РЕЙТИНГИ САЙТОВ НАУЧНЫХ ОРГАНИЗАЦИЙ РОССЕЛЬХОЗАКАДЕМИИ

С.Г. Сальников С.Г., вед. научный сотрудник

Всероссийского института аграрных проблем и информатики им. А.А.Никонова Россельхозакадемии

Адрес: *salnsg@gmail.com

ВВЕДЕНИЕ. Вебометрика (webometrics) как одна из областей наукометрии возникла сравнительно недавно и поначалу опиралась на уже имеющиеся и хорошо разработанные методы библиометрии. Впервые термин вебометрика был использован в работе [1] В работе [2] было дано и первое определение вебометрики, как метода изучения «… количественных аспектов создания и использования информационных ресурсов, структур и технологий в Веб-пространстве…»

Со времени своего возникновения вебометрика получила значительное развитие. Один из самых успешных и известных проектов в данном направлении – рейтинг сайтов научных организаций «Ranking Web of World Research Centers» [3] методология расчёта которого была предложена и который поддерживается лабораторией Cybermetrics Lab, входящей в состав CSIC – одного из крупнейших исследовательских центров Испании. С подробным и обстоятельным анализом данного рейтинга можно ознакомиться в [4].

Появление таких популярных рейтингов подтолкнуло многих исследователей к созданию аналогичных, но локальных (страновых, отраслевых) рейтингов. Причины такого положения дел достаточно очевидны: далеко не все научные организации конкретной страны попадают в рейтинги [3]. Примеры таких локальных рейтингов можно найти в работах [5]-[8].

В данной статье предложена модификация методики [3], описанная в [7,8] и примененная для построения рейтингов сайтов научных учреждений Республики Сербия и научных учреждений Северо-Запада России. В качестве основы была взята методика из работы [8].

МЕТОДИКА РАСЧЕТА. Согласно [3] определим вебометрические индикаторы S, V, R и Sc:

S – размер сайта (S-size) – общее количество страниц, обнаруживаемых на конкретном сайте конкретной поисковой машиной. В качестве таковых в работе использовались для сбора данных по индикатору S поисковые системы Яндекс, Google и Yahoo, соответствующие индикаторы обозначены SЯндекс, SGoogle и SYahoo;

V – видимость сайта (V-visibility) – количество уникальных гиперссылок с других веб-ресурсов на ресурсы данного конкретного сайта. В качестве сервисов, используемых для расчётов величины V, использовались сервисы поисковых служб Яндекс, Google и сервиса Solomono. Соответствующие конкретные индикаторы получили обозначение VЯндекс, VGoogle и VSolomono;

R – полнотекстовость сайта (R-rich files) – количество полнотекстовых файлов, то есть суммарное количество файлов с расширениями PDF, DOC и PPT, обнаруживаемых на конкретном сайте конкретным сервисом. В качестве сервисов, используемых для расчётов величины V, использовались сервисы поисковых служб Яндекс, Google. Соответствующие конкретные индикаторы получили обозначение RЯндекс и RGoogle;

Sc – научность сайта (Sc-scholar) – количество ссылок на сайт, обнаруживаемых сервисом Google Scholar (обозначение - ScGoogle).

Заметим, что в отличие от оригинальной методики [8], мы используем другие поисковые службы и сервисы для расчётов индикаторов V и R. Причина такого состояния дел состоит в том, что расчёты видимости сайта для системы Yahoo показывают крайне скудные результаты и, по сути, не в коей мере не соответствуют тому положению дел, которые демонстрируют расчеты этого индикатора при помощи соответствующих сервисов Яндекс и Google. С другой стороны, сервис Solomono предоставляет сведения, сопоставимые с показателями для Яндекс и Google, и потому был включен нами в расчет индикатора V вместо аналогичного показателя для сервиса Yahoo в методике [8].

В отличие от [8] мы были вынуждены при расчетах индикатора R ограничиться только службами Яндекса и Google, так как вычисления аналогичного показателя для Yahoo показало не только, как и в случае с индикатором V, значительное занижение этого показателя сервисом Yahoo, но и влекло за собой большие трудности технического характера. В результате было решено отказаться от данного локального индикатора вообще с соответствующими (очевидными) изменениями в расчетных алгоритмах и формулах.

В расчетах для российских сайтов, как в нашем случае, мы, как и в [8], в дополнение к расчетам индикатора Sc по данным сервиса Google Scholar (ScGoogle) возьмём также расчёты по ScЯндекс – количеству упоминаний конкретного сайта в Каталоге сайтов Яндекса.

