

УДК 004.942, 004.032.26

*А.В. Мансуров, А.С. Шатохин***Моделирование перспективной динамики
роста рейтинга Webometrics на примере
Алтайского государственного университета***A.V. Mansurov, A.S. Shatokhin***Simulation of Potential Dynamics of the Webometrics
Ranking on an Example of the Altai State University**

В данной работе проводится анализ технологии формирования рейтинга Webometrics, выполняется численное моделирование перспективной динамики роста рейтинга на примере Алтайского государственного университета с использованием искусственных нейронных сетей. На основании полученных результатов предлагается ряд практических рекомендаций, способствующих росту рейтинговой оценки вуза в глобальном рейтинге Webometrics.

Ключевые слова: рейтинг Webometrics, рейтинги высших учебных заведений, интернет-ресурсы учебных заведений, оценка связей.

Введение. Один из важных компонентов современного рынка образовательных услуг – составление глобальных рейтингов высших учебных заведений (вузов) мира [1–2]. Популярный в настоящее время рейтинг Webometrics (<http://www.webometrics.info/>) характеризует деятельность вузов в сети Интернет и является важной оценкой процесса информатизации образовательной и научно-исследовательской деятельности вуза, а также развития и функционирования Интернет-представительства вуза в сети [3]. Достаточно слабые позиции у российских вузов (по данным рейтинга Webometrics (январь 2012 г.), самый успешный российский вуз занимает 147-е место, всего 10 вузов имеют мировой рейтинг меньше 1000) служат поводом для анализа методики составления рейтинговой оценки и поиска шагов, способствующих росту рейтинговой оценки вуза.

В данной работе выполняется анализ методологии формирования рейтинга Webometrics и проводится численное моделирование перспективной динамики роста рейтинга на примере Алтайского государственного университета (по данным рейтинга на январь 2012 г. университет занимает 1969-е место в мире и 37-е место среди российских вузов) для относительного позиционирования в первой десятке вузов России, определяются необходимые для этого технические и организационные мероприятия.

1. Методика формирования рейтинговой оценки Webometrics. Рейтинговая оценка Webomet-

The paper analyses the methodology and numerical modeling webometrics ranking. The Neural Network approach is used to simulate the possible rise of the Altai State University Webometrics ranking. In accordance with the obtained results the researchers propose practical suggestions to raise the university's position to a much higher level in the Webometrics ranking.

Key words: Webometrics ranking, university rankings, academic websites, link analysis.

rics вычисляется с помощью некоторой свертки на основе нескольких показателей, которые входят в итоговую оценку с различными весовыми коэффициентами, задающими важность параметров [3]. Показатели формируются в виде промежуточной рейтинговой оценки (максимум соответствует первому месту в рейтинге) по конкретному критерию для каждого учебного заведения.

Показатель Visibility (V) – 50% представляет так называемый **импакт-фактор**, отвечающий за популярность информационного источника в сети Интернет. Он отражает количество уникальных внешних ссылок на веб-ресурсы учебного заведения по внутренним данным системы Majestic SEO (<http://www.majesticseo.com/>) на основе информации об обратных ссылках (Backlinks) и ссылающихся на исследуемый домен сторонних доменов (Referring Domains), а также данных о трафике, которые Majestic SEO получает от системы Alexa Internet (<http://www.alexa.com/>). Показатели Size (S) – 10%, Rich Files (R) – 10% и Scholar (Sc) – 30% представляют групповой показатель **активности**, отражающий в данном случае активность работы информационного представительства вузов в сети Интернет в виде количества страниц веб-ресурсов вуза, проиндексированных поисковой системой Google. В групповой показатель активности включается и научная активность вузов (рейтинг Scimago SIR), в том числе ее проявление в виде различных элек-

тронных документов и публикаций, доступных в сети Интернет. Под информационным представителем понимаются все веб-ресурсы вуза, располагающиеся в рамках одного домена (все субдомены основного домена) вуза.

Согласно публикациям авторов данной методики [4–5], итоговый рейтинг $RANK_j$ для каждого университета определяется путем линейного агрегирования основных рейтинговых показателей в соответствии с выражением:

$$RANK_j \sim Q_j = W_V * V_j + W_S * S_j + W_R * R_j + W_{SC} * Sc_j, \quad (1)$$

где весовые коэффициенты W_V , W_S , W_R , W_{SC} соответствуют проценту важности рейтинговых показателей; параметр j задает порядковый (не рейтинговый) номер каждого университета в процессе обработки данных.

Формирование и вычисление рейтингов происходит два раза в год – в январе и в июле. На веб-сайте проекта Webometrics Ranking (<http://www.webometrics.info/>) доступен полный массив данных параметров $RANK_j$, S_j , V_j , R_j и Sc_j за январь 2012 г. (на момент написания публикации более старые данные в полном объеме недоступны) для всех вузов, находящихся в базе данных проекта, т.е. более 20000 вузов мира.

