

# РЕЙТИНГИ КАК ЦИФРОВОЙ ИНСТРУМЕНТ ОЦЕНКИ ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ ОРГАНИЗАЦИЙ<sup>26</sup>

Меденников В.И., Флеров Ю.А.

Вычислительный центр им. А.А. Дородницына, ФИЦ ИУ РАН,  
Россия, г. Москва, ул. Вавилова, д.44-2

[dommed@mail.ru](mailto:dommed@mail.ru), [fler@ccas.ru](mailto:fler@ccas.ru)

*Аннотация: в работе предлагаются методики формирования рейтингов отечественных научно-исследовательских организаций на примере аграрных для оценки эффективности их деятельности с учетом активного, пассивного и полуактивного цифровых следов в интернет-пространстве. Приведено сравнение и анализ полученных рейтингов аграрных научно-исследовательских организаций.*

Ключевые слова: цифровой след, рейтинги, научно-исследовательские организации, интернет-пространство.

## Введение

На имидж большинства организаций в условиях цифровой экономики сегодня значительно влияет грамотное, эффективное представительство их в интернете, получившее понятие цифрового следа. Значение его для науки в инновационной экономике резко выросло, поскольку интернет помимо предоставления ученым качественно новых возможностей для широкого обмена идеями между собой и их цифрового взаимодействия с научными информационными ресурсами (ИР) обеспечивает также эффективную систему трансфера научных знаний в экономику при определенных условиях, способствует росту интеллектуального потенциала общества через совершенствование системы образования.

Обычно цифровой след подразделяют на два основных вида: пассивный и активный. К пассивному относят информацию, собираемую без ведома собственника. К активному относят информацию, публикуемую в интернете самим владельцем, либо уполномоченным им лицом. В реальной жизни это деление условно, поскольку информация о деятельности организаций может публиковаться из различных других источников, в том числе генерируемых автоматизированными методами. Назовем такой цифровой след полуактивным.

В наших исследованиях будем считать, что активный цифровой след в сфере аграрных научно-исследовательских организаций в России – это массив данных, который размещается на сайте каждой из них; пассивный – информация, собираемая и обрабатываемая при помощи наиболее известной программы анализа и аудита сайтов Site-Auditor [1]; полуактивный – информация, размещаемая ими в системе РИНЦ. Однако, в нашей стране наблюдается существенная недооценка активного и пассивного цифровых следов. Так, по этой причине не был выполнен указ Президента Российской Федерации от 2012 г., в соответствии с которым к 2020 г. свыше пяти ВУЗов России должны были попасть в топ-100 рейтингов ведущих мировых университетов, хотя были выделены из бюджета значительные средства – 80 млрд рублей [2]. Аналогично, в науке, по мнению Минобрнауки, основные усилия должны быть направлены на повышение наукометрических показателей ученых: «В настоящее время Минобрнауки России реализует ряд проектов, направленных на повышение видимости публикаций российских ученых, повышение показателей и рейтинга российских журналов и их вхождения в международные наукометрические базы данных» [3]. Следствием такого подхода является то, что научную общественность больше интересует необходимость постоянно помнить о количестве публикаций, рейтинге журналов и издательств, в которых публикуешься, о том, чтобы вовремя загрузить статью в БД, дабы увеличить цитируемость в соответствии с созданным механизмом принуждения, который заставляет ученых выбирать темы исследований в соответствии с указанными выше критериями, а не потребностями экономики, общества.

В связи с этим и возрастающей ролью интернет-технологий в современном цифровом обществе актуальными становятся исследования по разработке рейтингов аграрных научно-исследовательских организаций (НИО) на основе их активного, пассивного и полуактивного цифровых следов в интернет-пространстве.

---

<sup>26</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №20-07-00836 "Научные основы формирования единой цифровой платформы (единого информационного Интернет-пространства) аграрных научно-образовательных ресурсов на основе математического моделирования".

## 1 Методика формирования рейтингов на основе пассивного цифрового следа

Как уже отмечалось выше, для анализа пассивного цифрового следа воспользуемся помощью наиболее известной программы анализа и аудита сайтов Site-Auditor, определяющей 41 показатель, характеризующий исследуемый сайт по семи группам параметров (таб. 1).

