтов-посредников на их связность являются актуальными задачами, напрямую влияющими на присутствие тематических сообществ.

Управляемость процессами в Вебе подразумевает применение количественных математических методов для обоснования решений о целенаправленной деятельности. Решение таких задач дает принципиальную возможность создания механизмов самоорганизации фрагментов Веба. И хотя спектр таких задач применительно к Вебу представляется чрезвычайно широким, на сегодня эта тема является мало исследованной. Наиболее актуальными представляются постановки математических задач, решение которых существенно влияет на присутствие в Вебе сайтов и тематических сообществ.

Анализ открытых источников, посвященным программным средствам вебометрики, не позволил обнаружить некоммерческие российские разработки на эту тему. Необходима разработка программных средств для изучения Веба, применимых к российским веб-ресурсам и открытым для использования российскими исследователями.

Таким образом, актуальна проблема разработки общих подходов, методов и программных средств для исследования Веба, позволяющих строить формальные модели и ставить математические задачи, исследование и решение которых направлено на улучшение присутствия в Вебе российских сайтов и тематических сообществ. Предложенные в диссертационной работе решения поставленной проблемы опираются на результаты исследований университетских гиперссылок (*M. Thehwall, Statistical Cybermetrics Research Group, University of Wolverhampton*), теоретико-множественных и теоретико-графовых моделей Веба (*A-L. Barabasi, L. Vjörneborn, A. Broder*), ранжирования сайтов (*I.F. Aguillo, Cybermetrics Lab; Ю.И. Шокин, ИВТ СО РАН*).

Основные объекты исследований. Целевым множеством называется задаваемое прямым перечислением множество

регламентируемых тематических сайтов. *Сопутствующим множеством* называется множество сайтов, не входящих в целевое множество, на которые существуют гиперссылки с сайтов целевого множества. *Фрагмент Веба* – это объединение сайтов целевого и сопутствующего множеств и множество связывающих их гиперссылок. Основным объектом исследований в диссертационной работе является регламентируемый тематический фрагмент Веба. В качестве объектов для апробации математических моделей, методов и программ использовались академический фрагмент Веба (целевое множество – официальные сайты научных учреждений РАН), университетский фрагмент Веба (целевое множество – официальные сайты классических университетов Российской) и фрагмент Веба региональных органов государственной власти (целевое множество – сайты органов государственной власти Республики Карелия (РК)).

Цель работы заключается в разработке математических моделей, методов и программных средств для исследования взаимосвязей российских регламентируемых тематических веб-сайтов, направленных на улучшение их присутствия в Вебе.

Для достижения указанной цели **решается ряд задач:**

1. разработка метода построения концептуальной модели фрагмента Веба для регламентируемого тематического целевого множества, обеспечивающего автоматизированную реализацию моделей для целевых множеств различной тематики;
2. применение теоретико-графовых моделей и методов для исследования характеристик веб-графа, построенного на основе концептуальной модели фрагмента Веба, позволяющее дать оценки связности и определить основные направления для постановки задач улучшения присутствия в Вебе тематических сообществ;
3. исследование характеристик присутствия единичных сайтов с использованием известных алгоритмов ранжирования;
4. постановка, исследование и решение задач, улучшающих характеристики присутствия веб-сайтов и их сообществ в Вебе;
5. разработка структурных и функциональных требований, положенных в основу программного обеспечения для сбора, хранения и обработки вебометрической информации и его программная реализация.

Методы исследования. Для решения поставленных в работе задач используются методы концептуального моделирования, элементы теории множеств, теории графов, методов оптимизации, теории вероятностей и математической статистики, имитационное моделирование и численные

эксперименты на ЭВМ и методы проектирования программного обеспечения и баз данных.

Практической основой исследований являются:

1. созданные в процессе исследований вебометрические базы данных академического фрагмента Веба, университетского фрагмента Веба и фрагмента Веба региональных органов государственной власти РК;
2. массив данных о гиперссылках российского Веба, предоставленный Компанией Яндекс.

Научная новизна выполненных исследований заключается в разработке и развитии теоретических и практических разделов вебометрики, направленных на улучшение присутствия в Вебе российских регламентируемых тематических веб-ресурсов. Основные аспекты научной новизны:

1. Концептуальная модель фрагмента Веба для регламентируемого тематического целевого множества, представленная в виде теоретико-множественных отношений. Формальный характер модели обеспечивает возможность определения в её терминах формализованной процедуры построения концептуальной модели фрагмента Веба для заданного регламентируемого тематического целевого множества.
2. Метод автоматизированного построения реализации концептуальной модели реального фрагмента Веба, позволяющий дать численные оценки для каждого из составляющих подмножеств и степень их связности, классификацию входящих в них сайтов, а также существенно уменьшить размерность самих моделей посредством исключения подмножеств, сайты которых бесперспективны с точки зрения влияния на связность фрагмента.
3. Алгоритм ранжирования официальных веб-сайтов классических университетов России, основанный на содержательной интерпретации характеристик присутствия сайтов в Вебе.
4. Теоретико-графовые характеристики веб-графа для реальных фрагментов российского Веба, показывающие незначительную взаимосвязь веб-сайтов российских тематических сообществ.
5. Математические модели взаимосвязей сайтов посредством гиперссылок, применение которых позволяет улучшить характеристики присутствия веб-сайтов и их сообществ в Вебе.

Практическая значимость:

- Предложенные методы и реализованные на программном уровне алгоритмы позволяют строить научно обоснованные модели фрагментов Веба, которые могут быть использованы при разработке сценариев развития

комплексов веб-ресурсов организационных систем, направленных на улучшение их присутствия в Вебе.

- Результаты исследований внедрены для комплекса веб-ресурсов Карельского научного центра РАН в рамках проекта «Исследование и разработка научно-методических основ создания и развития инфраструктуры информационно-телекоммуникационной системы КарНЦ РАН».
- Результаты работы получили практическое применение в рамках исследований официальных сайтов органов государственной власти РК, что подтверждено соответствующим актом.
- Результаты исследований включены в Отчет о деятельности РАН в 2006 году (том 3, Основные исследования и разработки научных учреждений РАН, готовые к практическому применению).
- По результатам исследований прочитан обязательный спецкурс «Технологии специализированных баз данных для вебметрических исследований» для магистрантов факультета прикладной математики - процессов управления Санкт-Петербургского государственного университета (2009/2010 уч.год).
- Получено Свидетельство о государственной регистрации программы для ЭВМ «Информационная система для вебметрических исследований» Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ № 2010610941 от 29 января 2010 г.
- Широкая применимость предложенных моделей, методов и программного обеспечения подтверждается результатами апробации для тематических фрагментов Веба различной тематики.

Достоверность и обоснованность результатов, полученных в диссертации, основаны на большом объеме собранного, обработанного и исследованного фактического материала, применении системного подхода в исследованиях, корректности принятых при разработке моделей допущений. Разработанные и используемые методы и теоретические подходы находят свое подтверждение посредством их сравнения на различных фрагментах российского Веба и проверкой на множестве данных, предоставленных Компанией Яндекс. Адекватность ряда математических моделей подтверждается результатами компьютерного моделирования.

Личный вклад автора в проведенное исследование. В диссертацию включены результаты, которые получены лично автором.

Связь работы с научными темами и грантами.

В основу диссертации положены результаты, полученные автором в ходе исследований, проводимых по планам научно-исследовательских

работ Института прикладных математических исследований КарНЦ РАН «Исследование и разработка научно-методических основ создания и развития инфраструктуры информационно-телекоммуникационной системы КарНЦ» (тема НИР № 56, 2008-2010 гг.).

Исследования поддержаны:

- Компанией Яндекс (конкурс Интернет-математика 2007, проект «Математические модели согласованного поведения малых Интернет-сообществ»),
- инициативным грантом РФФИ (проект «Вебометрические исследования научных интернет-ресурсов российского Интернета», №08-07-00023а, 2008-2010 гг.),
- Программой фундаментальных исследований ОМН РАН «Алгебраические и комбинаторные методы математической кибернетики и информационные системы нового поколения» (проект «Задачи оптимальной маршрутизации трафика, распределения и защиты информационных ресурсов», 2009, 2010 гг.).

Апробация работы. Результаты исследований докладывались на 22 конференциях, научных школах, семинарах и совещаниях:

- I и V научно-методическая конференция «Университеты в образовательном пространстве региона: опыт, традиции и инновации», Петрозаводск, 1999, 2010 гг.
- Всероссийская научная школа «Математические методы в экологии», Петрозаводск, 2001 г.
- Международная конференция «Северная Европа в XXI веке: природа, культура, экономика», Петрозаводск, 2006 г.
- IX и XI Всероссийская научная конференция «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Переславль-Залесский, 2007 г., Петрозаводск, 2009 г.
- Седьмая международная Петрозаводская конференция «Вероятностные методы в дискретной математике», Петрозаводск, 2008 г.
- XV, XVI и XVII Всероссийская научно-методическая конференция «Телематика», Санкт-Петербург, 2008, 2009 и 2010 гг.
- Третья Всероссийская школа молодых ученых «Математические методы в экологии», Петрозаводск, 2008 г.
- II, III и IV Всероссийская научно-практическая конференция «Информационная среда вуза XXI века» Петрозаводск, 2008, 2009, 2010 гг.
- Научный семинар Хельсинкского института информационных технологий (*Helsinki Institute for Information Technology*), Хельсинки, 2009 г.
- Международная научно-практическая конференция «Теория активных систем - 2009», Москва, 2009 г.

