

УДК 004.738.5:51-7

МОДЕЛЬ УНИВЕРСИТЕТСКОГО ВЕБА

© 2010 г.

А.А. Печников

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск

pechnikov@krc.karelia.ru

Поступила в редакцию 17.09.2010

В качестве исследуемого фрагмента веба рассматривается множество официальных сайтов классических университетов России и всех остальных сайтов веба, на которые существуют гиперссылки с университетских сайтов. Исследования гиперссылок позволяют предложить модель фрагмента веба, основанную на нескольких структурно-составляющих подмножествах сайтов и ссылках между ними.

Ключевые слова: веб, вебометрика, гиперссылка, сайт, множество сайтов университетов, модель университетского веба.

Введение

К актуальным направлениям вебометрики [1] – одного из развивающихся направлений информатики – относятся исследования гиперссылок. Практическая применимость этих исследований успешно демонстрируется реализацией алгоритмов информационного поиска таких популярных систем, как Google и Яндекс [2, 3]. Теоретические исследования показывают, что изучение гиперссылок имеет достаточный потенциал в смысле как новых источников информации и коммуникации, так и ценности самих веб-страниц [4–7].

Многие исследователи отмечают одновременное наличие в вебе как хаоса, так и порядка, при этом если хаос носит разносторонний и всеобъемлющий характер, то признаки порядка проявляются на некоторых его фрагментах. Если к этому добавить перманентную ограниченность исследовательских ресурсов, то следствием из утверждения о хаосе и порядке является концентрация внимания исследователей на достаточно узких фрагментах веба, таких, например, как множество сайтов университетов Великобритании, с расчетом последующего переноса полученных результатов на более общие случаи. В работах [5, 7] описаны методы и программные средства, разработанные применительно к таким исследованиям. Серия работ, выполненных с использованием этих методов и программных средств, посвящена построению и анализу университетского веб-графа Великобритании [8–10]. Построена модель, названная моделью «короны» ('corona' graph model), основанная на анализе ссылочного взаимодействия

сайтов на уровне подсайтов определенной глубины вложенности (в 109 университетских сайтах выделено 7669 таких подсайтов). Показано, что существует сильно связанная компонента, содержащая 25% таких подсайтов, далее, в терминах модели «галстука-бабочки» [11], маленькое левое крыло (7–8%), весьма большое правое крыло бабочки (35–40%) и остальные подсайты – это «отростки».

В данной работе в качестве фрагмента веба рассматривается множество официальных сайтов классических университетов России и всех остальных сайтов веба, на которые существуют гиперссылки с университетских сайтов (такой фрагмент веба мы будем называть университетским вебом). Известно, что официальные сайты российских университетов занимают достаточно скромные места в рейтинге Webometrics Ranking of World Universities [12]. Например, официальный сайт Московского государственного университета находится в третьей сотне и ещё лишь три сайта российских вузов попали в первую тысячу Top 8000 Universities. Одним из индикаторов ранжирования, используемых в [12], является количество внешних гиперссылок, сделанных на университетские сайты с других сайтов веба. Результаты исследований, изложенные в данной статье, позволяют частично объяснить такую ситуацию слабой связанностью как университетских сайтов посредством гиперссылок между собой, так и между университетскими сайтами и наиболее значимыми сайтами университетского веба. На сегодня автору неизвестны открытые публикации, посвященные исследованиям университетского веба в изложенной постановке.

Программные средства и методология

В работе использована методология, ранее разработанная применительно к исследованиям академического веба [13]¹. Вначале определяется некоторое начальное множество сайтов, объединяемых по заданному признаку (в нашем случае – официальные сайты классических университетов России). Далее из него выделяется целевое множество исследований, включающее в себя только те сайты, которые могут быть отсканированы с помощью поискового робота (это не все сайты начального множества, но, как правило, большая его часть).