2. Исследование методики и моделирование перспективной динамики роста рейтинга. Исследование массива данных рейтинговых показателей S_j , V_j , R_j , Sc_j позволяет обнаружить разные области определения показателей. Диапазон изменений значений S_j , V_j , R_j находится в одинаковых границах (от 1 до 18742 ... 20279), а значения показателя Sc_j изменяются в интервале [1; 9576]. Либо формирование этого рейтингового показателя происходит по методике, отличной от заявленной [4–5], либо поведение исходных данных не позволяет сформировать ранжирование, аналогичное остальным показателям. Также можно предполагать, что чувствительность системы к изменению параметра Sc_j будет примерно в два раза выше, чем к остальным параметрам.

Слабое влияние параметров S_j и R_j (их вес в формировании итогового рейтинга составляет 10%) позволяет сделать предположение о том, что они могут быть существенны лишь для областей высокого рейтинга (малых значений $RANK_j$), а в области низкого рейтинга серьезного влияния на итоговый показатель могут не оказывать.

Для изучения зависимости $RANK_j = Rn(S_j, V_j, R_j, Sc_j)$ и установления характера влияния рейтинговых показателей S_j , V_j , R_j , Sc_j на $RANK_j$ проводится численный модельный эксперимент. Целевой вуз для численного эксперимента – Алтайский государст-

венный университет (домен asu.ru; рейтинг на январь 2012 г.: $RANK = 1969$; $S = 2136$; $V = 3509$; $R = 1145$; $Sc = 1975$). Для упрощения эксперимента предполагается, что показатели всех остальных вузов будут оставаться неизменными на момент вычисления следующих рейтингов Webometrics в ближайшем будущем. Тогда необходимым условием попадания в число первых 10 российских вузов является получение оценки $RANK < 1000$.

Рабочая выборка численного эксперимента формируется с учетом текущей рейтинговой позиции целевого вуза. Она включает в себя полный набор данных для рейтинга $RANK_j$ от 1 до 2500. Дальнейшие показатели отбираются равномерно группами по 200 значений с величиной шага между группами, равной 500. Формирование рабочей выборки заканчивается при достижении $RANK_j > 6000$ ввиду отсутствия дальнейшей целесообразности.

Применение выражения (1) для расчета итоговых значений $RANK_j$ на основе рейтинговых показателей методики обнаруживает явную неоднозначность и невозможность построения итогового рейтинга по вычисляемым значениям Q_j . Вероятно, формирование итогового рейтинга происходит в соответствии с более сложной процедурой, чем об этом заявляют авторы методики, а значит, использование (1) для моделирования динамики $RANK_j$ не представляется возможным и необходим другой подход. В этом случае процесс формирования итогового рейтинга можно представлять как некоторый «черный ящик», получающий на входе данные и формирующий выходной результат. Для исследования работы подобных систем часто применяется подход с использованием искусственных нейронных сетей (ИНС), имитирующих работу подобного «черного ящика».

Моделирование зависимости $RANK_j = Rn(S_j, V_j, R_j, Sc_j)$ выполняется с помощью ИНС пакета Neural Network Toolbox в среде MATLAB [6]. Используется ИНС типа «многослойный перцептрон» с двумя внутренними слоями и функцией активации «тангенциальная сигмоида» [7]. Обучение выполняется методом обратного распространения ошибки с градиентной оптимизацией Левенберга-Марквардта. Обучающая выборка генерируется из сформированной ранее рабочей выборки путем случайного отбора 60% данных, равномерно распределенных по всей области рабочей выборки. Оставшиеся 40% используются в процессе обучения для тестирования работы ИНС.

На входы ИНС подаются значения S_j , V_j , R_j , Sc_j , на выходе ИНС выдает значения итогового рейтинга $RANK_j$. Входные и выходные параметры автоматически перенормируются в рабочий диапазон ИНС [-1; 1]. Точность (или ошибка) восстановления $RANK_j$ по входным данным составляет примерно ± 30 , что в среднем соответствует погрешности от 1 до 3%.

Результаты моделирования с помощью ИНС динамики итогового рейтинга $RANK_j$ целевого вуза при независимой вариации каждого рейтингового показателя приведены на рисунке 1.