- VIII Международная конференция «Когнитивный анализ и управление развитием ситуаций», Москва, 2009 г.
- IV Международная научно-практическая конференция "Современные информационные технологии и ИТ-технологии", Москва, 2009 г.
- Семинар по теории управления организационными системами Института проблем управления им. В.А. Трапезникова РАН, 2009 г.
- VIII Всероссийская школа-семинар «Прикладные проблемы управления макросистемами», Апатиты, 2010 г.
- Семинар по информационному поиску факультета прикладной математики-процессов управления СПбГУ, 2010 г.
- Научный доклад на заседании Президиума Карельского научного центра РАН, 29 июня 2010 г.

Основные результаты, выносимые на защиту:

1. Концептуальная модель фрагмента Веба и основанные на ней метод и автоматизированная процедура построения модели фрагмента Веба для заданного регламентируемого тематического целевого множества.
2. Теоретико-графовая модель взаимосвязей веб-сайтов посредством гиперссылок, дающая оценки влияния на связность целевого множества иерархической подчиненности организаций-владельцев сайтов и веб-сайтов организаций, выполняющих коммуникационные функции для его участников.
3. Задачи взаимосвязей веб-сайтов посредством гиперссылок, поставленные и исследованные с использованием методов дискретной оптимизации, кооперативной теории игр и имитационного моделирования, решение которых позволяет улучшить характеристики присутствия веб-сайтов и их сообществ в Вебе.
4. Алгоритм ранжирования официальных веб-сайтов классических университетов России, основанный на содержательной интерпретации характеристик присутствия сайтов в Вебе, требующий вдвое меньше исходных данных по сравнению с известными алгоритмами.
5. Рекомендации для организаций-владельцев официальных сайтов, использование которых в административных регламентах позволяет улучшить присутствие их веб-ресурсов в Вебе. Реализация рекомендаций для сайтов органов государственной власти Республики Карелия обеспечила прирост тематического индекса цитирования от 10 до 15% в течение года.
6. Проблемно-ориентированная информационная система, включающая в себя робот-сборщик информации о сайтах российского Веба, и реляционную базу данных для хранения и обработки собираемой

Публикации. Результаты диссертации опубликованы в 36 печатных работах, из которых 11 – статьи в журналах и сборниках, входящих в Перечень изданий, рекомендованных ВАК РФ для опубликования результатов докторских диссертаций. Разработан сайт рабочей группы по вебметрике ИПМИ КарНЦ РАН. Получено Свидетельство о государственной регистрации программы для ЭВМ Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ.

Структура и объем работы. Диссертация состоит из введения, шести глав, заключения, приложений и списка использованных источников, включающего 198 наименований. Общий объем диссертации 304 страницы, из них основная часть работы содержит 266 страниц и включает 43 рисунка и 66 таблиц; 9 приложений имеют общий объем 15 страниц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** дан анализ исследуемой проблемы, её специфика и современное состояние, подходы, применяемые для её решения. Обосновывается актуальность темы диссертации, определены цели и задачи исследования, охарактеризована научная новизна и практическая значимость результатов, приводятся данные о внедрении и апробации результатов исследований.

В **первой главе** дается аналитический обзор современного состояния вебметрических исследований, позволивший изложить основные понятия предметной области, определить основные объекты исследования и используемые подходы и методы.

Во **второй главе** представлена формализованная концептуальная модель для регламентируемого тематического целевого множества веб-сайтов и основанные на ней метод автоматизированного построения фрагментов Веба на основе задаваемого целевого множества и оценки её реализации.

Веб-сайт (сайт) – совокупность *html*-страниц и веб-документов, связанных внутренними гиперссылками и обладающих единством содержания, идентифицируемый в Вебе по уникальному доменному имени.

Внешняя гиперссылка с сайта *s* на сайт *t* формально представляет собой четверку:

<адрес исходной страницы на сайте *s*, уровень исходной страницы, контекст гиперссылки, адрес целевой страницы на сайте *t*>.

Уровень страницы определяется следующим образом: начальная страница сайта имеет уровень 0, а уровень любой другой страницы сайта –

это минимальное количество внутренних гиперссылок, ведущих от начальной страницы к данной. Под *контекстом гиперссылки* понимается текст, сопровождающий её на странице.

Уникальной внешней гиперссылкой называется гиперссылка из множества гиперссылок, имеющих одинаковый контекст и адрес целевой страницы, сделанная с исходной страницы с наивысшим уровнем. Далее рассматриваются именно такие гиперссылки, поэтому как аналогичные будут использоваться термины «*гиперссылка*» или «*ссылка*». Обозначим множество всех гиперссылок, сделанных с сайта s на сайт t , как $\{<s,t>\}$.

Целевым множеством называется множество сайтов, задаваемое перечислением доменных имен, а *сопутствующим множеством* (по отношению к заданному целевому множеству) называется множество сайтов, не входящих в целевое множество, на которые существуют гиперссылки с сайтов целевого множества. Введем следующие обозначения: $T = \{t_1, t_2, \dots, t_n\}$ – целевое множество, $U = \{u_1, u_2, \dots, u_m\}$ – сопутствующее множество, $F = T \cup U$ – объединение целевого и сопутствующего множеств, E – множество гиперссылок, соединяющих сайты из F : $E = \bigcup_{\substack{\forall s \in F \\ \forall t \in F, t \neq s}} \{<s,t>\}$.

Фрагмент Веба – это пара $\langle F, E \rangle$.

Регламентируемый сайт – это сайт, для которого существует нормативный акт организации-владельца, содержащий определение цели и задач веб-ресурса, основных структурных составляющих, правил добавления и изменения информации и т.д. Множество регламентируемых сайтов будем называть *регламентируемым множеством*. *Тематическое множество официальных сайтов* – это множество официальных сайтов организаций, относящихся к заданной тематике (например, реализующих одинаковые виды деятельности). *Фрагмент Веба для регламентируемого тематического целевого множества* – это пара $\langle F, E \rangle$, для которой T является регламентируемым тематическим множеством официальных сайтов.

Концептуальная модель фрагмента Веба для регламентируемого тематического целевого множества представляет собой следующий набор множеств:

$$KM = \{T, B, K, S, Att, Char\},$$

где T – целевое множество, B – множество сайтов ближайших окрестностей, K – множество сайтов-коммуникаторов, S – множество сайтов оболочки, Att – множество отношений, определяющих структуру KM ; $Char$ – множество атрибутов объектов KM .

Ближайшей окрестностью веб-сайта называется множество сайтов его организации-владельца, являющихся элементами U , на которые

существуют гиперссылки с официального сайта. Обозначим \mathcal{B}_t ближайшую окрестность сайта $t \in \mathcal{T}$, тогда $\mathcal{B} = \bigcup_{t \in \mathcal{T}} \mathcal{B}_t$.

Веб-коммуникатором называется сайт, принадлежащий \mathcal{U} и имеющий входящие ссылки с «достаточно большого» количества сайтов \mathcal{T} и/или исходящие ссылки на «достаточно большое» количество сайтов из \mathcal{T} . Определим две функции: $insitecount(\mathcal{A}, s)$ – количество сайтов из множества \mathcal{A} , имеющих гиперссылки на заданный сайт s , а $outsitecount(s, \mathcal{A})$ – количество сайтов из \mathcal{A} , на которые существуют гиперссылки с сайта s . Обозначим нижнее и верхнее пороговые значения, как λ и μ , λ, μ – целые и $\lambda \leq \mu$.

В качестве нижнего порогового значения принимается

$$\lambda = \text{round} \left(\frac{\sum_{t \in \mathcal{T}} insitecount(\mathcal{T}, t)}{|\mathcal{T}|} \right),$$

(*round* – функция округления до ближайшего целого). Значение λ характеризует «среднюю степень» интереса к сайту из \mathcal{T} , проявляемую со стороны других сайтов из \mathcal{T} .

Замеры показывают, что около 80% сайтов из \mathcal{U} имеют лишь по одному сайту из \mathcal{T} , ссылающемуся на них, поэтому говорить о «среднем» интересе к сайтам \mathcal{U} со стороны сайтов \mathcal{T} как о среднем арифметическом не представляется возможным. Поэтому верхнее пороговое значение задается следующим образом:

$$\mu = \text{round} \left(\frac{\sum_{u \in \mathcal{U} \setminus \mathcal{B}} insitecount(\mathcal{T}, u)}{\left\| \left\{ u \in \mathcal{U} \setminus \mathcal{B}, insitecount(\mathcal{T}, u) \geq \lambda \right\} \right\|} \right).$$

Определим функцию тематической близости сайта $u \in \mathcal{U}$ к тематике целевого множества \mathcal{T} как $T_genre: \mathcal{U} \rightarrow \{0,1\}$, $T_genre(u)=1$, если u тематически близок к \mathcal{T} и $T_genre(u)=0$, если нет. В настоящее время значения $T_genre(u)$ определяются на основе прямого просмотра сайтов экспертом.

Тематическим сайтом-посредником называется сайт $u \in \mathcal{U} \setminus \mathcal{B}$, для которого выполняется условия $insitecount(\mathcal{T}, u) \geq \mu$ & $outsitecount(u, \mathcal{T}) \geq \lambda$ и $T_genre(u)=1$. Множество тематических сайтов-посредников обозначим $\mathcal{K}^{\leftrightarrow}$.