Полученная база данных гиперссылок позволяет сформировать так называемое сопутствующее множество, – это все сайты, на которые существует хотя бы одна гиперссылка с сайтов целевого множества. В сопутствующем множестве выделяется множество так называемых ближайших веб-окрестностей. Ближайшей веб-окрестностью (или просто окрестностью) официального сайта называется множество сайтов организации – владельца официального сайта (кроме самого официального сайта), на которые существуют гиперссылки с официального сайта. Исключим из сопутствующего множества множество ближайших окрестностей и полученное в результате этой операции множество разобьем на два подмножества. Первое из них назовем множеством веб-коммуникаторов, т.е. сайтов, на которые сделано достаточно большое количество ссылок с сайтов целевого множества (пороговое значение для «достаточно большого количества» существенно зависит от целевого множества и ряда других факторов, в случае университетских сайтов оно принято равным 7). Обязательным свойством веб-коммуникатора является наличие хотя бы одной гиперссылки, сделанной с него на сайты целевого множества. Второе подмножество, включающее все оставшиеся сайты, называется множеством-оболочкой.

Таким образом, весь университетский веб можно представить как объединение четырех множеств: целевого множества, множества веб-окрестностей, множества веб-коммуникаторов и множества-оболочки. Каждое из перечисленных множеств исследуется более детально с целью понимания его структуры и взаимосвязей между его элементами. Выявление интегрированных связей между множествами позволяет построить обобщенную модель фрагмента веба.

Для получения, хранения и обработки веб-метрической информации использовался комплекс программ WebSciRes (от слов Webometrics, Science и Research), разработанный в Институте прикладных математических исследований КарНЦ РАН. В состав WebSciRes входят поисковый робот для сбора исходящих с сайтов гиперссылок LPR (Link, Page, Robot) и база данных, предназначенная для их хранения и обработки DB OL (Data Base of OutLinks) [14].

Целевое множество

Основными объектами исследования в данной работе являются официальные сайты классических университетов России. Для определения начального множества в качестве основного источника информации использовался федеральный портал «Российское образование» (<http://www.edu.ru>). Имеющийся на портале сервис «Расширенный поиск университетов, институтов, академий» по поисковой фразе «государственный университет» и опциям «Вид – университет» и «Организационно-правовая форма – Государственный вуз» позволяет получить список, содержащий информацию о 141 вузе, из которых вручную можно выбрать информацию о классических университетах (включая и доменные имена их официальных сайтов). Не вдаваясь в дискуссию по поводу того, что такое «классический университет» (смотри, например, [15]), отметим лишь, что количество таких университетов в России около 80.

Сканирование всех сайтов классических университетов оказалось невозможным по таким причинам, как защищенность от сканирования, использование специальных технологий при разработке сайтов, таких как JavaScript и Adobe Flash, и ограниченность ресурсов исследователей и владельцев сайтов. Поэтому на сегодня целевое множество содержит 56 сайтов, для которых мы перечислим названия университетов-владельцев в алфавитном порядке: Адыгейский, Алтайский, Амурский, Астраханский, Башкирский, Белгородский, Братский, Брянский, Бурятский, Волгоградский, Воронежский, Горно-Алтайский, Дальневосточный, Ингушский, Иркутский, Кабардино-Балкарский, Казанский, Калмыцкий, Камчатский, Костромской, Кубанский, Магнитогорский, Марийский, Мордовский, Московский, Нижегородский, Новосибирский, Омский, Орловский, Пензенский, Пермский, Петрозаводский, Поморский, Самарский, Санкт-Петербургский, Северо-Осетинский, Сибирский федеральный, Смоленский, Ставропольский, Сургутский, Сыктывкарский,

¹ При поддержке Российского фонда фундаментальных исследований (грант № 08-07-00023а).

Тверской, Тольяттинский, Томский, Тульский, Тывинский, Тюменский, Ульяновский, Уральский, Хакаский, Череповецкий, Читинский, Чувашский, Югорский, Южно-Уральский и Якутский университеты.