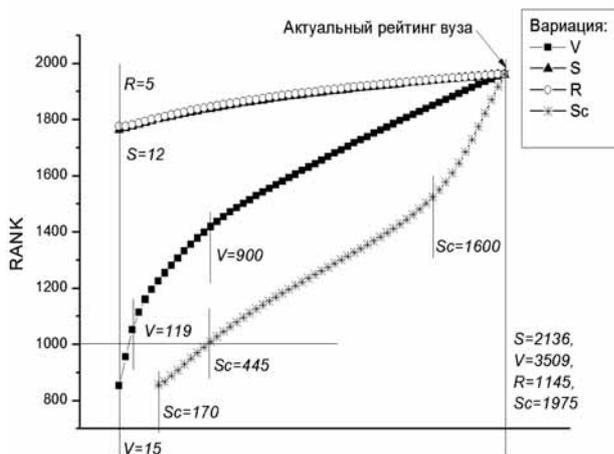


Рис. 1. Моделирование динамики итогового рейтинга целевого вуза при независимой вариации рейтинговых показателей

Можно отметить, что влияние показателей S и R на формирование итогового результата является незначительным, что согласуется с заявленной в методике важностью этих параметров. Их влияние должно быть существенно в области высоких рейтинговых оценок (при малых $RANK$). Вариация более весомых показателей V и Sc оказывает на итоговый рейтинг более заметный результат и позволяет выполнить условие $RANK < 1000$. Однако для получения высокого итогового рейтинга $RANK < 1000$ необходимо получить $V < 119$ и $Sc < 445$, что является практически неосуществимым условием для целевого вуза в виду его размеров, структуры и степени вовлеченности в мировую научную жизнь.

Вариация показателя V демонстрирует две области с различным поведением. Зависимость $RANK(V)$ для $V \in (900; 3509]$ практически линейная с малой скоростью изменения $RANK$ от V : $\Delta RANK = 588$ при изменении рейтингового показателя $\Delta V = 2609$ пунктов. Во второй области для $V \in [15; 900]$ зависимость становится более крутой и нелинейной, и для $\Delta V = 885$ изменение $\Delta RANK = 588$. Можно говорить о том, что в течение длительного времени рост показателя V будет вызывать незначительные изменения итогового рейтинга $RANK$, пока V не достигнет определенного значения, после которого изменения $RANK$ станут ощутимыми. Также модельная система на базе ИНС обнаруживает повышенную чувствительность к показателю Sc , что согласуется со сделанным ранее в тексте работы предположением. Повышенная чувствительность модельной системы может объясняться особенностями перенормирования данных показателя Sc на

входах и выходах ИНС в рабочий диапазон $[-1; 1]$, область определения которого в 2 раза меньше областей определения других рейтинговых показателей.

Результаты моделирования динамики итогового рейтинга целевого вуза при совместной вариации рейтинговых показателей приведены на рисунке 2. Учитывая приоритетность показателей V и Sc (80% влияния на итоговый рейтинг), вариация S и R рассматривается совместно. Согласно полученным результатам, одновременная работа над корректировкой рейтинга по основным показателям V и Sc является очень эффективным шагом. Один из возможных вариантов выполнения условия $RANK < 1000$ реализуется при $V < 1769$, $Sc < 989$. При учете влияния всех рейтинговых показателей, возможные варианты удовлетворения условия $RANK < 1000$ реализуются немного раньше, например, при $V < 1949$, $Sc < 1091$, $S < 1200$, $R < 635$. Таким образом, для получения $RANK \leq 1000$ необходимо добиться изменения рейтинговых показателей: $\Delta V \geq 1560$, $\Delta Sc \geq 884$, $\Delta S \geq 936$, $\Delta R \geq 510$.

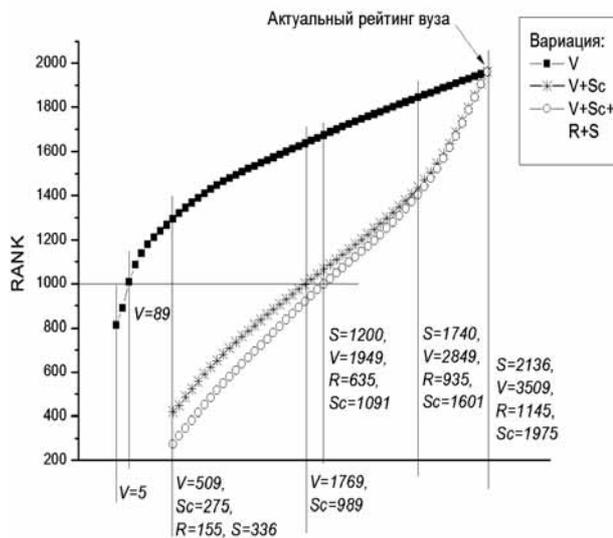


Рис. 2. Моделирование динамики итогового рейтинга целевого вуза при совместной вариации рейтинговых показателей

3. Мероприятия, направленные на рост рейтинговой оценки Webometrics. Частично предложения общего плана, способствующие росту рейтинга вуза, приведены авторами проекта Webometrics Ranking на веб-ресурсе проекта [8]. В соответствии со сделанными наблюдениями, основные усилия в первую очередь следует сосредоточить на корректировке рейтинговых показателей V и Sc . Учитывая комплексный характер планируемых мероприятий, действия обязательно затронут и оставшиеся показатели и R и S . Большую часть все-