Рейтинг сайта вычисляется по формуле

W=V+S+2*R+1,5*Sc, где

V=log10(1+V0); V0= (VЯндекс+VGoogle+VSolomono)/3

S=log10(1+S0); S0= (SЯндекс+SGoogle+SYahoo)/3

R=log10(1+R0); R0= (RЯндекс+RGoogle)/2

Sc=(log10(1+ScЯндекс)+log10(1+ScGoogle))/2

Небольшая модификация по сравнению с методикой [8] — ввод поправочных слагаемых «1+» — связана с тем, что в расчётах индикаторов для конкретных сайтов зачастую появляются нули. Следует также сразу сделать ряд замечаний по приведенному расчетному алгоритму. Мы целиком и полностью согласны с замечаниями из [8] по части применимости коммерческих поисковых систем для измерения приведённых выше вебометрических индикаторов. Значения аналогичных индикаторов для одного и того же сайта, выдаваемыми сервисными службами разных поисковых систем, могут различаться на порядки (!) Причины такого положения дел многообразны: здесь и закрытость алгоритмов коммерческих поисковых систем, используемых при расчётах тех или иных показателей. И принудительное сокращение объёма информации, которая хранится поисковыми сервисами. И политика самих администраторов сайтов, которые могут запретить доступ к ряду страниц на своём сайте, что заставляет программы-роботы (боты) собирающие информацию с сайтов соблюдать такие правила, что, очевидно, существенно искажает реальную картину. Вызывает сомнение также принцип усреднения аналогичных показателей. Анализ данных показателей на малых сайтах (с обозримым числом его страниц) показывает, что вместо, например,

V0= (VЯндекс+VGoogle+VSolomono)/3

разумнее было бы взять в расчёты

V0= max(VЯндекс,VGoogle,VSolomono)

Аналогичные соображения верны и для других индикаторов, представленных выше. Действительно, если алгоритмы поиска всех страниц сайта достаточно адекватны и если поисковый сервис ищет (или это декларирует) все страницы конкретного сайта (или какую-то пропорционально равную их часть), то индикаторы выше должны достаточно хорошо коррелировать друг с другом. И действительно корреляция SЯндекс и SGoogle составляет 0,82 тогда как для пар (SЯндекс , SYahoo) и (SGoogle , SYahoo) она составляет только 0,43 и 0,37 соответственно. Приведённые значения показателей и тот факт, что все значения индикатора SYahoo для конкретных сайтов всегда (за редким исключением) меньше (иногда существенно) соответствующих индикаторов SЯндекс и SGoogle для тех же сайтов, позволяет сделать предположение, что индикатор SYahoo недостаточно корректно отражает требуемые показатели и попросту неадекватен поставленным целям. Однако, в силу уже имеющейся традиции, заложенной в локальных рейтингах [5-8] и (опять таки) в силу того, что значения индикатора SYahoo почти всегда невелики, было решено оставить этот индикатор в расчётных формулах.

Ещё хуже обстоят дела с показателем V. Если корреляция между показателями VGoogle и VSolomono приемлема (0,63) то в парах (VЯндекс , VGoogle) и (VЯндекс , VSolomono) она неудовлетворительна (0,42 и 0,38 соответственно).

Для пары (RЯндекс , RGoogle) ситуация достаточно благополучна, корреляция показателей достигает 0,88. Но тогда вновь возникает вопрос о том, насколько уместно усреднение по данным показателям?

Один из способов разрешения представленных выше противоречий (достаточно эффективный и универсальный) – построение порядковых рейтингов по каждому из показателей и выведение общего рейтинга как суммы частичных рейтингов, был предложен в [7]. Мы вернёмся к данному вопросу в анализе результатов.

ЦЕЛЕВОЕ МНОЖЕСТВО РЕЙТИНГА. В качестве объекта исследования выступили официальные сайты 105 научных организаций, входящих в состав Россельхозакадемии. Список наименований этих организаций будет приведен ниже вместе с рейтингом этих сайтов. Веб-адреса их официальных сайтов можно найти в специализированном каталоге на сайте ВИАПИ им. А.А. Никонова [9].

РЕЙТИНГ САЙТОВ. В приведенной ниже таблице 1 представлен ранжированный список всех 105 сайтов вместе со значениями индикатора и наименованиями учреждений.