Таблица 1. Группы и показатели анализа сайтов программой Site-Auditor

№ п/п	Показатель	Группа	Краткие пояснения к показателю
1	Baidu	Индексация	Количество страниц сайта, проиндексированных поисковой системой Baidu
2	Bing		Количество страниц сайта, проиндексированных поисковой системой Bing
3	Google		Количество страниц сайта, проиндексированных поисковой системой Google
4	Google.Images		Количество изображений на страницах сайта, проиндексированных поисковой системой Google
5	Seznam		Количество страниц сайта, проиндексированных поисковой системой Seznam
6	Wada		Количество страниц сайта, проиндексированных поисковой системой Wada
7	Яндекс		Количество страниц сайта, проиндексированных поисковой системой Яндекс
8	DMOZ	Каталоги	Наличие сайта в каталоге DMOZ
9	Mail.ru		Наличие сайта в каталоге Mail.ru
10	Seznam		Наличие сайта в каталоге Seznam
11	Рамблер ТОП		Наличие сайта в каталоге Рамблер ТОП
12	Яндекс		Наличие сайта в каталоге Яндекс
13	AdGuard	Проблемы	Проблемы на сайте, обнаруженные сервисом AdGuard
14	Spamhaus (IP)		Проблемы на сайте, обнаруженные сервисом AdGuard Проблемы на сайте, обнаруженные сервисом Spamhaus (IP)
15	Spamhaus (Домен)		Проблемы на сайте, обнаруженные сервисом Spamhaus (Домен)
16	Инфицирован (Google)		Проблемы на сайте, обнаруженные Google
17	Инфицирован (Яндекс)		Проблемы на сайте, обнаруженные Яндексом
18	Яндекс.АГС-40		Проблемы на сайте, обнаруженные сервисом Яндекс.АГС-40
19	Яндекс.Минусинск		Проблемы на сайте, обнаруженные сервисом Яндекс.Минусинск
20	Alexa global	Рейтинги	Рейтинги сервиса Alexa (глобальные)
21	Alexa local		Рейтинги сервиса Alexa (локальные)
22	SimWeb global		Рейтинги сервиса Sim Web (глобальные)
23	SimWeb local		Рейтинги сервиса Sim Web (локальные)
24	Яндекс ТИЦ		Показатель ТИЦ от Яндекса
25	Backtweets	Социальные сервисы	Найдено ссылок на сайта в сервисе Backtweets
26	Facebook		Найден сайт в Facebook
27	Facebook share		Найдены ссылки на сайт в Facebook
28	Facebook комментарии		Количество упоминаний сайта в комментариях Facebook
29	Facebook ссылки		Количество ссылок на сайт в Facebook
30	Google Plus		Найдено упоминаний сайта в Google Plus
31	Мой мир (mail.ru)		Найдено упоминаний сайта в «Мой мир»
32	Alexa	Ссылки на сайт	Найдено ссылок на сайт сервисом Alexa
33	Google		Найдено ссылок на сайт поисковиком Google
34	Linkpad сайты		Найдено сайтов со ссылками на сайт сервисом Linkpad
35	Linkpad ссылки		Найдено ссылок на сайт сервисом Linkpad

№ п/п	Показатель	Группа	Краткие пояснения к показателю
36	Majestic сайты		Найдено сайтов со ссылками на сайт сервисом Majestic
37	Majestic ссылки		Найдено ссылок на сайт сервисом Majestic
38	Яндекс блоги		Найдено Яндексом упоминаний сайта в блогах
39	Bing	Ссылки с сайта	Ссылки с сайта, найденные сервисом Bing
40	Linkpad сайты		Ссылки с сайта (на сайты) найденные Linkpad
41	Linkpad ссылки		Ссылки с сайта (все), найденные сервисом Linkpad

Анализ полученной сводной таблицы по всем НИО показал, что для дальнейшего рассмотрения оставлены 18 показателей, поскольку остальные имеют либо нулевые значения, либо отличны от нуля для небольшого количества сайтов (таб. 2).

Таблица 2. Группы и показатели, используемые для анализа сайтов и сопоставления рейтингов

№ п/п	Показатель	Группа
2	Bing	Индексация
3	Google	
4	Google.Images	
7	Яндекс	
20	Alexa global	Рейтинги
21	Alexa local	
22	SimWeb global	
23	SimWeb local	
24	Яндекс тИЦ	Ссылки на сайт
32	Alexa	
33	Google	
34	Linkpad сайты	
35	Linkpad ссылки	
36	Majestic сайты	
37	Majestic ссылки	
39	Bing	Ссылки с сайта
40	Linkpad сайты	
41	Linkpad ссылки	

Вновь полученная таблица по всем НИО с 18 столбцами была загружена в среду R для последующего анализа, очистки, выявления выбросов, их удаления и пополнения получившихся пропусков в данных штатными средствами пакета статистической обработки информации R [4].

Анализ информации, выдаваемой для сайтов НИО программой Site-Auditor, показал искажение данных для некоторых сайтов домена ниже второго (совпадение данных с вышестоящим доменом), а также невыдачу рейтингов и ссылок для доменов уровня ниже второго сайтом Alexa. В результате Кемеровский НИИ сельского хозяйства и Сибирский НИИ растениеводства и селекции из последующего рассмотрения исключили с присвоением оценки 0 и постановкой в рейтинговом списке на последнее место. Также показатели под номером 20, 21 и 32 из таб. 2 удаляются, считаются пропусками и будут в дальнейшем восстанавливаться как пропуски.

Рассмотрим проблему выбросов и пропусков в данных. Как выяснилось выше, основные кандидаты на пропуски – это значения рейтингов и ссылок, полученные с сервиса Alexa для сайтов, созданных на доменах уровня ниже второго. Сопоставление рейтингов, которые дают сервисы Alexa, SimWeb и Яндекс (тИЦ) показывает противоречивые данные. Там, где одни сервисы дают определённые значения, другие предлагают нулевые значения и наоборот. Таким образом, можно предположить, что нулевые значения должны быть исключены и соответствующие данные нужно считать пропущенными за исключение сервиса Яндекс тИЦ, у которого выходные значения невелики, поэтому и нулевые его значения можно отставить. Таким образом, исключим данные (и в последующем будем пополнять) для нулевых значений четырёх показателей с кодами 20-23 (глобальные и локальные рейтинги Alexa и SimWeb), а также явные выбросы для показателя 39 с

сайта Почвенного института им. В.В. Докучаева, а также данные по показателю 40 (ссылки с сайта, найденные сервисом Linkpad) для сайта НИВИ Нечернозёмной зоны РФ. Наконец, удалим все нулевые данные. В последующем мы будем пополнять такие данные как пропуски.

Загрузка измененных данных в Site-Auditor приводит к тому, что среда сервиса data.world предупреждает нас о том, что 35 числовых значений в нашем файле выглядят как выбросы. Простое решение проблемы с удалением проблемных данных приводит к тому, что во вновь полученном файле число ошибок возрастает до 38, что связано с изменением статистических характеристик оставшихся данных при выбросе проблемных. Решение данной задачи сервис data.world не дает, что вновь заставило нас обратиться к возможностям чистки данных в среде [4], для чего необходимо данные нормализовать.