го комплекса мероприятий составляют продвиженческие мероприятия, направленные на популяризацию веб-ресурсов вуза в сети Интернет:

1) перекрестный обмен ссылками на веб-ресурсы вуза с аналогичными веб-ресурсами местных, региональных, национальных и иностранных (с которыми установлены партнерские отношения) учебных заведений;

2) размещение ссылок на веб-ресурсы вуза на страницах местных, региональных и прочих Интернет-проектов – порталов, Интернет-СМИ, справочников, веб-ресурсов органов государственной власти и коммерческих структур (PR-материалы, специализированное размещение ссылок, приуроченное к некоторому мероприятию и т.п.), внесение адресов веб-ресурсов вуза в каталоги и справочники в сети Интернет;

3) представление (дублирование) контента веб-ресурсов вуза на английском и иных, специфичных для региона, где располагается вуз, языках;

4) создание на базе домена вуза и веб-проектов вуза научных порталов, специализированных проектов и персональных страниц преподавателей и сотрудников вуза, посвященных научной и исследовательской деятельности, а также общественной деятельности как вуза в целом, так и его отдельных групп сотрудников. Предоставление в открытом доступе электронных вариантов публикаций сотрудников вуза, курсов лекций и материалов практических занятий, создание открытой электронной библиотеки (репозитария);

5) стимулирование научной деятельности сотрудников вуза, развитие международного сотрудничества в области научной деятельности, рост вовлеченности сотрудников вуза в международные проекты, рост числа публикаций сотрудников вуза в ведущих мировых научных изданиях. Данный пункт направлен на удовлетворение условиям попадания в рейтинг Scimago SIR.

Эти инициативы позволят увеличить генерируемые системами Majestic SEO и Alexa информационные показатели об обратных ссылках (Backlinks)

и ссылающихся на исследуемый домен сторонних доменов (Referring Domains), способствовать росту показателя TrafficRank. Английский язык, являющийся де-факто языком международного общения, позволит привлечь к ресурсам вуза аудиторию из других стран, не говорящих по-русски.

К техническим мероприятиям можно отнести следующие:

1) регистрация и добавление академического домена вуза в системы Majestic SEO и Alexa, добавление кодов Majestic Badge и виджетов от Alexa на страницы веб-ресурсов вуза. Это повысит успех его нахождения и индексирования данными системами, позволит контролировать и просматривать расширенную статистику;

2) установка на рабочие места пользователей компьютерной сети вуза с выходом в сеть Интернет программного обеспечения Alexa ToolBar или Alexa StatusBar для увеличения формируемой системой характеристики о трафике;

3) оптимизация и децентрализации веб-ресурсов вуза – переход к отдельному веб-серверу (вирт. хостинг или выделенный сервер) со своим субдоменом в общем академическом домене вуза для каждого отдельного подразделения. Это даст возможность для более автономной и успешной работы по наполнению и поддержанию контента в актуальном состоянии, а также напрямую будет способствовать росту количества страниц, способных быть проиндексированными поисковыми системами.

Заключение. В работе выполнено рассмотрение процесса формирования рейтинговой оценки вузов Webometrics Ranking. Для поиска потенциальных возможностей повысить рейтинг вуза произведено моделирование перспективной динамики роста рейтинга Алтайского государственного университета с использованием искусственных нейронных сетей, определены одни из возможных границ области, удовлетворяющей решению поставленной задачи, а также необходимые технические и организационные мероприятия, способствующие будущему росту рейтинга вуза.

Библиографический список

1. Артюшина И.А., Шутин В.А. Глобальные рейтинги на повестке дня // Вопросы образования. – 2008. – №2.

2. Карпенко О.М., Бершадская М.Д., Вознесенская Ю.А. Интернет-сайты российских вузов: динамика развития по оценке международного рейтинга университетов Webometrics в 2007–2009 гг. // Инновации в образовании. – 2010. – №1.

3. Ranking Web of World Universities. Methodology [Электронный ресурс]. – URL: <http://www.webometrics.info/methodology.html>.

4. Aguillo I.F., Ortega J.L., Fernandez M. Webometric ranking of world universities: introduction, methodology,

and future developments // Higher Education in Europe. – 2008. – Vol. 33, №2/3.

5. Ortega J.L., Aguillo I.F. Mapping world-class universities on the web // Information Processing and Management. – 2009. – Vol. 45 (2).

6. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6. – М., 2002.

7. Осовский С. Нейронные сети для обработки информации / пер. с польск. И.Д. Рудинского. – М., 2002.

8. Ranking Web of World Universities. Best Practices [Электронный ресурс]. – URL: http://www.webometrics.info/best_practices.html.