Таб. 1 Рейтинги научных учреждений Россельхозакадемии		
НАИМЕНОВАНИЕ	РЕЙТИНГ	МЕСТО
ВИР	14,93	1
ВНИИ экономики сельского хозяйства	14,48	2
ВИАПИ им. А.А. Никонова	14,16	3
ВНИИ птицеперерабатывающей промышленности	14,08	4
НИИСХ Юго-Востока	12,63	5
ВНИИГиМ	12,57	6
ВНИИ селекции плодовых культур	12,15	7
ГОСНИТИ	11,93	8
ВНИИМК	11,81	9
ВНИИ коневодства	11,79	10
СЗНИИМЭСХ	11,57	11
ВНИИССОК	11,48	12
ВСТИСП	11,44	13
ВНИИ жиров	11,27	14
Ставропольский НИИСХ	11,26	15
ВНИОПТУСХ	11,17	16
ВНИТИП	10,74	17
СНИИЖК	10,01	18
Почвенный институт им. В.В. Докучаева	9,97	19
УрНИВИ	9,74	20
ВНИИ цветоводства и субтропических культур	9,70	21
Агрофизический НИИ	9,69	22
СКЗНИИСиВ	9,42	23
ВИЭСХ	9,41	24
ВНИИ охотничьего хозяйства и звероводства»	9,24	25
ВНИИ овощеводства	9,19	26
ВНИИ агролесомелиорации	9,14	27
ВНИИ экономики и нормативов	8,95	28
ВНИВИ патологии, фармакологии и терапии	8,92	29
ВНИИ лекарственных и ароматических растений	8,90	30
ВНИИ пивовар., безалкогол. и винодельческой промышленности	8,86	31
КНИИСХ им. П.П.Лукьяненко	8,84	32
ВИЭВ	8,84	33
ВНИИМП им. В.М.Горбатова	8,84	34
ВНИИ механизации сельского хозяйства	8,79	35
РНИИСП	8,65	36
Краснодарский НИВИ	8,55	37
ВИУА им Прянишникова	8,26	38
ВНИИМС	8,13	39
ГОСНИИ хлебопекарной промышленности	7,78	40
НИИ хранения и переработки сельскохозяйственной продукции	7,59	41
Татарский НИИ сельского хозяйства	7,47	42
ВНИТИ биологической промышленности	7,45	43
ВНИИГРЖ	7,40	44

ВНИИОБ	7,13	45
НИИ пчеловодства	7,01	46
ВНИИФБиП	6,69	47
ВНИИПБТ	6,69	48
Башкирский НИИ сельского хозяйства	6,68	49
ВНИИ риса	6,64	50
Костромской НИИСХ	6,51	51
ВНИИ крахмала и крахмалопродуктов	6,51	52
Прикаспийский ЗНИВИ	6,40	53
ВНИИСС	6,26	54
ЮУНИИПОК	6,19	55
ВНИИ консервной и овощесушильной промышленности	6,00	56
Нижне-Волжский НИИСХ	5,99	57
Северо-Кавказский ЗНИВИ	5,96	58
ВНИХИ	5,96	59
НИИ аридного земледелия	5,93	60
ВНИИ зерна и продуктов его переработки	5,89	61
ВНИИ виноградарства и виноделия	5,89	62
НИИДП	5,75	63
Донской НИИСХ	5,75	64
НИИСХ Центрально-Черноземной полосы им. В.В.Докучаева	5,70	65
Владимирский НИИСХ	5,70	66
ВНИИ ветеринарной вирусологии и микробиологии	5,62	67
ПНИИСС им. П.Н. Константинова	5,59	68
ВНИИР	5,55	69
ВНИИ люпина	5,49	70
ВНИИ сельскохозяйственной микробиологии	5,49	71
ВНИИ зернобобовых и крупяных культур	5,47	72
НИИ пушного звероводства и кролиководства им.В.А.Афанасьева	5,44	73
ВНИИ кормов имени В. Р. Вильямса	5,35	74
НИИММП	5,15	75
ВНИМС	5,10	76
ВНИИМЖ	5,05	77
ВНИИТиН	4,97	78
ВНИИ льна	4,92	79
ВНИИОЗ	4,78	80
ВНИИМЗ	4,55	81
ВНИИМЛ	4,53	82
ВНИИ табака, махорки и табачных изделий	4,30	83
Ульяновский НИИСХ	4,27	84
Приморский НИИСХ	4,16	85
Саратовская НИВС	4,11	86
НИИЭОАПК ЦЧР РФ РАСХН	4,01	87
НИИСХ ЦРНЗ	3,96	88
СКНИИЖ	3,79	89
ВНИИ ветеринарной санитарии, гигиены и экологии	3,47	90
Архангельский НИИСХ	3,45	91
Смоленский НИИСХ	3,40	92
ВНИИ земледелия и защиты почв от эрозии	3,36	93