Тематическим сайтом-коллектором называется сайт $u \in \mathcal{U} \setminus \mathcal{B}$, для которого выполняется условие $insitecount(\mathcal{T}, u) \geq \mu$ & $\lambda > outsitecount(u, \mathcal{T}) \geq 1$ и $T_genre(u)=1$. Множество тематических сайтов-коллекторов обозначим \mathcal{K}^{\leftarrow} .

Тематическим сайтом-индуктором называется сайт $u \in \mathcal{U} \setminus \mathcal{B}$, для которого выполняется условие $\mu > \text{insitecount}(\mathcal{T}, u) \geq \lambda$ & $\text{outsitecount}(u, \mathcal{T}) \geq \lambda$ и $T_genre(u) = 1$. Множество тематических сайтов-индукторов обозначим $\mathcal{K}^{\rightarrow}$.

Сайты-посредники, коллекторы и индукторы, для которых $T_genre(u) = 0$, составляют множество \mathcal{K}^{Other} («нетематические» веб-коммуникаторы).

Множество веб-коммуникаторов – это $\mathcal{K} = \mathcal{K}^{\leftrightarrow} \cup \mathcal{K}^{\leftarrow} \cup \mathcal{K}^{\rightarrow} \cup \mathcal{K}^{Other}$.

Множество сайтов оболочки $\mathcal{S} = \mathcal{U} \setminus (\mathcal{B} \cup \mathcal{K})$ составляют сайты из \mathcal{U} , не являющиеся сайтами ближайших окрестностей или веб-коммуникаторами.

\mathcal{S} разбивается на три непересекающихся подмножества: $\mathcal{S} = \mathcal{S}^{\mu} \cup \mathcal{S}^{\lambda} \cup \mathcal{S}^{\lambda}$; \mathcal{S}^{μ} – насыщенная оболочка ($\forall s \in \mathcal{S}^{\mu}: \text{insitecount}(\mathcal{T}, u) \geq \mu$), \mathcal{S}^{λ} – вязкая оболочка ($\forall s \in \mathcal{S}^{\lambda}: \mu > \text{insitecount}(\mathcal{T}, u) \geq \lambda$), \mathcal{S}^{λ} – разреженная оболочка ($\forall s \in \mathcal{S}^{\lambda}: \lambda > \text{insitecount}(\mathcal{T}, u) \geq 1$).

Множество отношений Att , определяющих структуру модели, это:

$Att = \{ \mathcal{T}\mathcal{T}, \mathcal{T}\mathcal{B}, \mathcal{B}\mathcal{T}, \mathcal{T}\mathcal{K}^{\leftrightarrow}, \mathcal{T}\mathcal{K}^{\leftarrow}, \mathcal{T}\mathcal{K}^{\rightarrow}, \mathcal{K}^{\leftarrow}\mathcal{T}, \mathcal{K}^{\rightarrow}\mathcal{T}, \mathcal{T}^{\mu}\mathcal{S}, \mathcal{T}^{\lambda}\mathcal{S}, \mathcal{T}\mathcal{S} \}$, где

$\mathcal{T}\mathcal{T} \subseteq \mathcal{T} \times \mathcal{T}$ – отношение, задающее структуру \mathcal{T} через множество гиперссылок, связывающих его сайты: $\mathcal{T}\mathcal{T} = \bigcup_{\substack{\forall s \in \mathcal{T} \\ \forall t \in \mathcal{T}, t \neq s}} \{ < s, t > \}$;

$\mathcal{T}\mathcal{B} \subseteq \mathcal{T} \times \mathcal{B}$ – отношение, задающее связи между \mathcal{T} и \mathcal{B} : $\mathcal{T}\mathcal{B} = \bigcup_{\substack{\forall t \in \mathcal{T} \\ \forall b \in \mathcal{B}}} \{ < t, b > \}$;

$\mathcal{B}\mathcal{T} \subseteq \mathcal{B} \times \mathcal{T}$ – обратное отношение по отношению к $\mathcal{T}\mathcal{B}$;

отношения $\mathcal{T}\mathcal{K} \subseteq \mathcal{T} \times \mathcal{K}$ и $\mathcal{K}\mathcal{T} \subseteq \mathcal{K} \times \mathcal{T}$ задаются аналогично $\mathcal{T}\mathcal{B}$ и $\mathcal{B}\mathcal{T}$, где в качестве \mathcal{K} могут быть $\mathcal{K}^{\rightarrow}$, \mathcal{K}^{\leftarrow} или $\mathcal{K}^{\leftrightarrow}$;

$\mathcal{T}\mathcal{S} \subseteq \mathcal{T} \times \mathcal{S}$ – отношение, задающее связи между \mathcal{T} и множеством \mathcal{S} , где в качестве \mathcal{S} могут быть \mathcal{S}^{μ} , \mathcal{S}^{λ} или \mathcal{S}^{λ} : $\mathcal{T}\mathcal{S} = \bigcup_{\substack{\forall t \in \mathcal{T} \\ \forall s \in \mathcal{S}}} \{ < t, s > \}$;

$\mathcal{S}\mathcal{T} = \emptyset$ – отношение, задающее связи между \mathcal{T} и \mathcal{S} .

Множество атрибутов объектов \mathcal{KM}

$Char = \{ Distr, Ind \}$.

Показатель тематического распределения сайтов множества $distr(\mathcal{A})$ характеризует, какова доля элементов из \mathcal{A} , тематически близких к тематике целевого множества. В качестве \mathcal{A} принимаются \mathcal{K}^{Other} , \mathcal{B} , \mathcal{S}^{μ} , \mathcal{S}^{λ} , и \mathcal{S}^{λ} , поэтому

$Distr = \{ distr(\mathcal{K}^{Other}), distr(\mathcal{B}), distr(\mathcal{S}^{\mu}), distr(\mathcal{S}^{\lambda}), distr(\mathcal{S}^{\lambda}) \}$.

Множество показателей степени связности Ind характеризует множества отношений Att по степени связности самого целевого множества T , а также T и всех подмножеств множеств B, K, S , составляющих \mathcal{KM} :

$$Ind = \{ind(TT), ind(TB), ind(BT), ind(TK^{\leftarrow}), ind(TK^{\rightarrow}), ind(K^{\leftarrow}T), ind(K^{\rightarrow}T), ind(S^{\downarrow}S), ind(TS^{\downarrow}), ind(TS^{\uparrow}), ind(S^{\downarrow}T), ind(S^{\uparrow}T)\}.$$

Здесь $ind(TT) = |TT|/|T|$. Остальные показатели равны суммарному количеству гиперссылок, входящих в соответствующее множество, деленному на мощность этого множества, или суммарному количеству гиперссылок, исходящих с соответствующего множества, деленному на мощность этого множества.

Процедура построения \mathcal{KM} состоит из следующих этапов:

1. задание целевого множества T ;
2. формирование сопутствующего множества \mathcal{U} для заданного T ;
3. формирование множества ближайших окрестностей B ;
4. вычисление пороговых значений λ и μ и задание функции T_genre ;
5. формирование множества подмножеств коммуникаторов K и множества-оболочки S ;
6. формирование множества отношений Att ;
7. определение множества значений атрибутов $Char$.

Первый этап выполняется вручную, остальные этапы выполняются с использованием информационной системой для вебометрических исследований (ИС4ВИ), разработанной в Институте прикладных математических исследований Карельского научного центра РАН.

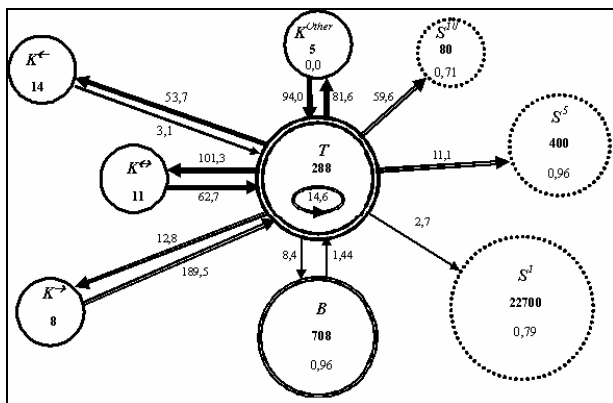


Рис. 1. Диаграмма концептуальной модели академического фрагмента Веба

На рис. 1 приводится диаграмма концептуальной модели, реализованная для академического фрагмента Веба. В качестве T принято

288 официальных сайтов учреждений РАН. Каждое из подмножеств \mathcal{KM} обозначено соответствующим кружком, в котором дано обозначение подмножества, его мощность и значение $dist$ (для некоторых из них). Стрелки задают отношения между подмножествами и помечаются значениями ind . Похожие результаты были получены в случае реализации концептуальной модели для университетского фрагмента Веба и для фрагмента Веба органов государственной власти РК.

Анализ реализаций концептуальной модели позволяет сделать ряд общих выводов об организации реальных фрагментов Веба. Большинство сайтов характеризуются тематической близостью по отношению к \mathcal{T} , но степень их влияния на присутствие в Вебе сайтов целевого множества, реализуемая через гиперссылки, различна. Сайты, входящие в \mathcal{S} , составляя от 80 до 96% всех сайтов реальных фрагментов Веба, такого влияния практически не оказывают вследствие отсутствия гиперссылок, сделанных с них на сайты \mathcal{T} . Сайты множества ближайших окрестностей также слабо влияют на связность \mathcal{T} , поскольку в основном имеют гиперссылки, сделанные на головные сайты организаций, входящих в \mathcal{T} , а не на другие сайты из \mathcal{T} . Высокую степень связности с \mathcal{T} демонстрируют сайты множеств $\mathcal{K}^{\rightarrow}$, $\mathcal{K}^{\leftrightarrow}$ и \mathcal{K}^{\leftarrow} .