Первоначальная визуальная проверка построением гистограмм и последующее уточнение полученных визуальных оценок при помощи теста на нормальность Шапиро-Уилка показали, что для пяти показателей 03, 35, 37, 40 и 41 достаточно типового логарифмического преобразования. Для остальных переменных нормализация была проведена при помощи преобразования Бокса-Кокса

$$x'(\lambda) = \frac{x^\lambda - 1}{\lambda} \quad (1)$$

В случае  $\lambda = 0$  использовали логарифмическое преобразование  $x'(0) = \ln(x)$ . Для нахождения оптимального коэффициента  $\lambda$  была использована идея [5] о его нахождении при помощи принципа максимального правдоподобия. В результате чего количество выбросов сократилось до двух при первоначальной значительной величине. Результаты преобразования для показателей 02, 04, 34 и 39 оказались неудовлетворительными без какого-либо разумного объяснения, поэтому были исключены из дальнейшего рассмотрения.

После приведения всех переменных к нормальному виду остаётся проблема пополнения данных, а именно замена нулевых значений и выбросов на подходящие значения. Согласно стандартной процедуре [5] проблема «борьбы с пропусками» в данных включает в себя 3 типовых шага:

- идентификация недостающих данных;
- исследование закономерностей появления отсутствующих значений;
- формирование наборов данных, не содержащих пропуски (в результате удаления или замены соответствующих фрагментов).

Причём, только первый этап чётко и однозначно формализуем, второй же и третий требуют неформального подхода в каждой конкретной ситуации. В нашем случае получилось 128 сайтов хотя бы с одним пропущенным значением и 55 без таковых.

При этом сайты с номерами 4, 13, 32, 34, 45, 68, 96, 123, 142, 150 и 182 из последующих расчётов были исключены с нулевой оценкой и размещением в конце рейтингового списка в силу превышения количество позиций с пропущенными данными над количеством позиций сохранённых. И лишь для оставшихся 172 сайтов была проведена процедура пополнения пропущенных данных [6].

Основные возможности для анализа пропущенных данных в статистической среде R предоставляет пакет “mice” [7]. Из множества методов заполнения пропущенных значений воспользуемся наиболее популярным и стабильным из них – так называемым методом «случайного леса», основное преимущество которого в нашем случае крайне актуально: метод не требует нормальности всех используемых переменных, поскольку ряд переменных не являются нормальными даже после проведенных преобразования.

После заполнения пропущенных значений из вновь полученного массива данных исключим подраздел «Рейтинги» для формирования на оставшемся массиве рейтингов сайтов и сравнением их с теми же глобальными рейтинговыми списками, которые уже существуют, а именно с ТИЦ (Яндекс) и рейтинг-листами Alexa и SimWeb. Тогда внутри каждой из оставшихся подгрупп – «Индексация», «Ссылки с сайта» и «Ссылки на сайт» веса показателей при суммировании их значений будут произведены в соответствии с распределением стандартных отклонений каждой их переменной по принципу «большее стандартное отклонение – меньший весовой коэффициент и наоборот». Идея такого решения состоит в том, что так подобранными коэффициентами удаётся, как правило, существенно снизить стандартного отклонение результата, что в принципе всегда желательно для любых статистических результатов. Подбор же параметров из каких-то других соображений или просто суммирование с равными весами (вычисление среднего значения по группе) приводит обычно к худшим результатам, чем тот вариант, о котором написано выше, для чего воспользуемся результатами работы [8]. Веса будем выбирать по найденной там формуле:

$$v_i = \frac{1}{k-1} \cdot \frac{\sigma^2 - \sigma_i^2}{\sigma^2}, \quad (2)$$

где  $k$ - количество показателей,  $\sigma_i^2$  - дисперсия  $i$ -го показателя,  $\sigma^2 = \sum_{i=1}^{i=k} \sigma_i^2$ . Легко видеть, что

$\sum_{i=1}^{i=k} v_i = 1$  и, что чем больше дисперсия конкретной переменной, тем меньше её весовой коэффициент.

Результаты таких расчётов весовых коэффициентов в виде % для каждой из трёх групп параметров– «Индексация», «Ссылки на сайт» и «Ссылки с сайта» - приведены в таблице ниже (таб. 3). При этом веса групп проводились с использованием экспертных оценок [6].

Таблица 3. Расчётные коэффициенты показателей и групп

Код	Наименование	Группа	Вес в группе	Вес группы
03	Индексация в поисковике Bing	Индексация	45,35	15,0
07	Индексация в поисковике Яндекс		54,65	
Итого по подгруппе			100,00	
32	Найдено ссылок на сайт сервисом Alexa	Ссылки на сайт	22,82	70,0
33	Найдено ссылок на сайт поисковиком Google		23,99	
35	Найдено ссылок на сайт сервисом Linkpad		10,33	
36	Найдено сайтов со ссылками на сайт сервисом Majestic		21,58	
37	Найдено ссылок на сайт сервисом Majestic		21,28	
Итого по подгруппе			100,00	
40	Ссылки с сайта (на сайты) найденные Linkpad	Ссылки с сайта	81,54	15,0
41	Ссылки с сайта (все), найденные сервисом Linkpad		18,46	
Итого по подгруппе			100,00	
ИТОГО по всем группам				100,0

Теперь формализуем методику оценки аграрных НИО на основе их пассивного цифрового следа, для чего введем следующие обозначения.