Северо-Западный НИИ молочного и лугопастбищного хозяйства	3,12	94
ВНИИ фитопатологии	3,07	95
Краснодарский НИИ овощного и картофельного хозяйства	2,82	96
Удмуртский ГНИИ сельского хозяйства	2,75	97
Всероссийский Институт Гельминтологии им. К.И. Скрябина	2,51	98
РНИС Шелководства	2,41	99
ВНИИ сельскохозяйственной радиологии и агроэкологии	2,31	100
Чувашский НИИСХ	2,29	101
НИИ пищеконц. промышленности и спец. пищевой технологии	2,23	102
Самарская НИВС	2,01	103
Сибирский НИИ птицеводства	1,39	104
Пермский НИИСХ	1,34	105

АНАЛИЗ РЕЗУЛЬТАТОВ. Расчёты рейтинга позволяют сделать ряд обобщений.

Так, группировка по индикатору V выглядит следующим образом (таблица 2):

Таб 2 Распределение сайтов по количеству ссылок				
Количество ссылок на сайт				
	До 10	От 10 до 100	От 100 до 1000	Свыше 1000
Сайтов в группе	8	44	45	8
В % от общего количества	7,6%	41,9%	42,9%	7,6%

Как видно распределение сайтов по группам достаточно симметрично. Наряду с явными аутсайдерами (1-ая группа, 8 сайтов или 7,6% от общего количества сайтов), есть такая же по размерам (4-ая) группа лидеров (8 сайтов, 7,6%). Остальные сайты (89 сайтов, 84,8%) попадают в группы с удовлетворительными (2-я группа) показателями (44 сайта, 41,9%) или хорошими показателями (3-я группа) индикатора (45 сайтов, 42,9%).

Аналогичная картина наблюдается для индикатора S (таблица 3):

Таб. 3 Распределение сайтов по количеству страниц				
Количество страниц на сайте				
	До 10	От 10 до 100	От 100 до 1000	Свыше 1000
Сайтов в группе	10	50	36	9
В % от общего количества	9,5%	47,6%	34,3%	8,6%

Как видно, распределение сайтов по группам вновь достаточно симметрично, наряду с аутсайдерами (1-ая группа, 10 сайтов или 9,5% от общего количества сайтов) есть примерно такая же по размерам (4-ая) группа лидеров (9 сайтов, 8,6%), остальные сайты попадают в группы с удовлетворительными (2-я группа) показателями (50 сайта, 47,6%) или хорошими показателями (3-я группа) индикатора (36 сайтов, 34,3%).

Группировка по индикатору R выглядит качественно иначе (таблица 4):

Таб. 4 Распределение сайтов по количеству размещённых полнотекстовых файлов				
Количество полнотекстовых файлов на сайте				
	До 10	От 10 до 100	От 100 до 1000	Свыше 1000
Сайтов в группе	58	34	13	-
В % от общего количества	55,2%	32,4%	12,4%	-

Как видно, сайтов, на которых размещено 1000 и более полнотекстовых файлов в целевом множестве сайтов нет вообще. Больше половины сайтов (58 сайтов или 55,2% от общего количества сайтов в целевом множестве) входят в первую группу. Около трети сайтов (34 сайта или 32,4%) входят во вторую группу и только примерно 1/8 часть организаций (13 организаций или 12,4%) входят в 3-ую группу и представлены достаточно большим (от 100 до 1000) количеством полнотекстовых файлов на своих сайтах.

Группировка по индикатору Sc во многом аналогична таковой для показателя R (таблица 5):

Таб. 5 Распределение сайтов по индикатору научности				
Значение индикатора научности сайта				
	До 10	От 10 до 100	От 100 до 1000	Свыше 1000
Сайтов в группе	65	38	2	-
В % от общего количества	61,9%	36,2%	1,9%	-

Вновь, сайтов с показателем более 1000 нет вообще. Почти 2/3 сайтов (66 сайтов или 61,9% от общего количества сайтов в целевом множестве) входят в 1-ую группу. Около 1/3 сайтов (38 сайтов или 36,2%) входят во 2-ую группу и только 1/50 часть организаций (2 организации или 1,9%) входят в 3-ую группу и представлены большим (от 100 до 1000) количеством ссылок на них в каталогах Яндекса и Google Scholar.