В третьей главе рассмотрено применение теоретико-графовых методов к исследованию связности регламентируемых тематических фрагментов Веба, как одной из основных характеристик присутствия \mathcal{T} в Вебе.

Веб-графом $G_{FE}(V, E)$, заданным на фрагменте Веба $\langle \mathcal{F}, \mathcal{E} \rangle$ называется ориентированный граф без петель и кратных дуг, множество вершин которого является объединением целевого множества и множества веб-коммуникаторов $V = \mathcal{T} \cup \mathcal{K}$, а множество дуг E задается следующим образом: $\forall v_1, v_2 \in V: (v_1, v_2) \in E \Leftrightarrow \exists \langle v_1, v_2 \rangle \in \mathcal{T}\mathcal{T} \cup \mathcal{T}\mathcal{K} \cup \mathcal{K}\mathcal{T} \cup \mathcal{K}\mathcal{K}$.

Исследуются вопросы связности целевого множества на административном каркасе и административных подмножествах. Для организаций-владельцев сайтов целевого множества, имеющих иерархическую структуру, иерархический граф обозначен $G_H = (\mathcal{D}^T, E_H)$, где \mathcal{D}^T – это множество наименований организаций-владельцев сайтов целевого множества, а E_H – множество дуг. E_H определяется следующим образом: для $\forall u, d \in \mathcal{D}^T: (u, d) \in E_H$ тогда и только тогда, когда организация d находится в отношении подчиненности к организации u .

Пусть даны $ds, dt \in \mathcal{D}^T$, причем организация dt находится в отношении подчиненности к организации ds , т.е. $(ds, dt) \in E_H$, и сайтам организаций ds, dt

соответствуют вершины $s, t \in T$. Определим *каркасную функцию связности сверху вниз* (Skeleton Force of Connectivity) $SFC^+(s, t)$ следующим образом:

$$SFC^+(s, t) = \{1 | (\exists (s, t) \in E_k) \& (ds, dt) \in E_H, 0 | (\neg \exists (s, t) \in E_k) \& (ds, dt) \in E_H\}.$$

Обозначим через $\mathcal{D}_{ds} = \{dt_1, dt_2, \dots\}$, $\mathcal{D}_{ds} \subset \mathcal{D}^T$ множество организаций, которые находятся в отношении подчиненности к организации ds (\mathcal{D}_{ds} – т.н. *административное подмножество*). Обозначим s сайт организации ds , а \mathcal{T}_s – множество сайтов, соответствующих организациям из \mathcal{D}_{ds} (любой организации $dt_i \in \mathcal{D}_{ds}$ взаимнооднозначно соответствует сайт $t_i \in \mathcal{T}_s$). Определим функцию связности сверху вниз между сайтом s и всеми сайтами $t \in \mathcal{T}_s$:

$$SFC^+(s, \mathcal{T}_s) = \frac{\sum_{t \in \mathcal{T}_s} SFC^+(s, t)}{|\mathcal{T}_s|}$$

Каркасная функция связности снизу вверх $SFC^-(s, t)$ и *функция связности снизу вверх* $SFC^-(s, \mathcal{T}_s)$ определяются аналогично. *Функцию связности* для пары сайтов s и t можно задать как $SFC(s, t) = \alpha SFC^+(s, t) + \beta SFC^-(s, t)$, где α и β являются коэффициентами, отражающими важность иерархии сверху и вниз снизу вверх ($\alpha, \beta > 0$, $\alpha + \beta = 1$). *Функция связности* $SFC(s, \mathcal{T}_s)$ для сайта s и подмножества \mathcal{T}_s определяется аналогично.

Исследование реальных фрагментов Веба показывает существенное влияние административного каркаса на связность целевого множества. Значения функции связности для академического фрагмента Веба приведены на рис. 2.

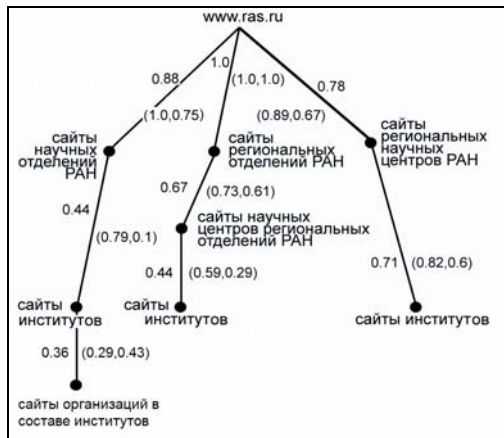


Рис. 2. Значения функции связности на административном каркасе

Дуги помечены следующим образом: слева от дуги (или над ней) стоит значение $SFC(s, \mathcal{T}_s)$ для соответствующих s и \mathcal{T}_s ; справа от дуги (или под ней) пара значений в скобках обозначает $(SFC^+(s, \mathcal{T}_s), SFC(s, \mathcal{T}_s))$.

Графом административного подмножества вершины $s \in \mathcal{T}$ называется граф $G_s = G(\mathcal{T}_s, E_s)$, где множество дуг $E_s \subseteq E$ содержит только те дуги из E , которые связывают вершины из \mathcal{T}_s . В качестве «слабой» характеристики связности принята функция CFC_s (*Community Force of Connectivity*)

$$CFC_s = \frac{|E_s|}{|\mathcal{T}_s| \times (|\mathcal{T}_s| - 1)}.$$

Исследования различных подграфов $G_s = G(\mathcal{T}_s, E_s)$ показывают незначительную связность административных подмножеств академического фрагмента Веба и фрагмента Веба органов государственной власти РК. Например, для институтов Отделения математических наук ($s = omn.ras.ru$, $|\mathcal{T}_s| = 10$) имеем $CFC_s = 0,161$.

Предложена следующая процедура исследования компонент сильной связности (*KCC*) веб-графа $G_{FE}(V, E)$. Исследуется подграф $G^{\mathcal{T}}_{FE}(\mathcal{T}, E^{\mathcal{T}}) \subseteq G_{FE}(V, E)$, построенный на целевом множестве \mathcal{T} , в качестве множества дуг которого принимается подмножество дуг $E^{\mathcal{T}} \subseteq E$, соединяющих только вершины, входящие в множество \mathcal{T} . Вычисляются полустепени захода (*Indegree*) и исхода (*Outdegree*) вершин из \mathcal{T} , формируются все *KCC*, изолированные вершины, вершины, имеющие только входящие и только исходящие дуги, определяется максимальная *KCC* и ее диаметр. Построенные конструкции и вычисленные характеристики позволяют дать содержательную интерпретацию взаимосвязей сайтов и охарактеризовать присутствие в Вебе сайтов целевого множества.

Далее исследуется веб-граф $G_{FE}(V, E)$, что позволяет дать содержательную интерпретацию взаимосвязей сайтов и оценить влияние веб-коммуникаторов на изменение связности сайтов \mathcal{T} . Изначально важная роль сайтов-коммуникаторов может быть оценена как с точки зрения их влияния на прирост взаимосвязей сайтов \mathcal{T} , так и изменения диаметра *KCC*.

Приведем некоторые результаты, полученные для университетского фрагмента Веба ($|\mathcal{T}| = 56$). Значения полустепеней захода и исхода для некоторых вершин \mathcal{T} : *Indegree*: сайт Новосибирского госуниверситета – 15, Московского – 14, Санкт-Петербургского – 12; *Outdegree*: сайт Бурятского госуниверситета – 43, Горно-Алтайского – 35, Нижегородского – 30. Статистической зависимости между значениями *Indegree* и *Outdegree* не обнаруживается.

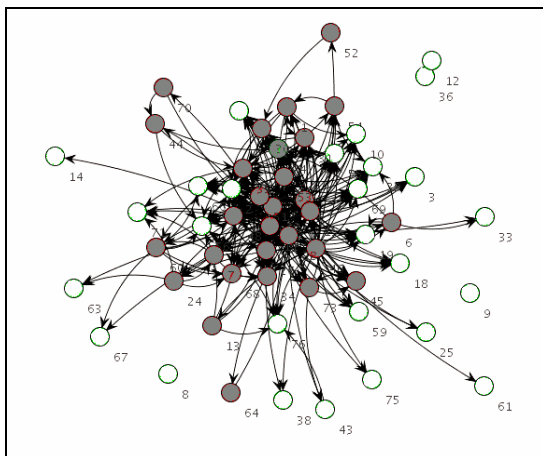


Рис. 3. Веб-граф $G_{FE}^T(T, E^T)$ университетского фрагмента Веба

В университетском веб-графе $G_{FE}^T(T, E^T)$ обнаружена единственная КСС, содержащая 29 вершин с диаметром 5; из остальных 27 вершин у 21 вершины имеются исходящие, но нет входящих дуг, 4 являются изолированными, а 2 имеют исходящие и входящие дуги, но не входят в КСС. На рис. 3 вершины, составляющие КСС, изображены серыми кружками, а остальные вершины – светлые кружки.