Общее количество страниц, проанализированных роботом LPR, превышает 800000, на них обнаружено 985000 исходящих внешних гиперссылок, из которых более 40000 уникальных.

На различных страницах одного сайта могут встречаться гиперссылки на один и тот же внешний адрес, имеющие одинаковый контекст. Распространенный пример – ссылка на вышестоящую организацию или на сайт организации – разработчика официального сайта. Уникальной гиперссылкой называется гиперссылка из множества всех гиперссылок с одинаковым адресом и контекстом, которая находится на странице, имеющий максимальный уровень; при этом уровень начальной страницы сайта считается наивысшим. Далее мы будем рассматривать только уникальные гиперссылки, поэтому термин «уникальная» в большинстве случаев будем опускать.

Обозначим целевое множество буквой T , а множество уникальных гиперссылок, сделанных сайтами из T друг на друга, обозначим буквой E . Как уже известно, $|T| = 56$. Исследования показали, что $|E| = 460$. Таким образом, в среднем на каждый официальный университетский сайт приходится примерно по 8 уникальных гиперссылок, сделанных с сайтов других университетов. Однако распределение гиперссылок по сайтам весьма неравномерно.

Первая пятерка сайтов по числу гиперссылок, сделанных на другие сайты из T , выглядит так: Амурский госуниверситет (78 ссылок), Горно-Алтайский госуниверситет (69), Бурятский госуниверситет (58), Нижегородский госуниверситет (52) и Чувашский университет (38). В этом предложении и далее из экономии места для обозначения официального сайта мы используем название его университета-владельца, надеясь, что это не вызывает особой путаницы. 24 университетских сайта не имеют ни одной ссылки на другие университетские сайты.

Абсолютным лидером по ссылкам, сделанным на него с других сайтов из T , является сайт Новосибирского госуниверситета (84 ссылки). В первую пятерку входят также Южно-Уральский госуниверситет (25), Московский госуниверситет (24), Дальневосточный госуниверситет (19) и Томский госуниверситет (19). Половина сайтов университетов получила не более 5 ссылок от своих коллег.

В расстановке «встречных» гиперссылок никакой закономерности не наблюдается. К примеру, на сайт Нижегородского университета сделано лишь три гиперссылки из T , причем «встречные» ссылки обнаружены только между Нижегородским и Казанским университетами.

Целевое множество сайтов и множество связывающих их гиперссылок можно рассматривать как ориентированный граф без петель с кратными дугами (веб-граф) $G(T, E)$. Оказалось, что граф содержит 31 компоненту сильной связности, причем максимальная компонента $G(T^*, E^*)$ содержит 26 сайтов, а остальные 30 компонент содержат по одному сайту. Диаметр $diam(G(T^*, E^*)) = 6$.

T^* содержит официальные сайты Алтайского, Амурского, Астраханского, Бурятского, Волгоградского, Воронежского, Горно-Алтайского, Дальневосточного, Марийского, Мордовского, Нижегородского, Омского, Пензенского, Пермского, Петрозаводского, Поморского, Сибирского федерального, Сыктывкарского, Тверского, Томского, Тывинского, Ульяновского, Читинского, Чувашского, Югорского и Якутского университетов.

Ближайшие веб-окрестности

Понятие «ближайшая веб-окрестность официального сайта» весьма близко по смыслу к понятию «веб-ресурсы организации» хотя следует заметить, что и для этого понятия отсутствует точное определение (или, что будет точнее, оно во многом зависит от контекста исследований).

Обозначим сопутствующее множество буквой U , а $B_t \subset U$ – ближайшую веб-окрестность официального сайта $t \in T$, тогда множество всех ближайших окрестностей целевого множества $B = \cup B_t, \forall t \in T, B \subset U$.