$P_2^m$  – критерий оценки  $m$ -го НИО на основе их пассивного цифрового следа;

$d_{rm}^2$  – объем  $r$ -го показателя оценки сайта на основе пассивного цифрового следа  $m$ -го НИО;

$q_{rm}^2$  – значение  $r$ -го показателя оценки сайта на основе пассивного цифрового следа  $m$ -го НИО;

$\omega_r^2$  – вес значения  $r$ -го показателя оценки сайта на основе пассивного цифрового следа;

$$q_{rm}^2 = d_{rm}^2 / \max_m d_{rm}^2.$$

Тогда получим

$$P_2^m = \sum_k \omega_k^2 q_{km}^2. \quad (3)$$

Окончательные результаты приведены в сравнительной таблице 12.

Как было отмечено выше, представляет определённый интерес сравнение полученных рейтингов с теми же глобальными рейтинговыми списками, которые уже существуют, а именно с ТИЦ (Яндекс) и рейтинг-листами Alexa и SimWeb. Для примера ниже приведена точечная диаграмма для нашего вновь построенного рейтинга на основе пассивного цифрового следа и рейтинга, построенного на основе показателей системы SimWeb local (рис. 1).

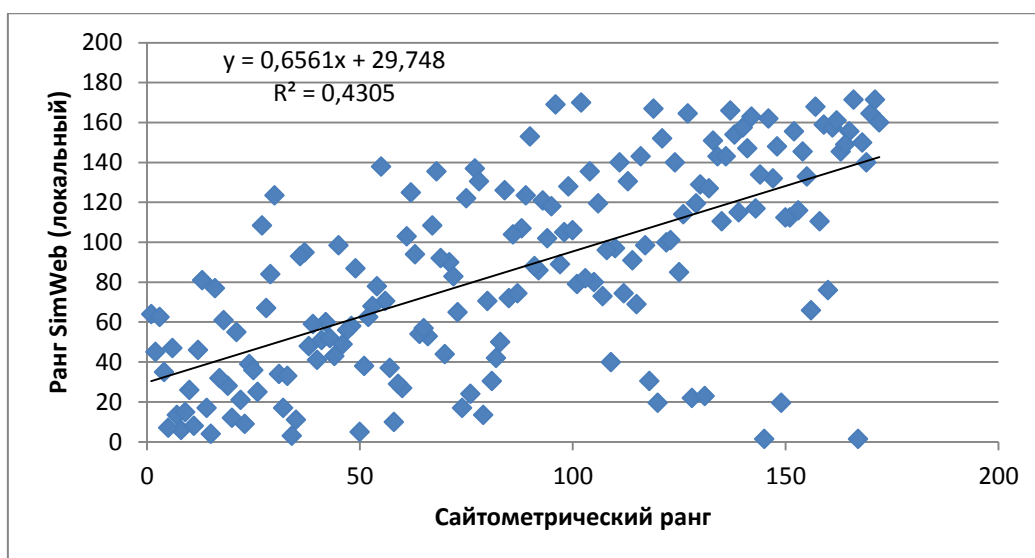


Рис. 1. Сравнение рейтингов на основе пассивного цифрового следа и SimWeb (local)

Визуально тесная связь рейтингов достаточно очевидна, что подтверждает и большое значение параметра  $R^2$ . Впрочем, использование этого параметра в таких расчётах не вполне корректно, так как мы имеем дело с ранговыми величинами. Правильнее в данной ситуации для расчётов использовать корреляционные коэффициенты Кендалла или Спирмена.

Как следует из анализа, все связи умеренно сильны и высоко значимы (на уровне  $<0,1\%$ ). При этом корреляция между вновь созданным рейтингом НИО и рейтингом, построенным на основе тИЦ Яндекса, более сильная, чем между вновь построенным рейтингом и остальными рейтингами (0,63 для тИЦ Яндекса и в границах от 0,39 до 0,51 для остальных рейтингов). Этого, впрочем, следовало ожидать, так как тИЦ Яндекса, как локальный рейтинг, должен лучше отражать местную (региональную) специфику сайтов.

## 2 Методика формирования рейтингов на основе полуактивного цифрового следа

Информация, необходимая для формирования рейтингов на основе полуактивного цифрового следа, бралась с сайта Российского индекса научного цитирования (РИНЦ). Состав же информации приведён в таб. 4. Веса всех показателей в окончательных расчётах рейтингов брались одинаковыми и равными 0,2. Поскольку Ингушский НИИСХ не был представлен в РИНЦ, то в окончательном рейтинге он получил нулевую оценку. Проведенная реформа науки серьезно сказалась на полуактивном цифровом следе многих НИО. Так, для 45 из них информация в РИНЦ размещена общим количеством в составе информации о федеральном научном центре (ФНЦ), либо федеральном исследовательском центре (ФИЦ), в который они попали в результате реформ. В дальнейшем будем именовать их как ФЦ. Тем не менее, данные НИО были оценены отдельно, так же, как и соответствующий ФЦ. Такой подход представляется вполне логичным, поскольку данные собирались для организаций только усреднёнными по работникам, исходя из предположения, что при объединении НИО в ФЦ многие показатели публикационной активности не ведут себя аддитивно. Грамотней было бы оценивать отдельные НИО с весовыми коэффициентами, пропорциональными количеству научных сотрудников в них, но отсутствие подобной информации подтолкнуло к упрощённому варианту, указанному выше.