В веб-графе $G_{FE}(V, E)$ мощность КСС увеличивается до 70 вершин, а её диаметр остается равен 5. Количество «вовлеченных» сайтов из T в этом случае равняется 17.

По результатам исследований веб-графов реальных фрагментов Веба можно сделать следующие выводы:

1. степень связности целевых множеств невысока, следствием чего является их незначительное присутствие в Вебе,
2. административная иерархия организаций-владельцев сайтов находит хорошее отражение в Вебе и повышает присутствие сайтов в Вебе,
3. связь между сайтами организаций, подчиненных одной и той же вышестоящей организации, является весьма незначительной, что существенно снижает их присутствие в Вебе.
4. существенно влияние сайтов-коммуникаторов на увеличение связности сайтов целевого множества для всех рассмотренных фрагментов Веба.

В четвертой главе исследуются вопросы ранжирования сайтов тематического сообщества, связанные с их индивидуальными характеристиками присутствия в Вебе. В первоначальной постановке задача была сформулирована *Cybermetrics Lab*. Используются вебометрические

индикаторы, измеряемые с помощью поисковых машин (*Google, Yahoo, Live Search* и *Exalead*): S – *размер сайта* (количество страниц, обнаруживаемых на сайте поисковыми машинами), V – *видимость сайта* (количество гиперссылок на сайт с других веб-ресурсов), R – *количество полнотекстовых файлов* и Sc – *научность сайта* (количество ссылок на сайт, обнаруживаемых *Google Scholar*). Значения для S , V и R определяются как сумма результатов замеров по каждой машине (для S минимальное и максимальное значение отбрасываются). По каждому индикатору производится ранжирование сайтов по убыванию их значений. Для обозначения ранга по заданному индикатору используются обозначения $RankV$, $RankS$, $RankR$ и $RankSc$ соответственно (наивысший ранг равен 1). Интегральный показатель W вычисляется по формуле $W = K_v \cdot RankV + K_s \cdot RankS + K_r \cdot RankR + K_{sc} \cdot RankSc$, где $K_v=4$, $K_s=2$, $K_r=1$ и $K_{sc}=1$ – весовые коэффициенты при рангах по соответствующим индикаторам (на 2008 год).

Поясним содержательный смысл индикаторов присутствия через оценки возможности попадания пользователя на заданный сайт. Все множество пользователей Веба по отношению к данному сайту можно разделить два подмножества: первое – это пользователи, знающие данный сайт и обращающиеся к нему по адресу, а второе – это все остальные пользователи, которые имеют потенциальную возможность узнать о сайте различными способами. Индикаторы характеризуют возможность попадания на сайт через один из четырех таких способов: S – обнаружение с помощью поисковой машины и переход на сайт, V – переход с другого сайта по ссылке, R – обнаружение с помощью поисковой машины полнотекстового файла данного сайта и переход на сайт и Sc – переход через ссылку, найденную поисковой машиной для научных публикаций. Следовательно, критерий, основанный на S , V , R и Sc , характеризует возможность попадания пользователей из второго подмножества на заданный сайт. Для российских университетских сайтов показана сильная попарная статистическая зависимость между S , R и Sc , поэтому для каждого ранжируемого сайта достаточно вычислить значение одного из интегральных показателей, например, количества страниц как средневзвешенного значения

$$Sint(i) = P_{\text{Яндекс}} \cdot S_{\text{Яндекс}}(i) + P_{\text{Rambler}} \cdot S_{\text{Rambler}}(i) + P_{\text{Google}} \cdot S_{\text{Google}}(i),$$

где i – номер сайта, а $P_{\text{Яндекс}}$, P_{Rambler} и P_{Google} – показатели популярности поисковых систем. Производится нормирование всех $Sint(i)$:

$$Sintnorm(i) = \frac{Sint(i)}{\sum_i Sint(i)}.$$

Для каждого сайта измеряется значение CY – тематического индекса цитирования Яндекса и производится нормирование :

$$CY_{norm}(i) = \frac{CY(i)}{\sum_i CY(i)}.$$

Интегральный показатель вычисляется по формуле $WRR(i) = P_s \cdot S_{intnorm}(i) + P_v \cdot CY_{norm}(i)$, где P_s и P_v – вероятности реализации того или иного способа попадания на сайт. Далее производится ранжирование сайтов по убыванию значений $WRR(i)$.

Для реализации процедуры ранжирования используются данные *LiveInternet* «Сайты Рунета». В качестве значений $P_{Яндекс}$, $P_{Rambler}$ и P_{Google} берутся данные по статистике переходов. Для вычисления значений P_s и P_v берутся переходы с поисковых систем SE , переходы с каталогов и рейтингов Cat и количество переходов по ссылкам L . P_s и P_v вычисляются по формулам:

$$P_s = \frac{SE}{SE + Cat + L}, \quad P_v = \frac{Cat + L}{SE + Cat + L}.$$

Сопоставление результатов ранжирования по методу *Cybermetrics Lab* и предложенному алгоритму показывает, что полученные результаты близки между собой, например, в первую двадцатку как по одному, так и по другому алгоритму входят четырнадцать сайтов одних и тех же университетов.

Измерения вебометрических индикаторов показывают существенное отставание по всем индикаторам от зарубежных аналогов.

Пятая глава посвящена математическим задачам, разработка и решение которых ведут к улучшению присутствия в Вебе сообществ сайтов в смысле некоторых заданных критериев. Первая из задач относится к дискретным оптимизационным задачам и позволяет найти такой способ расстановки гиперссылок в заданном множестве сайтов, который повышает их присутствие в Вебе с точки зрения поисковых машин. На основе предложенной модели исследованы вопросы согласованного поведения реальных веб-сообществ. Выявлены подмножества российского Веба, поведение которых является близким к согласованному. Во второй задаче исследуется вопрос о том, каков должен быть вклад организаций-владельцев заданного множества сайтов на создание сайта-коммуникатора, существенно улучшающего его связность. Построена и исследована математическая модель дележа затрат на создание веб-коммуникатора, основанная на методах кооперативной теории игр. В третьей задаче исследуется вопрос о том, при каких условиях сайты получают дополнительную выгоду (в смысле количества посещений пользователями) от участия в множестве сайтов, ссылающихся на один и тот же сайт-коммуникатор и имеющих обратные гиперссылки с него. Получены решения, позволяющие каждому участнику сообщества оценивать полезность его участия в нем. На практике решение указанных задач могут

быть положены в основу административных регламентов, направленных на улучшение присутствия в Вебе регламентируемых тематических сообществ сайтов.

Рассмотрим веб-сообщество со следующими характеристиками:

n – количество сайтов-участников,

c_i – значимость i -го участника, $c_i > 0, \forall i = 1..n$,

m_i – максимально возможное количество ссылок от i -го на других участников, $m_i > 0, \forall i = 1..n$.

Определим матрицу ссылок $X=(x_{ij}), i, j=1..n$, где $x_{ij}=1$, если существует ссылка от i -го участника к j -му, и $x_{ij}=0$, если ссылки не существует. Пусть $F(X)$ – функция, характеризующая некоторый интегральный показатель значимости всех участников и зависящая от того, каким образом расставлены ссылки между ее участниками.

Тогда задача заключается в нахождении матрицы X , удовлетворяющей заданным ограничениям и доставляющей оптимальное значение целевой функции:

$$F(X) = \underset{x_{ij}}{opt} \quad (1).$$

Основные предположения, принятые для задания вида функции приращения значимости, заключаются в следующем:

- чем больше ссылок на ресурс, тем он становится «значимее»,
- чем больше значимость ресурса i , тем больше возрастает значимость ресурса j , если $x_{ij}=1$,
- чем больше исходящих ссылок от ресурса i , тем меньше приращение значимости каждого ресурса j , для которого $x_{ij}=1$.

Считаем, что внешние влияния на участников системы не существенны, и поэтому изменение значимости j -го участника может быть представлено следующей формулой: $\forall j = \overline{1, n}: \tilde{c}_j = c_j + \sum_{i=1}^n x_{ij} c_i \alpha_i$.

Здесь $\alpha_i = \beta / \sum_{k=1}^n x_{ik}$ или содержательно – если пользователь попал на заданную страницу, то он равновероятно отправится далее по любой из имеющихся на ней ссылок и при этом вероятность того, что он не закончит работу в Вебе на этой странице, равна так называемому «дампинг-фактору» $\beta, 0 < \beta < 1$. По аналогии с Брином и Пейждем чаще всего принимается $\beta=0,85$ (Brin S., Page L. *The Anatomy of a Large-Scale Hypertextual Web Search Engine // Computer Networks and ISDN Systems. – 1998. – № 30. – P. 107-117*).

Тогда изменение значимости j -го участника выражается следующей формулой:

$$\forall j = \overline{1, n}: \tilde{c}_j = c_j + \sum_{i=1}^n \frac{x_{ij}}{\sum_{k=1}^n x_{ik}} c_i \beta \quad (2).$$

Введем следующие ограничения:

$$x_{ii} = 0, \quad i = \overline{1, n} \quad (3)$$

$$x_{ij} = 0, 1, \quad i = \overline{1, n}, \quad j = \overline{1, n} \quad (4)$$

$$\sum_{j=1}^n x_{ij} \leq m_i, \quad i = \overline{1, n} \quad (5)$$

$$\sum_{j=1}^n x_{ij} \geq 1, \quad i = \overline{1, n} \quad (6)$$

Ограничения (3) учитывают, что сайту нельзя дать ссылку на самого себя, (4) – ссылки с i -го сайта либо существуют, либо нет. Ограничение (5) означает, что количество исходящих ссылок ограничено некоторыми реальными соображениями, например, оно меньше n , т.к. в этом случае поисковые системы обнуляют значимости всех участников сообщества. Ограничение (6) означает, что участником может быть сайт, который обязательно делится своей значимостью с другими, но обратное не обязательно.