Каждая ближайшая окрестность разбивается на два непересекающихся множества: прямую и скрытую окрестности $B_t = B_t^{dir} \cup B_t^{hid}$.

Прямая веб-окрестность официального сайта B_t^{dir} – это все сайты организации, имеющие доменные имена в зоне доменного имени данного сайта $t \in T$, на которые существуют ссылки с официального сайта (соответственно, $B^{dir} = \cup B_t^{dir}, \forall t \in T$).

К примеру, B_{UNN}^{dir} официального сайта Нижегородского университета содержит 39 сайтов, на которые сделано 126 уникальных ссылок с сайта www.unn.ru. К ним относятся такие сайты, как «Сетевое взаимодействие вузов по основным направлениям Болонского процесса»

(br.unn.ru, 2 уникальные ссылки), сайт Научно-исследовательского физико-технического института ННГУ (nifti.unn.ru, 3 ссылки), сайты кафедры информатики и автоматизации научных исследований (iani.unn.ru, 2 ссылки), сайты многих факультетов и подразделений университета.

Исследования показали, что для всего целевого множества T $|B^{dir}| = 1620$. На сайты из B^{dir} сделано 7550 ссылок с сайтов из множества T , что в дальнейшем будем записывать с использованием функции $Tlinkscout(X)$, значение которой равно количеству уникальных гиперссылок, сделанных со всех сайтов целевого множества T на сайты некоторого подмножества $X \subset U$. Для данного случая $Tlinkscout(B^{dir})=7550$.

Скрытая веб-окрестность официального сайта B_i^{hid} – это все сайты с уникальными доменными именами, которые на официальном сайте организации $t \in T$ позиционированы тем или иным способом как ее веб-ресурсы и на которые существуют гиперссылки с этого сайта, однако при этом они имеют доменные имена в других доменных зонах. Соответственно, $B^{hid} = \cup B_t^{hid}, \forall t \in T$.

Поиск скрытых веб-окрестностей является нетривиальной задачей, и найти все веб-ресурсы университета, доменные имена которых не ассоциируются с именем официального сайта, зачастую очень непросто. Задача изначально существенно упрощается, если на первых страницах сайта есть ссылка на раздел сайта типа «Веб-ресурсы университета», что на сегодня встречается достаточно часто. В качестве примера можно привести сайты из B_{VSU}^{hid} Воронежского университета (доменное имя официального сайта – www.vsu.ru), к которым относится сайт кафедры криминалистики (vrncrime.ru, 1 ссылка), несколько персональных сайтов преподавателей университета.

Такой способ поиска позволил обнаружить, что $|B^{hid}|=150$, $Tlinkscout(B^{hid})=1070$.

Таким образом, обнаружено, что $|B|=1770$ и $Tlinkscout(B)=8620$.

Веб-коммуникаторы

Множество веб-коммуникаторов $K \subset U$ разбивается на три подмножества $K=K^{ind} \cup K^{col} \cup K^{med}$:

- подмножество сайтов-индукторов K^{ind} («мало входящих ссылок, много исходящих») – индуктор ссылается на 7 и более сайтов целевого множества;

- подмножество сайтов-коллекторов K^{col} («много входящих, мало исходящих») – на кол-

лектор ссылаются 7 и более сайтов целевого множества;

- подмножество сайтов-посредников K^{med} («много входящих, много исходящих») – на посредника ссылаются не менее 7 сайтов, а он, в свою очередь, ссылается не менее, чем на 7 сайтов целевого множества.

Проведенные исследования позволили обнаружить 3 сайта-индуктора, 15 коллекторов и 10 посредников.

К индукторам относятся следующие сайты:

- сайт «Междисциплинарный научный сервер» (www.scientific.ru, на который сделано 10 ссылок с 5 сайтов из T , а он имеет 59 ссылок на 25 университетских сайтов);

- сайт проекта «Рутения» (www.ruthenia.ru, 4/4, 63/37);

- сайт отделения ГПНТБ СО РАН (www.prometeus.nsc.ru, 5/4, 27/7).