Таблица 4. Показатели для формирования рейтингов на основе полуактивного цифрового следа

№ п/п	Показатель	Вес в рейтинге
1	Индекс Хирша организации	0,2
2	Средневзвешенный импакт-фактор журналов, в которых публиковались статьи сотрудников НИО (за последние 5 лет)	0,2
3	Среднее число публикаций в расчёте на одного автора (за последние 5 лет)	0,2
4	Среднее число цитирований в расчёте на одну публикацию (за последние 5 лет)	0,2
5	Среднее число цитирований в расчёте на одного автора (за последние 5 лет)	0,2

Как видно из таб. 4, информация собиралась, как за последние 5 лет, так и за весь период деятельности соответствующей НИО (индекс Хирша). Поэтому единственный вопрос об адекватной

оценке параметров при слиянии НИО в один ФЦ остаётся для индексов Хирша, так как, очевидно, что при слиянии этот показатель (в отличие от остальных) может только возрастать. Исходя из формулы расчета данного индекса, теоретически индекс Хирша при объединении двух организаций варьируется в широких пределах – от максимума двух значений до суммы этих значений.

Анализ показал, что практически подобные ситуации не встречаются. Более того, скорость роста индекса Хирша организаций при их слиянии много меньше ожидаемой. Причина такого положения заключается в том, что индекс Хирша вычисляется на основе рангового распределения числа цитирований публикаций, которые очень часто ведут себя как степенные (аналог широко известного эмпирического закона Ципфа). Как показывают простые статистические проверки на массиве данных РИНЦ для аграрных НИО, данный закон тоже выполняется с хорошей точностью.

Для примера ниже приведена частотная кривая публикаций ВИАПИ имени А.А. Никонова и её приближение соответствующей степенной функцией, выполненное штатными средствами электронных таблиц Excel (рис. 2). Как видно, имеет место очень хорошее ( $R^2=0,9622$ ) приближение фактического распределения первых 80 (по количеству цитирований) публикаций ВИАПИ соответствующей эмпирической степенной функцией.

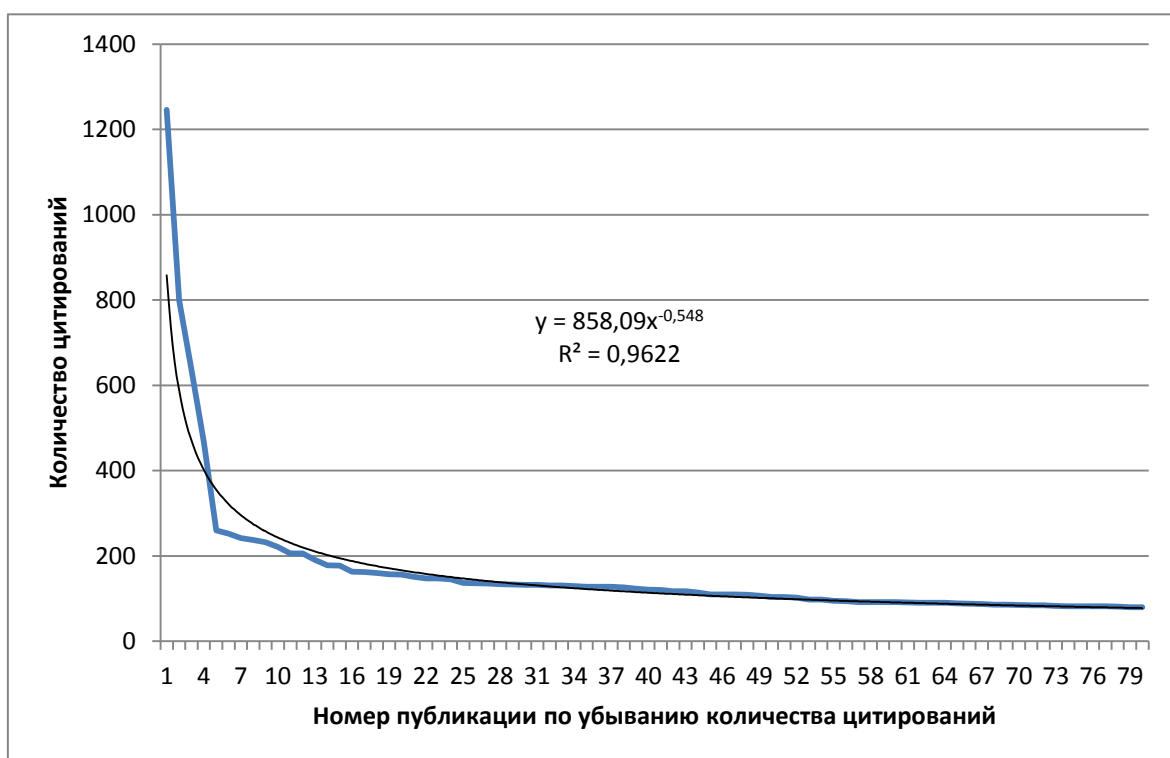


Рис. 2. Распределение публикаций ВИАПИ по частоте цитирования и соответствующее степенное приближение этого распределения

Подобный подход был применен для восстановления данных для тех организаций, которые не имеют в настоящее время самостоятельных страниц в РИНЦ. Сгенерированные таким образом восстановленные данные с использованием статистической среды R явились базой в дальнейшем для последующих расчётов рейтингов НИО. Полный массив восстановленных данных можно посмотреть в [6]. Данные для расчёта рейтингов приводились к интервалу  $[0,1]$ , взвешивались с одинаковыми весами и полученные величины суммировались. По полученным оценкам НИО были ранжированы.

Формализованная методика оценки аграрных НИО на основе их полуактивного цифрового следа выглядит следующим образом.