Рассматриваются различные варианты задачи в зависимости от вида целевой функции (1) и дополнительных содержательных ограничений.

В первом случае договоренность участников сообщества можно сформулировать так: распределение ссылок должно привести в первую очередь к увеличению значимости наименее значимых ресурсов, при этом суммарный прирост значимости по всем участникам системы должен быть максимальным. Тогда целевая функция принимает вид

$$F(X) = \sum_{j=1}^n \tilde{c}_j \lambda_j \rightarrow \max_X \quad (1'),$$

где коэффициенты $0 \leq \lambda_j \leq 1$ имеют следующий содержательный смысл, – чем больше начальное значение c_j , тем меньше значение λ_j .

Во втором случае договоренность участников можно сформулировать так: распределение ссылок среди участников должно привести к минимальному отклонению полученных значимостей каждого участника от нового среднего значения по всему сообществу:

$$F(X) = \sum_{j=1}^n \left(\frac{\sum_{k=1}^n \tilde{c}_k}{n} - \tilde{c}_j \right)^2 \rightarrow \min_{x_j} \quad (1'').$$

Исследованы свойства моделей (1'), (3-6) и (1''), (3-6), учитываемые при разработке точных и приближенных алгоритмов решения задач.

В рамках конкурса Интернет-математика 2007 выполнен проект по выявлению согласованного поведения реальных веб-сообществ. Были взяты 20 реальных сообществ, содержащих от 7 до 84 участников, относящихся к различной тематике. Использовались предоставленные компанией Яндекс наборы данных о ссылках сайтов друг на друга по состоянию на 7 декабря 2007 года, содержащие информацию о 2714279 сайтах.

В модели (1'), (3-6) и (1''), (3-6) был введен ряд дополнительных ограничений, появившихся в процессе их адаптации к реальным сообществам. Например, обе модели изначально трактуются как регламентируемые, – если известен оптимальный способ расстановки ссылок между участниками, то он и будет директивно реализован, – поэтому неравенства (5) заменяются на равенства (5')

$$\sum_{j=1}^n x_{ij} = m_i, \quad i = \overline{1, n} \quad (5').$$

В качестве значений $\tilde{c}_i, i = \overline{1, n}$ были взяты значения *тематического индекса цитирования Яндекса (mИЦ)* соответствующих сайтов. Поскольку $\tilde{c}_i, i = \overline{1, n}$ – это значения, которые являются результатом согласованных действий участников сообщества, из (2) и (5') следует, что начальные значения $c_i, i = \overline{1, n}$ являются решениями системы линейных уравнений:

$$\begin{cases} c_1 + \sum_{i=1}^n \frac{x_{ij}}{L_i} \cdot c_i = \tilde{c}_1 \\ \dots\dots\dots \\ c_k + \sum_{i=1}^n \frac{x_{ij}}{L_i} \cdot c_i = \tilde{c}_k \\ \dots\dots\dots \\ c_n + \sum_{i=1}^n \frac{x_{ij}}{L_i} \cdot c_i = \tilde{c}_n \end{cases}'$$

где $L_i, i = \overline{1, n}$ – общее количество исходящих ссылок от i -го участника сообщества. Обнаружено, что в рамках социальной модели наиболее согласовано поведение сообщества *Религия: Православие*, а в рамках кооперативной модели – сообщество сайтов министерств Правительства РФ.

С использованием методов кооперативных игр рассматривается вопрос о вкладе организаций-владельцев сообщества сайтов в создание общего сайта-коммуникатора. Предполагается, что веб-граф $G_{T,FE}^T(T, E^T)$ –

сильно связный и веса его дуг равны 1. Обозначим $d(i,t)$ длину кратчайшего пути из i в t , $i,t \in \mathcal{T}$. Критерием доступности сайта t на целевом множестве \mathcal{T} является средняя длина пути, которая задается следующим образом:

$$midd(t) = \frac{\sum_{i \in \mathcal{T}, i \neq t} d(i,t)}{n-1}.$$

Предположим, что игроки-владельцы сайтов договорились создать веб-коммуникатор h , с которого обязательно будут сделаны гиперссылки с весом 1 на любой сайт из \mathcal{T} и с каждого сайта из \mathcal{T} будет сделана гиперссылка на h , имеющая вес 1. При добавлении сайта h средняя длина пути в вершину $t \in \mathcal{T}$ из всех других вершин, кроме h и её самой, равна

$$midd^h(t) = \frac{\sum_{i \in \mathcal{T}, i \neq t} d^h(i,t)}{n-1}.$$

При этом верно неравенство $midd^h(i) \leq 2$ (на сайт i существуют ссылки с некоторых сайтов из \mathcal{T} и длина пути равна 1, либо кратчайший путь проходит через сайт-коммуникатор h и длина пути равна 2). Выигрыш i -го участника от создания веб-коммуникатора h равен $v(i) = midd(i) - midd^h(i)$.

Обозначим Z стоимость разработки и поддержки сайта h , а взнос каждого игрока z_i , $Z = \sum_{i \in \mathcal{T}} z_i$. Требуется ответить на вопрос, каковы должны быть значения z_1, z_2, \dots, z_n справедливые (в некотором смысле) для каждого игрока-владельца сайта целевого множества.

Считаем коалицией множество владельцев сайтов из \mathcal{T} , участвующих в создании h , причем h будет ссылаться только на участников коалиции, и только они будут ссылаться на коммуникатор. Зададим выигрыш коалиции S как $v(S) = \sum_{i \in S} v(i)$, где значение характеристической функции для i -го участника $v(i) = midd(i) - midd^h_S(i)$ рассчитывается с учетом того, что коммуникатор создается только для членов коалиции S . Именно поэтому в качестве вычитаемого принято значение $midd^h_S(i)$, т.е. средняя длина пути в вершину i из всех других вершин коалиции S , кроме h и её самой.

Решение основано на разделении платы на создание веб-коммуникатора пропорционально компонентам вектора Шепли, строящемуся с учетом среднего вклада каждого участника в выигрыш гранд-коалиции. Плата за создание сайта-коммуникатора z_1, z_2, \dots, z_n делится между членами сообщества \mathcal{T} пропорционально величинам

$$\Phi_i = \sum_{S \subseteq \mathcal{T}} \frac{(|S|-1)!(n-|S|)!}{n!} (v(S) - v(S \setminus \{i\})).$$

Задача обобщается на случай взвешенного графа, когда его дуги имеют веса, большие 0.

Рассмотрена математическая модель, связанная с увеличением посещаемости сайтов за счет участия в *сообществе динамического каталога (СДК)*. В СДК имеется головной сайт, на первой странице которого расположена видимая часть динамического каталога, содержащего ссылки на всех остальных участников сообщества, а остальные сайты, называемые рядовыми участниками, имеют прямые ссылки на головной сайт со своей первой страницы. Возможный выигрыш рядового сайта от участия в СДК заключается в увеличении его посещаемости за счет роста переходов на него с каталога головного сайта.

Попадая на сайт любого участника СДК, пользователь выполняет одно из следующих действий:

- если это головной сайт, то пользователь может остаться на этом сайте и сеанс заканчивается, либо перейти через каталог на рядового участника (сеанс заканчивается);

- если это рядовой участник, то пользователь может остаться на этом сайте (сеанс заканчивается), либо перейти головной сайт (сеанс заканчивается).

Введем следующие обозначения:

p_i – вероятность попадания пользователя на i -й сайт СДК, $\sum_{i=0}^n p_i = 1$;

p_i^0 – вероятность перехода с i -го рядового сайта на головной сайт (вероятность того, что пользователь, попав на i -й рядовой сайт, останется на нем, равна $1 - p_i^0$);

q_0 – вероятность того, что пользователь, попав на головной сайт, останется на нем;

q_j^{cat} – вероятность перехода на любой рядовой сайт с j -й позиции каталога,

$$\sum_{j=0}^n q_j^{cat} = 1.$$

На головном сайте расположена видимая часть каталога, содержащая k позиций для ссылок ($k \leq n$); $q_1^{cat} \geq q_2^{cat} \geq \dots \geq q_k^{cat}$, а для $\forall j = \overline{k+1, n}$, $q_j^{cat} = 0$.

Пусть r_i – позиция рядового участника в каталоге в некоторый момент времени, $i = \overline{1, n}$, $\forall i, j = \overline{1, n} : r_i \neq r_j$. Позиция i -го участника изменяется за один сеанс следующим образом:

- если пользователь переходит с i -го рядового сайта на головной сайт, то r_i уменьшается, а r_{i-1} — увеличивается на единицу (либо r_i не изменяется, если он уже был равен 1);
- если пользователь попадает с головного сайта на i -го участника, то r_i увеличивается, а r_{i+1} — уменьшается на единицу (либо r_i не изменяется, если он уже был равен n);
- если пользователь за сеанс не делает перехода с головного сайта на рядовой (или наоборот), то позиции всех рядовых сайтов остаются прежними.