Наконец, возвращаясь к обсуждению порядкового рейтинга, которое было начато в конце раздела «Методика расчёта», заметим, что пересчёты по такому (достаточно обобщённому и универсальному) алгоритму были нами произведены. Вновь составленный список-рейтинг хорошо коррелирует с нашим первоначальным рейтингом (коэффициент корреляции равен 0,73), что говорит о достаточной объективности представленного выше рейтинга и о том, что он в целом отражает некие общие, универсальные тенденции в составе сайтов научных организаций Россельхозакадемии. Для подтверждения данного тезиса приведём таблицу, сопоставляющую первые десятки организаций по обоим рейтингам (таблица 6):

Таб. 6. Сравнительные рейтинги научных организаций Россельхозакадемии		
Место	Наименование	
1	ВИР	ВИР
2	ВНИИ экономики сельского хозяйства	ВНИИ экономики сельского хозяйства
3	ВИАПИ им. А.А. Никонова	ВИАПИ им. А.А. Никонова
4	ВНИИПП	ВНИИПП
5	ВНИИ селекции плодовых культур	НИИСХ Юго-Востока
6	ГОСНИТИ	ВНИИГиМ
7	ВНИИССОК	ВНИИ селекции плодовых культур
8	ВСТИСП	ГОСНИТИ
9	ВНИИ коневодства	ВНИИМК
10	НИИСХ Юго-Востока	ВНИИ коневодства

Как видно, первая четвёрка сайтов в рейтингах совпадает полностью, а общее совпадение в первой десятке для обоих рейтингов составляет 80%.

ВЫВОДЫ. Очевидный вывод из приведенного выше рейтинга и его анализа таков:

Существует только 4 сайта (из 105) – сайты ВИР, ВНИЭСХ, ВИАПИ и ВНИИПП - в должной степени соответствующие понятию «сайта научной организации» и с достаточно большим отрывом опережающие по всем своим показателям остальные сайты научных организаций РАСХН в рейтинге сайтов.

Ещё 12 сайтов (в списке они занимают места с 5 по 16) можно признать удовлетворяющими данному званию в должной мере. Остальные сайты (места в рейтинге с 17 по 75-80) к сайтам научных организаций могут быть отнесены только условно и только по некоторым признакам (индикаторам) из ряда таковых. Наконец, примерно одна треть сайтов (25-30 сайтов из конца списка) не удовлетворяют ни одному из признаков научности сайта и состояние дел с ними - как с сайтами научных организаций - должно быть признано неудовлетворительным.

СПИСОК ЛИТЕРАТУРЫ

[1] Almind T., Ingwersen P. Infometric analyses on the World Wide Web: methodological approaches to ‘webometrics’ //Journal of Documentation. 1997. V. 53. N 4. P. 404-426

[2] Bjornerborn L., Ingwersen P. Perspectives of webometrics // Scientometrics. 2001. V. 50. N 1. P. 65-82.

[3] Ranking Web of World Research Centers [Электронный ресурс] – Режим доступа: *http://research.webometrics.info

[4] Карпенко О.М., Бершадская М.Д., Вознесенская Ю.А. Международный рейтинг университетских веб-сайтов [Электронный ресурс] – Режим доступа: *www.edit.muh.ru/content/mag/trudy/09_2008/06.pdf

[5] Шокин Ю.И., Клименко О.А., Рычкова Е.В., Шабальников И.В. Рейтинг сайтов научных организаций СО РАН// Вычислительные технологии 2008. Т. 13. N 3. С. 128-135.

[6] Ханчук А.И., Наумова В.В. Информационное пространство Дальневосточного отделения РАН// Информационные технологии. Вестник ДВО РАН 2009. N 4. С. 122-129.

[7] Шокин Ю.И., Клименко О.А., Рычкова Е.В. Рейтинг сайтов научных организаций Республики Сербия [Электронный ресурс] – Режим доступа: *http://www.ict.nsc.ru/userfiles/file/presentation/MIT-09/Srbija_webometrics.pdf

[8] Мазалов В.В., Печников А.А. О рейтинге официальных сайтов научных учреждений Северо-Запада России// Управление большими системами. Выпуск 24. С. 130-146.

[9] Сайты научных организаций Россельхозакадемии [Электронный ресурс] – Режим доступа: *http://www.viapi.ru/services/codificator/institute/

Назад в раздел

Поделиться...

© «Битрикс», 2001-2006