$P_5^m$  – критерий оценки  $m$ -го НИО на основе их полуактивного цифрового следа;

$d_{hm}^5$  – объем  $h$ -го показателя полуактивного цифрового следа  $m$ -го НИО;

$q_{hm}^5$  – значение  $h$ -го показателя полуактивного цифрового следа  $m$ -го НИО;

$\omega_{hm}^5$  – вес значения  $h$ -го показателя полуактивного цифрового следа  $m$ -го НИО;

$q_{hm}^5 = d_{hm}^5 / \max_m d_{hm}^5$ . Тогда получим

$$P_5^m = \sum_h \omega_h^5 q_{hm}^5. \quad (4)$$

Окончательные результаты приведены в сравнительной таблице 12.

### 3 Методика формирования рейтингов на основе активного цифрового следа

Как было уже отмечено во введении, роль активного цифрового следа помимо традиционной имиджевой роли и роли в виде визитной карточки НИО в современных условиях приобретает существенно большее значение, поскольку при грамотном, научно-обоснованном представительстве НИО в интернете помимо предоставления ученым качественно новых возможностей для широкого обмена идеями между собой и их цифрового взаимодействия с научными ИР обеспечивает также эффективную систему трансфера научных знаний в экономику, способствует росту интеллектуального потенциала общества через совершенствование системы образования [6]. В работе показано также, что такая триединая роль научно-обоснованного представительства НИО в интернете наиболее полно реализуется при формировании единого информационного интернет-пространства научно-образовательных ресурсов, интегрирующего следующие их виды, представленные на сайтах НИО, ВУЗов, информационно-консультационных служб АПК: разработки, публикации, консультационная деятельность, нормативно-правовая информация, дистанционное обучение, пакеты прикладных программ (ППП), базы данных (БД). Именно данные виды представления научных знаний наиболее востребованы в экономике АПК. Данный вывод был сделан на основании обширного обследования 22 регионов России [9, 10, 11].

Кроме информационных научно-образовательных ресурсов (ИНОР) на сайтах указанных организаций в последние годы в той или иной мере начинают развиваться электронные биржи труда (ЭБТ) и торговые площадки (ЭТП), отражающие современные тенденции оказания услуг в интернете в виде данных сервисов. Тогда к проявлениям активного цифрового следа отнесем семь видов ИНОР, а также ЭБТ и ЭТП. Поскольку все ИР на сайтах хранятся в гетерогенном (неоднородном) виде, например, информация об электронной торговой площадке и электронной бирже труда представлена в форме от простейших досок объявлений до полноценных площадок и бирж, то их нужно привести к некоторому однородному виду за счет онтологического моделирования их, использования единого словаря параметров, форматов записи и представления данных и стандартизованной системы классификаторов. Такой подход позволит разработать типовую методику оценки активного цифрового следа НИО и их рейтингов, что, в конечном счете, даст оценки результатов научной деятельности НИО. Такая методика была разработана на основе анализа состояния и объемов ИР, представляющих активный цифровой след, на сайтах аграрных НИО, для чего была подготовлена соответствующая анкета, отражающая 79 показателей деятельности [6].

Приведение к однородному виду заключалось в учете современных тенденций развития интернет-технологий, когда провайдеры начинают предоставлять услуги по хранению контента сайтов в мощных системах управления базами данных (СУБД). ИР при этом могут храниться, с одной стороны, в виде каталогов, либо в виде полноформатного электронного представления, с другой стороны, в виде неупорядоченного списка, либо в виде упорядоченного электронного представления (с возможностью навигации, например, на основе СУБД по тематической рубрикации ГРНТИ, авторам, организациям, ключевым словам и т.д.).

Интегральный критерий оценки активного цифрового следа конкретной НИО определен как сумма взвешенных групп, общая сумма весов которых равна 1, следующих частных критериев: оценки видов представления научных ИР, оценки ИР по состоянию ЭТБ, оценки ИР по состоянию ЭБТ.

Значения весов показателей в виде % частных критериев определены на основе экспертных оценок, анкетирования преподавателей РГАУ-МСХА им. К.А. Тимирязева (таб. 5-10).

Таблица 5. Показатели уровня интеграции ИР

№ п/п	Наименование	Вес
1	Неупорядоченный список	10
2	Упорядоченное электронное представление	90
Итого		100

Таблица 6. Показатели форм хранения ИР

№ п/п	Наименование	Вес
1	Каталог	30
2	Полноформатное электронное представление	70
Итого		100

Таблица 7. Показатели видов представления научных ИР

№ п/п	Наименование	Вес
1	Разработки	30
2	Публикации	20
3	БД	5
4	ППП	5
5	Дистанционное обучение	5
6	Консультанты	30
7	Нормативно-правовая информация	5
Итого		100

Таблица 8. Показатели критерия оценки ИР по состоянию ЭТП

№ п/п	Наименование	Вес
1	Неструктурированная доска объявлений	5
2	Структурированная доска объявлений	10
3	Автоматизация поиска торгового партнера по заданному показателю	20
4	Автоматизация информационных процессов всех торговых операций	25
5	Полная автоматизация электронной торговли	40
Итого		100

Таблица 9. Показатели критерия оценки ИР по состоянию ЭБТ

№ п/п	Наименование	Вес
1	Неструктурированная доска объявлений	10
2	Структурированная доска объявлений	20
3	Электронная биржа труда (автоматизированный поиск)	60
4	Ссылки на другие биржи труда	10
Итого		100

Таблица 10. Частные критерии оценки активного цифрового следа

№ п/п	Наименование	Вес
1	Критерий оценки видов представления научных ИР	70
3	Частный критерий оценки ИР по состоянию ЭТП	15
4	Частный критерий оценки ИР по состоянию ЭБТ	15
Итого		100

Тогда на основании приведенных данных дадим математическое описание методики.