Обозначим $\Pi = (\pi_{ij})$ матрицу предельных вероятностей, где π_{ij} — это вероятность того, что $r_i = j$. Зная матрицу Π можно определить значения q_i — вероятностей перехода с каталога головного сайта на любой рядовой сайт СДК:

$$\forall i = \overline{1, n} : q_i = \sum_{j=1}^n q_j^{cat} \pi_{ij} .$$

Для i -го рядового сайта доход от участия в СДК как вероятность перехода на i -й сайт с головного сайта равен $d_i = p_0(1 - q_0)q_i$, потери от участия в СДК как вероятность перехода с i -го сайта на головной сайт — $u_i = p_i p_i^0$, а функция выигрыша от участия в сообществе:

$$F_i = p_0(1 - q_0) \sum_{j=1}^n q_j^{cat} \pi_{ij} - p_i p_i^0 .$$

Для нахождения матрицы Π построена система уравнений, содержащая $n^2 + 2n$ уравнений и n^2 неизвестных, при этом n^2 уравнений являются уравнениями третьего порядка относительно неизвестных π_{ij} . На неизвестные также накладываются условия $0 \leq \pi_{ij} \leq 1$ для $\forall i, j = \overline{1, n}$.

Для случая системы из двух участников получено достаточное условие выигрыша каждого из участников

$$p_0 > \frac{\max\{p_1 p_1^0, p_2 p_2^0\}}{(1 - q_0) q_2^{cat}} .$$

Для общего случая разработана имитационная модель СДК. Результаты моделирования показывают, что чем больше значение вероятности p_0 , тем больше рядовых участников оказываются в выигрыше. Причем наибольший выигрыш получают участники сообщества, у которого вероятность попадания пользователя на него не самая большая, но и не самая маленькая. При достаточно малых значениях вероятности p_0 ни один рядовой участник не получает положительного выигрыша, то есть если головной сайт не слишком популярен по сравнению с остальными

рядовыми участниками, то нет смысла участвовать в таком СДК.

В шестой главе описаны основные программные средства, используемые в диссертационной работе для сбора и обработки вебометрической информации. В состав специализированного комплекса программ, названного Информационной системой для вебометрических исследований (*ИСЧВИ*), входят поисковый робот *LPR* (от *Link, Page, Robot*), задачей которого является сбор внешних гиперссылок, и база данных для вебометрических исследований, представляющая множество взаимосвязанных реляционных таблиц и набор операций над ними.

Единицей анализа *LPR* является страница, переданная веб-сервером клиенту по *http*-запросу и имеющая *mime*-тип «*text/html*». Начальным адресом сканируемого сайта является доменное имя сайта. Рассматриваются гиперссылки, заданные в тегах `<a>` в значении параметра *href*, либо в тегах `<frame>` в значении параметра *src*. Если там обнаруживается абсолютный адрес вида

`[http[s]://]доменное_имя[:порт]/путь_к_странице`

и если доменное имя отлично от адреса сайта, то такая ссылка считается внешней гиперссылкой. В качестве контекста внешней ссылки рассматривается текст, расположенный между тегами `<a>` и ``.

Анализ страницы начинается с *http*-запроса к странице и чтения заголовков, по которым выясняются следующие моменты:

1. Существует ли такая страница, если нет, то ссылка на страницу помечается как неработающая, и анализ данной страницы завершается.
2. Не является ли данная страница перенаправлением на другую страницу. Если есть перенаправление, то анализируется ссылка-перенаправление. Если это ссылка на другой сайт, то она записывается как внешняя ссылка, а анализируемая внутренняя ссылка удаляется. Если это ссылка на внутреннюю страницу, то переписывается ссылка, и данный этап завершается.
3. Является ли данный объект *html*-страницей. Если нет, то данная ссылка помечается как «документ» и данный этап завершается.

Далее считывается код страницы, который проверяется на наличие перенаправлений на уровне страницы и в случае перенаправления подвергается обработке как в п.2. Если перенаправлений нет, то из кода извлекаются теги фреймов и гиперссылок, из которых отбирается вся требуемая информация о ссылках, включая контекст, которая записывается в отдельные таблицы *ВНУТРЕННИЕ ССЫЛКИ* и *ВНЕШНИЕ ССЫЛКИ*, соответствующие сканируемому сайту. Все ссылки, являющиеся ссылками на документы, записываются в таблицу внутренних ссылок с пометкой «документ». Остальные ссылки с пометкой «непроверенная страница», записываются в эту же таблицу, если их там еще нет.

Процесс сканирования сайта идет как поиск вширь: начальная страница, потом непроверенные страницы первого уровня, потом второго и т.д. Если по каким-то причинам сканирование было прервано, то оно может быть возобновлено с точки прерывания. *LPR* может быть в любой момент остановлен пользователем благодаря интерактивному режиму отслеживания работы робота.

В базе данных реализовано более 15 операций, соответствующих основным задачам вебометрических исследований. Программный комплекс *ИСАВИ* разработан на языке *PHP* и работает под управлением веб-сервера Apache с интегрированным модулем *PHP* и СУБД *MySQL*. Реализован прототип компьютерной программы визуализации веб-графов, который в качестве исходных данных использует информацию из базы данных *ИСАВИ*.

Проведенное сканирование более 3000 сайтов академического, университетского фрагментов Веба и фрагмента Веба органов государственной власти Карелии позволило сформировать базы данных, являющихся экспериментальной основой для исследований реальных фрагментов Веба.

ЗАКЛЮЧЕНИЕ

В диссертации сформулирована и решена важная научная задача разработки и развития теоретических и практических разделов вебометрики, направленных на улучшение присутствия в Вебе сайтов, входящих в состав множеств, являющихся тематическими регламентируемыми веб-ресурсами. Проведенные исследования представляют собой новый подход к исследованию Веба, базирующийся на разработанных формализованных методах и специализированных программных средствах для детального обследования фрагментов Веба, заключающийся в построении и анализе описательных и формальных моделей, выявлении наиболее значимых компонент и постановке и решении математических задач рационального поведения веб-сайтов, входящих в данные компоненты. Основные научные и практические результаты работы состоят в следующем:

1. Предложен метод автоматизированного построения фрагментов Веба на основе задаваемого целевого множества, веб-ресурсы которого относятся к одной тематике и являются регламентируемыми, т.е. создаются и развиваются по заранее сформулированным правилам, утвержденным в виде нормативных документов организаций-владельцев ресурсов. Основу метода составляет формализованная концептуальная модель фрагмента Веба для регламентируемого тематического целевого множества, представленная в виде теоретико-множественных отношений.

Формальный характер модели обеспечивает возможность определения в её терминах формализованной процедуры построения концептуальной модели фрагмента Веба для заданного регламентируемого тематического целевого множества, предполагающей программную реализацию большинства этапов.

Концептуальная модель, реализованная на реальных фрагментах Веба, позволила дать численные оценки для каждого из подмножеств и степени их связности, классификацию входящих в них сайтов, а также существенно уменьшить размерность самих моделей посредством исключения подмножеств, сайты которых бесперспективны с точки зрения влияния на связность фрагмента. В случае академического множества мощность исследуемого множества сайтов уменьшается в 22 раза, а количества связывающих их гиперссылок – в 17 раз; в случае университетского Веба – в 9 и 14 раз соответственно.

2. Показано, что теоретико-графовые модели и методы позволяют дать оценки связности фрагментов Веба и определить основные направления для постановки задач улучшения присутствия в Вебе тематических сообществ. Установлена невысокая степень связности целевых множеств академического и университетского Веба. Показано, что административный каркас играет системообразующую роль в организации академического Веба в целом, однако его возможности для улучшения связности использованы не более чем на 60%. Установлено существенное влияние сайтов-коммуникаторов на увеличение связности сайтов целевого множества для всех рассмотренных фрагментов Веба.

3. Исследование задач ранжирования сайтов на университетском и академическом российском Вебе показывает, что на результаты ранжирования в значительной степени влияет выбор единиц анализа, а не использование тех или иных поисковых систем для измерения индикаторов. Сопоставление результатов ранжирования по предложенному в диссертации алгоритму с результатами ранжирования по известному алгоритму *Cybermetrics Lab* показывают их сильную положительную статистическую взаимосвязь. Показано, что для случая российского университетского Веба предложенный алгоритм требует в два раза меньше исходных данных, чем алгоритм *Cybermetrics Lab*.

4. Поставлены и исследованы математические задачи, основанные на применении методов математического моделирования, разработка и решение которых ведут к улучшению присутствия в Вебе сообществ сайтов в смысле некоторых заданных критериев, и предложены алгоритмы их решения. Получены решения, позволяющие каждому участнику сообщества оценивать полезность его участия в нем. На практике решение указанных задач могут быть положены в основу

административных регламентов, направленных на улучшение присутствия в Вебе регламентируемых тематических сообществ сайтов.

5. Создана проблемно-ориентированная информационная система для вебометрических исследований, включающая в себя робот-сборщик информации, сканирующий исследуемые сайты российского Веба, и реляционную базу данных для хранения и обработки собираемой информации, в которой реализован набор операций, соответствующий основным задачам вебометрических исследований. Сформированы базы данных внешних гиперссылок для академического и университетского фрагментов Веба и фрагмента Веба органов власти Карелии, часть из которых доступна для пользователей Веба.