Примерами коллекторов являются сайт РФФИ (www.rfbr.ru, 153/37, 1/1) и сайт Научной электронной библиотеки (www.elibrary.ru, 37/22, 1/1).

Сайт «Российское образование. Федеральный образовательный портал» представляется наилучшим посредником (www.edu.ru, 255/44, 48/40).

Общее количество веб-коммуникаторов $|K|=28$ и $Tlinkscout(K) = 1170$, в том числе $Tlinkscout(K^{ind}) = 20$, $Tlinkscout(K^{col}) = 790$ и $Tlinkscout(K^{med}) = 360$.

Множество-оболочка

Множество-оболочка $S \subset U$ также разбивается на три подмножества, которые называются насыщенной, вязкой и разреженной оболочками.

Насыщенной оболочкой S^{rich} называется множество сайтов U , не входящих в целевое множество, на которые сделаны ссылки не менее чем с 7 сайтов из T , из которого удалены все сайты, входящие в множество веб-коммуникаторов K . Заметим, что насыщенная оболочка содержит 8 сайтов поисковых систем и 3 сайта интернет-статистики. Основные характеристики S^{rich} : $|S^{rich}| = 79$, $Tlinkscout(S^{rich}) = 2080$, то есть на один сайт из насыщенной оболочки в среднем приходится почти 26 гиперссылок с сайтов из T , что демонстрирует большой интерес университетских сайтов к сайтам насыщенной оболочки, не встречающийся, однако, большого взаимного интереса. К примеру, на сайт Рособразования (www.ed.gov.ru) сделано 68 ссылок с 28 университетских сайтов, а на сайты из T с сайта Рособразования – 5 ссылок на 5 сайтов.

Вязкая оболочка S^{swamp} аналогична насыщенной оболочке, но на ее сайты сделаны ссылки не менее чем с 4 и не более чем с 6 сайтов из T . Обнаружено, что $|S^{swamp}| = 245$ и $Tlinkscout(S^{swamp}) = 2140$.

Разреженная оболочка S^{vac} содержит сайты, на которые сделаны гиперссылки менее чем с 4 сайтов из T . Это самое большое множество из всех рассматриваемых, $|S^{vac}| = 14400$ и $Tlinkscout(S^{vac}) = 25800$. Понятно, что разреженная оболочка представляет наименьший интерес, например, для исследования вопросов связности веб-графа всего фрагмента веба.

Обсуждение полученных результатов

Таким образом, построена модель университетского веба, базирующаяся на 9 выявленных множествах сайтов:

- целевое множество официальных университетских сайтов T : $|T| = 56$, $Tlinkscout(T) = 460$ (1.14 % от всех уникальных гиперссылок);
- множество прямых ближайших окрестностей B^{dir} : $|B^{dir}| = 1620$, $Tlinkscout(B^{dir}) = 7550$ (18.7% ссылок),
- множество скрытых ближайших окрестностей B^{hid} : $|B^{hid}| = 150$, $Tlinkscout(B^{hid}) = 1070$ (2.7% ссылок),
- множество индукторов K^{ind} : $|K^{ind}| = 3$, $Tlinkscout(K^{ind}) = 20$ (0.05% ссылок),
- множество коллекторов K^{col} : $|K^{col}| = 15$, $Tlinkscout(K^{col}) = 790$ (1.96% ссылок),
- множество посредников K^{med} : $|K^{med}| = 10$, $Tlinkscout(K^{med}) = 360$ (0.9% ссылок),
- множество – насыщенная оболочка S^{rich} : $|S^{rich}| = 79$, $Tlinkscout(S^{rich}) = 2080$ (5.2% ссылок),
- множество – вязкая оболочка S^{swamp} : $|S^{swamp}| = 245$, $Tlinkscout(S^{swamp}) = 2140$ (5.3% ссылок),
- множество – разреженная оболочка S^{vac} : $|S^{vac}| = 14400$, $Tlinkscout(S^{vac}) = 25800$ (64.1% ссылок).