$i$  – код уровня интеграции научных ИР,  $i \in I$  (таб. 5);

$l$  – код формы хранения научных ИР,  $l \in L$  (таб. 6);

$n$  – код вида представления научных ИР,  $n \in N$  (таб. 7);

$m$  – номер НИО,  $m \in M$  ;

$P_j^m$  – частный критерий оценки активного цифрового следа  $m$ -ой НИО по  $j$ -му показателю,  $j \in J$

(таб. 10);

$P^m$  – интегральный критерий оценки активного цифрового следа  $m$ -ой НИО;

$\alpha_i^1$  – вес значения показателя  $i$ -го уровня интеграции научных ИР;

$\alpha_l^2$  – вес значения показателя  $l$ -й формы хранения научных ИР;

$\alpha_n^3$  – вес значения показателя  $n$ -го вида представления научных ИР;

$\beta_j$  – вес  $j$ -ого значения частного критерия оценки активного цифрового следа (таб. 10);

$v_{i\ln}^m$  – объем ИР  $i$ -го уровня интеграции,  $l$ -ой формы хранения,  $n$ -го вида представления  $m$ -ой НИО;

$\lambda_{i\ln}^m$  – значение критерия оценки ИР  $i$ -го уровня интеграции,  $l$ -ой формы хранения,  $n$ -го вида представления  $m$ -ой НИО;

$$\lambda_{i\ln}^m = v_{i\ln}^m / \max_m v_{i\ln}^m ;$$

$d_{sm}^3$  – значение  $s$ -го показателя критерия оценки сайта по состоянию ЭТП  $m$ -ой НИО (таб. 8);

$\omega_s^3$  – вес значения  $s$ -го показателя критерия оценки сайта по состоянию ЭТП (таб. 8);

$d_{gm}^4$  – значение  $g$ -го показателя критерия оценки сайта по состоянию ЭБТ  $m$ -ой НИО (таб. 9);

$\omega_g^4$  – вес значения  $g$ -го показателя критерия оценки сайта по состоянию ЭБТ (таб. 9). Тогда:

$$P^m = \sum_j \beta_j P_j^m, \text{ где } P_1^m = \sum_{i,l,n} \lambda_{i\ln}^m \alpha_i^1 \alpha_l^2 \alpha_n^3, P_3^m = \sum_s \omega_s^3 d_{sm}^3, P_4^m = \sum_g \omega_g^4 d_{gm}^4. \quad (5)$$

Результаты расчётов рейтингов НИО по оценке активного цифрового следа первых десяти НИО представлены в таб. 11, где под номерами столбцов понимаются следующие показатели: 1 – разработки, 2 – публикации, 3 – базы данных, 4 – ППП, 5 – дистанционное обучение, 6 – консультации, 7 – нормативно-правовая информация, 8 – ЭТП, 9 – ЭБТ, 10 – интегральный рейтинг. Первые места со значительным отрывом заняли ВНИИ экономики сельского хозяйства, ВНИТИ птицеводства и НИИ садоводства Сибири имени М.А. Лисавенко.

При этом 40 НИО получили нулевую оценку [6], в частности, широко известные: ВНИИ кукурузы, ВНИИ овцеводства и козоводства, ВНИИ селекции и семеноводства овощных культур, ВНИИ ветеринарной вирусологии и микробиологии. В основном в данную группу входят региональные НИО, что отражает отношение региональных органов власти к науке и трансферу научных знаний в экономику своих регионов. Заметим, что ВИАПИ им. А.А. Никонова по этому критерию занял только тридцать шестое место, в отличие от оценки пассивного цифрового следа, по которой он занял первое место. Это связано с исключением с сайта по распоряжению директора шести видов представления аграрных знаний, за исключением публикаций, объем которых и ‘вытянул’ НИО до десятого места.

Таблица 11. Рейтинги по оценке активного цифрового следа первых 10 НИО

НИУ	Частные рейтинги									10
	1	2	3	4	5	6	7	8	9	
ВНИИЭСХ	2	2	3	2	93	21	111	53	109	1
ВНИТИ птицеводства	64	90	101	100	93	1	111	4	109	2
НИИСС	1	55	101	100	93	104	111	4	109	3
ГОСНИТИ	4	7	101	100	93	104	11	53	109	4
ВИК им. В.Р. Вильямса	51	37	101	100	93	3	12	53	109	5
СибФТИ	12	21	1	1	93	104	111	4	109	6
СИБНИИП	10	160	101	100	93	2	111	53	109	7
ИАЭП	71	6	101	100	93	14	111	53	109	8
ВНИИМП	14	4	101	100	93	104	9	53	109	9
ВИАПИ	165	5	101	100	93	104	111	142	18	10

#### 4 Сравнение рейтингов НИО на основе пассивного (R1), полуактивного (R2) и активного (R3) цифровых следов

В таб. 12 приведены результаты расчетов рейтингов НИО на основе пассивного (R1), полуактивного (R2) и активного (R3) цифровых следов, а также полученные интегральные рейтинги (R4), в которых первые места со значительным отрывом заняли ВНИИ экономики сельского

хозяйства (оценка – 0,79), ВНИТИ птицеводства (оценка – 0,60) и НИИ садоводства Сибири имени М.А. Лисавенко (оценка – 0,54). Последние места заняли: Ивановский НИИСХ (оценка - 0,03), Дальневосточный НИИ защиты растений (оценка - 0,02), Челябинский НИИСХ (оценка - 0,02).