СПИСОК ТРУДОВ

Статьи, опубликованные в реферируемых журналах из Перечня ВАК:

1. Печников А.А. Математические модели размещения ссылок в локализованной системе Интернет-ресурсов // Системы управления и информационные технологии. – 2007. – №2(28). – С. 92-96.
2. Печников А.А. О некоторых подходах к моделированию кликообществ // Системы управления и информационные технологии. – 2008. – №3(33). – С. 15-18.
3. Печников А.А., Чуйко Ю.В. Исследование согласованного поведения малых Интернет-сообществ // Телекоммуникации. – 2008. – №10. – С. 8-12.
4. Печников А.А. Вебометрические исследования Web-сайтов университетов России // Информационные технологии. – 2008. – №11. – С. 74-78.
5. Мазалов В.В., Печников А.А. О рейтинге официальных сайтов научных учреждений Северо-запада России // Управление большими системами. – М.: ИПУ РАН. – 2009. – Вып. 24. – С.130-146.
6. Печников А.А., Луговая Н.Б., Чуйко Ю.В., Косинец И.Э. Разработка инструментов для вебометрических исследований гиперссылок научных сайтов // Вычислительные технологии. – 2009. – Том 14, №5. – С. 66-78.
7. Печников А.А., Луговая Н.Б., Чуйко Ю.В. О связности множества официальных сайтов РАН // Вопросы современной науки и практики. Университет им. В.И.Вернадского. – 2009. – № 12(26) – С. 154-158.
8. Мазалов В.В., Печников А.А., Фалько И.А. О построении рейтинга научных журналов // Управление большими системами. – М.: ИПУ РАН. – 2009. – № 27 – С. 47-52.
9. Воронин А.В., Печников А.А. Исследования сайтов органов власти Республики Карелия // Век качества: Связь, сертификация, управление, экономика. – 2010. – №3. – С. 2-4.

10. Мазалов В. В., Печников А. А., Чирков А. В., Чуйко Ю. В. Задача дележа затрат на создание веб-коммуникатора как кооперативная игра / Управление большими системами. Специальный выпуск 30.1 "Сетевые модели в управлении". - М.: ИПУ РАН. - 2010. - № 30.1. - С. 187-196.
11. Печников А.А. Модель университетского Веба / А.А. Печников // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2010. – №6. – С. 208 - 214.

Статьи в других научных изданиях:

12. Мазалов В.В., Печников А.А., Раковский И.В. Создание информационно-коммуникационной среды поддержки развития общественного участия и межмуниципальной кооперации // Петрозаводск: ЗАО «Копистар Оптима». – 2005. – 33 с.
13. Печников А.А. Задача рационального размещения ссылок в регламентируемой локализованной системе интернет-ресурсов // Методы математического моделирования и информационные технологии. Труды ИПМИ КарНЦ. – Петрозаводск. – 2006. – вып. 7. – С.176-182.
14. Печников А.А., Чуйко Ю.В. Математические модели согласованного поведения малых Интернет-сообществ // Интернет-математика. – Екатеринбург: Издательство Уральского университета. – 2007. – С. 164-170.
15. Печников А.А., Илюкевич О.Г. Рейтинг официальных web-сайтов университетов России и Финляндии: сравнительный анализ // Информационные ресурсы России. 2008. – № 3 (103) – С. 25-28.
16. Печников А.А., Луговая Н.Б. Являются ли сайты конференций RCDL научными веб-коммуникаторами? // Тр. XI Всероссийской научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Петрозаводск. – 2009. – С. 329-332.
17. Печников А.А. Об исследованиях российского научного Веба // Материалы Всеросс. конф. с международным участием «Знания – Онтологии – Теории» (ЗОНТ-09, 22-24 октября 2009 г., Новосибирск). – Новосибирск: Институт математики им. С.Л. Соболева. – 2009. – Т.2. – С. 197-201.
18. Печников А.А. О некоторых подходах к исследованию российского научного Веба // Теория активных систем – 2009. Тр. междунар. научно-практ. конф. (17-19 ноября 2009 г., Москва). Том II. – 2009. – С. 140-144.
19. Чуйко Ю.В., Печников А.А. Исследование связности российского научного Веба / Ю.В. Чуйко, А.А. Печников // Когнитивный анализ и управление развитием ситуаций (CASC'2009). Тр. Междунар. конф. (17-19 ноября 2009 г., Москва). – С.283-286.
20. Печников А.А. О некоторых результатах вебометрических исследований университетского Веба // IV Междунар. научно-практ. конф.

«Современные информационные технологии и ИТ-образование». Сборник избранных трудов. – М.: ИНТУИТ.РУ. – 2009. – С. 530-537.

Тезисы докладов на научных конференциях:

21. Осолоткина Е.Ю., Печников А.А. Экспертные оценки предметной области в INTERNET // Университеты в образовательном пространстве региона: опыт, традиции и инновации. Тез. докладов научно-методической конф. (Петрозаводск, 18-20 мая 1999 г.). – 1999. – Часть 1. – С. 40.

22. Осолоткина Е.Ю., Печников А.А. Экологический аспект в Internet-моделях муниципальных образований // Математические методы в экологии. Тез. докладов Всеросс. научной школы (Петрозаводск, 10-16 июня 2001 г.). – 2001. – С.186.

23. Печников А.А. Вэбметрика интернет-ресурсов Карельского научного центра РАН // Северная Европа в XXI веке: природа, культура, экономика. Материалы Междунар. конф., посвященной 60-летию КарНЦ РАН (24-27 октября 2006 г.). – Том 2. – Петрозаводск: КарНЦ РАН. – 2006. С. 146-147.

24. Печников А.А. Моделирование клик-сообществ // Обзорение прикладной и промышленной математики. – 2008. – том 15, вып. 3. – С. 566-567.

25. Луговая Н.Б., Печников А.А. Измерения научных сайтов // Телематика'2008. Тр. XV Всероссийской научно-методической конф., Санкт-Петербург, 23-26 июня 2008. – Том 1 – С.166.

26. Луговая Н.Б., Печников А.А. Исследование экологических сайтов // Математические методы в экологии. Тез. докладов Третьей Всероссийской школы молодых ученых (Петрозаводск, 24-29 августа 2008 г.). 2009. – С.106.

27. Печников А.А. О некоторых результатах вебметрических исследований университетских сайтов // Информационная среда вуза XXI века: материалы II Всероссийской научно-практической конф., Петрозаводск, 15-18 сентября 2008. – С. 122-123.

28. Косинец И.Э., Луговая Н.Б., Печников А.А. Об одном подходе к исследованию ресурсов российского сегмента сети Интернет // Информационная среда вуза XXI века: материалы II Всероссийской научно-практической конф., Петрозаводск, 15-18 сентября 2008. – С. 15-18.

29. Печников А.А., Луговая Н.Б. Структурные исследования научного Веба // Телематика'2009. Тр. XVI Всероссийской научно-методической конф. (22-25 июня 2009 г., С-Петербург), Том 2. – 2009. – С.340.

30. Луговая Н.Б., Печников А.А. Разработка инструментов для вебметрических исследований // Телематика'2009. Тр. XVI Всероссийской научно-метод. конф. (22-25 июня 2009 г., С-Петербург), Том 2. – С.341.

31. Косинец И.Э., Печников А.А. О подходе к классификации внешних гиперссылок // Информационная среда вуза XXI века: материалы III Всероссийской научно-практической конф. (21-25 сентября 2009 г.). – Петрозаводск. – 2009. – С. 121-124.
32. Печников А.А. О связях между сайтами Российской академии наук и учреждений высшего образования // Информационная среда вуза XXI века: материалы III Всероссийской научно-практической конф. (21-25 сентября 2009 г.). – Петрозаводск. – 2009. – С. 181-182.
33. Печников А.А. О веб-коммуникациях между официальными сайтами вузов Карелии // Университеты в образовательном пространстве региона: опыт, традиции и инновации. Материалы научно-методической конф. (Петрозаводск, 16-17 февраля 2010 г.), Часть II. – С. 81-83.
34. Мазалов В.В., Печников А.А. Задача дележа затрат на создание веб-коммуникатора как кооперативная игра // VIII Всероссийская школа-семинар «Прикладные проблемы управления макросистемами». Материалы докладов (Апатиты, 29 марта-2 апреля 2010 г.) – С.37-38.
35. Косинец И.Э., Печников А.А. Об анализе веб-ресурсов органов власти // Информационная среда вуза XXI века: материалы IV Всеросс. научно-практ. конф. (20-24 сентября 2010 г.). Петрозаводск, 2010. – С. 131-133.
36. Печников А.А. Концептуальная модель фрагмента Веба и примеры ее реализации // Информационная среда вуза XXI века: материалы IV Всеросс. научно-практ. конф. (20-24 сентября 2010 г.). Петрозаводск, 2010. – С. 172-173.

Веб-ресурсы

37. Вебометрика. Институт прикладных математических исследований КарНЦ РАН [Электронный ресурс]. – Режим доступа: <http://webometrics.krc.karelia.ru>.

Свидетельства о регистрации программ:

38. Свидетельство о государственной регистрации программы для ЭВМ «Информационная система для вебометрических исследований» Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ № 2010610941 от 29 января 2010 г. Авторы: Печников А.А., Луговая Н.Б.

Автореферат отпечатан с разрешения диссертационного совета
Д 212.190.03 (протокол № 60/2 от «31» января 2011 г.)

Заказ № _____. Тираж 100 экз.

Карельский научный центр РАН
Редакционно-издательский отдел
185003, г. Петрозаводск, пр. А. Невского, 50.