Анализ модели в целом показывает очень слабую связность университетского веба. Заметим, что более 70% гиперссылок с сайтов целевого множества сделаны на сайты множества оболочки, с которых отсутствуют ответные ссылки. Содержательно это означает, что университеты демонстрируют в вебе большой интерес к сайтам, для владельцев которых сами они не представляют интереса (или, по крайней мере, этот интерес не находит отражения в вебе).

Количество гиперссылок между сайтами целевого множества составляет чуть более 1% от

всех исходящих гиперссылок, а компонента сильной связности подграфа, построенного на целевом множестве, содержит менее половины официальных сайтов университетов. Это свидетельствует о том, что реальные связи между университетами не находят своего отражения в вебе.

Множество ближайших окрестностей имеет большое количество гиперссылок, сделанных на его сайты с сайтов целевого множества, однако обратные ссылки обнаруживаются далеко не всегда. Содержательно это соответствует тому факту, что университетский сайт содержит гиперссылки на сайты факультетов, кафедр, подразделений, однако отсутствуют ссылки с этих сайтов на головной сайт организации.

Сайты-коммуникаторы, несомненно, оказывают большое влияние на связность университетского веба. Однако результаты исследований демонстрируют их весьма незначительное количество, что также свидетельствует о слабом отражении реальных коммуникаций между университетами в вебе. В частности, в качестве сайтов-коммуникаторов не обнаружено ни одного сайта всероссийских и международных конференций, проводимых на базе классических университетов России, хотя в них принимают участие значительное количество преподавателей, что могло бы фиксироваться в виде гиперссылок на их основные места работы. И наоборот, страницы преподавателей (или хотя бы кафедр) должны содержать гиперссылки на сайты конференций, в которых они принимали участие.

Полученные результаты позволяют предложить ряд рекомендаций для владельцев и разработчиков университетских сайтов, реализация которых направлена на улучшение представительства университетов в вебе и, в частности, позиций в рейтингах университетских сайтов:

1) обратить внимание на желательность создания гиперссылок на официальных сайтах университетов на сайты других университетов в случае наличия соответствующих информационных поводов (проведение совместных проектов, мероприятий, наличие важных документов и электронных ресурсов и т.д.);

2) включить в регламенты университетских веб-ресурсов положение о том, что на самостоятельных сайтах подразделений обязательно должна создаваться ссылка на головной университетский сайт;

3) обратить внимание на желательность создания гиперссылок на официальных сайтах университетов на наиболее популярные сайты-коммуникаторы при наличии информационных поводов;

4) обратить внимание на желательность создания гиперссылок на официальных сайтах конференций на сайты университетов, представители которых участвуют в конференциях (или хотя бы входят в организационные и программные комитеты), и наоборот.

Заключение и возможные направления исследований

Построенная модель университетского веба демонстрирует возможность описания достаточно обширного фрагмента веба минимальным количеством понятий, позволяет сделать ряд предложений, направленных на улучшение присутствия в вебе университетских сайтов, и определяет большое количество направлений для дальнейших исследований.

Очевидно, что, несмотря на свои небольшие размеры и малое количество «внутренних» гиперссылок, большое значение в модели имеет множество T . Для объяснения полученных результатов по компоненте сильной связности веб-графа $G(T, E)$ требуется провести качественные исследования гиперссылок, ориентированные на их классификацию по контексту, и ответить на вопрос, заданный в работе [7]: «Что здесь делает эта ссылка?».