Таблица 12. Рейтинги первой десятки НИО на основе пассивного (R1), полуактивного (R2) и активного (R3) цифровых следов, а также их интегральные рейтинги

НИО	R1	R2	R3	R4	НИО	R1	R2	R3	R4
ВНИИЭСХ	9	1	1	1	СибФТИ	33	70	8	6
ВНИТИ птицеводства	23	30	2	2	СИБНИИП	100	46	4	7
НИИСС	54	149	3	3	ИАЭП	11	8	16	8
ГОСНИТИ	16	12	5	4	ВНИИМП	10	102	11	9
ВИК им. В.Р. Вильямса	42	9	7	5	ВИАПИ	1	2	36	10

Анализ результатов сравнения показывает, что большинство НИО практически не уделяют внимания своему имиджу. Так, из лидеров рейтинга R4 из первой десятки попали в лидеры первой же десятки рейтинга R1 лишь четыре НИО: ВНИИЭСХ, СибФТИ, СИБНИИП и ВИАПИ. Аналогично, лишь четыре НИО из лидеров рейтинга R4 попали в лидеры первой же десятки рейтинга R2. Наиболее яркий пример, как указывалось выше – ВИАПИ им. А.А. Никонова, который по критерию R3 занял только тридцать шестое место, связанное с волонтеристским решением директора НИО Петрикова А.В. об исключении с сайта большинства видов представления аграрных знаний, за исключением публикаций. Хотя и это уже большой прогресс, поскольку многие годы он требовал от разработчиков сайта лишь форму витринного, когда на сайте отображается информация о контактах, структуре НИО, да новостная лента.

## Заключение

Исследования показали, что если в настоящее время ВУЗы нормативно вынуждены развивать свои сайты, уделяя особое внимание лишь тем ресурсам, которые требуют регулирующие органы и которые оставляют в интернете в основном полуактивный цифровой след, то к сайтам НИО никаких требований нет. Более того, как показали исследования, прошедшая реструктуризация НИО в некоторых случаях оказывает отрицательное воздействие на состояние их сайтов, их информационное наполнение. Объединяющие под своим научным руководством ФИЦ и ФНЦ, в лучшем случае, поддерживают свои прежние сайты, на которых лишь располагают краткую информацию о присоединившихся НИО (руководство, структура, контакты). При этом сайты присоединившихся НИО чаще всего не поддерживаются или вообще ликвидируются, а находящаяся на них информация о публикациях, научных разработках и т.д. не переносится на сайты головных организаций. В результате ценная информация о научно-исследовательской деятельности НИО не доходит до потребителя. Показательным в этом плане является отношение руководства РАН к представительству в интернете самой академии. Так, на сайте РАН ([http://www.ras.ru/win/db/show\\_org.asp?P=.oi-3017.ln-ru.dl-pr-org.uk-10](http://www.ras.ru/win/db/show_org.asp?P=.oi-3017.ln-ru.dl-pr-org.uk-10)) на апрель 2021г. отделением сельскохозяйственных наук представлена информация о подведомственных организациях. И хотя за последние годы произошла реорганизация НИО, на сайте последние обновления датированы 2016-2017гг. со старым списком НИО. Нет научных центров. Поиск организаций, тематики их и персон не работает.

## Литература

1. Site-Auditor. URL: <https://freesoft.ru/windows/siteauditor> (дата обращения: 18.02.2021).
2. Вузы из проекта «5–100» так и не вошли в топ-100 международных рейтингов. URL: [https://www.rbc.ru/society/18/02/2021/602cbdf9a7947765cbb58e5?from=from\\_main\\_10](https://www.rbc.ru/society/18/02/2021/602cbdf9a7947765cbb58e5?from=from_main_10) (дата обращения: 18.02.2021).
3. Российские ученые должны быть обеспечены всеми необходимыми информационными ресурсами. [Электронный ресурс]. URL: [https://minobrnauki.gov.ru/ru/press-center/card/?id\\_4=2241](https://minobrnauki.gov.ru/ru/press-center/card/?id_4=2241) (дата обращения 01.05.2021).
4. The R Project for Statistical Computing. URL: <https://www.R-project.org/> (дата обращения 01.05.2021).
5. Мاستицкий С.Э., Шутиков В.К. Статистический анализ и визуализация данных с помощью R. – Электронная книга, адрес доступа: <http://r-analytics.blogspot.com/>
6. Меденников В.И., Муратова Л.Г., Сальников С.Г. Эффективность использования информационных интернет-ресурсов научно-исследовательских учреждений аграрного направления. – М.: Аналитик, 2017. – 237с.

7. *Stef van Buuren, Karin Groothuis-Oudshoorn*. mice: Multivariate Imputation by Chained Equations in R // Journal of Statistical Software. 2011. 45(3). P. 1-67.
8. *Светуньков И.С., Светуньков С.Г.* Методы и модели социально-экономического прогнозирования. – М.: Юрайт, 2014. Т.1 – 351с.
9. *Меденников В.И., Муратова Л.Г., Сальников С.Г.* Модели и методы формирования единого информационного интернет-пространства аграрных знаний. – М.: ГУЗ, 2014. – 426с.
10. *Zatsarinny Alexander, Ereshko Felix, Medennikov Viktor*. Scientific and methodological approaches to the generation of the internet information space of scientific and educational resources. // Proceedings of the 1st International Conference of Information Systems and Design Moscow, Russia, December 5, 2019.
11. *Medennikov Viktor, Raikov Alexander*. Creating the requirements to the national platform "Digital Agriculture" // Proceedings of the 8th International Scientific Conference on Computing in Physics and Technology. Moscow region, Russia, November 09-13, 2020. (Scopus: CEUR Workshop Proceedings, 2020, 2763). С. 13-18. doi: [https://dx.doi.org/10.30987/conferencearticle\\_5fce27715a3742.47428784](https://dx.doi.org/10.30987/conferencearticle_5fce27715a3742.47428784).