Исходя из построенной модели университетского веба, можно определить и другие направления дальнейших исследований. По-видимому, к первоочередным можно отнести исследования влияния веб-коммуникаторов на теоретико-графовые характеристики веб-графа, множеством вершин которого является объединение $T \cup K$, а множеством дуг – объединение множества E и всех дуг, сделанных с множества K на множество T . Основной вопрос можно сформулировать так: насколько велико влияние коммуникаторов на рост максимальной компоненты связности и изменение (уменьшение) ее диаметра?

Представляет интерес дальнейшая детализация взаимодействия целевого множества и ближайшего окружения, а также классификация по типам сайтов оболочки.

Список литературы

1. Almind T., Ingwersen P. Informetric analyses on the World Wide Web: Methodological approaches to «webometrics» // *Journal of Documentation*. 1997. № 53 (4). P. 404–426.
2. Brin S., Page L. The anatomy of a large scale hypertextual web search engine // *Computer Networks and ISDN Systems*. 1998. № 30 (1–7). P. 107–117.

3. Индекс цитирования. [Электронный ресурс] – Режим доступа: <http://help.yandex.ru/catalogue/?id=873431>.

4. Cronin B., Snyder H.W., Rosenbaum H. et al. Invoked on the web // *Journal of the American Society for Information Science and Technology*. 1998. № 49 (14). P. 1319–1328.

5. Thelwall M. Extracting macroscopic information from web links // *Journal of the American Society for Information Science and Technology*. 2001. № 52 (13). P. 1157–1168.

6. Flake G.W., Lawrence S., Giles C.L., Coetzee F.M. Self-organization and identification of web communities // *IEEE Computer*. 2002. № 35. P. 66–71.

7. Thelwall M. What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation // *Information Research*. Vol. 8. № 3, April 2003. [Электронный ресурс] – Режим доступа: <http://informationr.net/ir/8-3/paper151.html>.

8. Björneborn L. Small-world link structures across an academic web space: a library and information science approach. PhD dissertation. Royal School of Library and Information Science. 2004. 399 p. [Электронный ресурс] – Режим доступа: <http://vip.db.dk/lb/phd/phd-thesis.pdf>.

9. Björneborn L. Small-world connectors across academic web spaces // *Proceedings of the AoIR-ASIST 2004 Workshop on Web Science Research Methods*. 5th Annual Conference of the Association of Internet Researchers, Brighton, UK. Invited keynote presentation. – 19 September 2004 [Электронный ресурс] – Режим доступа: <http://cybermetrics.wlv.ac.uk/AoIRASIST/bjorneborn.html>.

10. Björneborn L. Hvor er verden dog lille - også på webben [It's a small world - also on the Web] // *Bibliotekspresen*. 2004. № 17. P. 494–495.

11. Broder A., Kumar A., et al. Graph structure in the web // *Journal of Computer Networks*. 2000. № 33(1–6). P. 309–320.

12. Ranking Web of World Universities [Электронный ресурс]. – Режим доступа: <http://www.webometrics.info>.

13. Печников А.А., Луговая Н.Б., Чуйко Ю.В., Косинец И.Э. Вебметрические исследования внешних связей научных сайтов // *Вычислительные технологии*. 2009. Том 14, № 5. С. 66–78.

14. Луговая Н.Б., Печников А.А. Разработка инструментов для вебметрических исследований // «Телематика'2009». Труды XVI Всероссийской научно-методической конференции (22–25 июня 2009 г., С.-Петербург). Том 2. С. 341.

15. Щипунов А.А., Башкатов В.З., Воробейчиков Э.С., Хасанов В.Я. Классический университет и глобальная информационная структура // *Вестник Томского государственного университета*. 2000. № 269. С. 126–130.

UNIVERSITY WEBSITE MODEL*А.А. Печников*

We investigate a fragment of the Web that includes a set of official sites of Russian classical universities and all other sites of the Web, for which there are hyperlinks from the university sites. Hyperlink studies allow one to propose a Web fragment model based on several structurally united subsets of sites and hyperlinks between them.

Keywords: Web, webometrics, hyperlink, site, set of university sites, university website model.