

SEO И IR. ОШИБКИ КОТОРЫХ УДАЛОСЬ ИЗБЕЖАТЬ, И КОТОРЫХ НЕ УДАЛОСЬ

SEO AND IR. MISTAKES THAT WE MANAGED AND DID NOT MANAGE TO AVOID

Ефимов Александр Александрович – заместитель начальника отдела информационно-аналитического сопровождения Уральского федерального университета (Российская Федерация), e-mail: alexander.efimov@urfu.ru

Efimov Alexander – deputy of head, Department of information and analytical support, Ural Federal University (Russian Federation), e-mail: alexander.efimov@urfu.ru

Аннотация: В статье автор приводит пример ситуации, при которой правильно настроенный и наполняемый институциональный репозиторий, может значительно потерять позиции в отраслевых поисковых системах из-за локальных кратковременных проблем. Автор также указывает на то, что кратковременные проблемы могут не только привести к потере результата, но и потребовать очень много времени на возвращение показателей к ожидаемому уровню.

Abstract: In the article author gives examples of the situation when correctly setted and constantly filled institutional repository may considerably loose positions in special search engines due to local short-term problems. Author also points out that these problems may not only cause the loss of results, but also require a lot of time to return the indexes to the expected level.

Ключевые слова: репозиторий, поисковая оптимизация, поисковая система, агрегатор, высокая доступность, временной штраф, сообщество, опыт.

Keywords: SEO, IR, CORE, repository, Search Engine, high availability, community, experience.

Институциональный репозиторий – электронный архив для длительного хранения, накопления и обеспечения долговременного и надежного открытого доступа к результатам научных исследований, проводимых в учреждении [1]. Подразумевая долгосрочное накопление и хранение информации в такой консервативной организации как университет, мы автоматически получаем достаточно консервативный сервис, отражающий научную, публикационную и иные активности университета. При этом, несмотря на понятные требования стабильности, преемственности, защищённости и пр., к сервису также предъявляются требования и по актуальности механизмов агрегирования контента, удобству редактирования данных и метаданных, эффективному представлению данных и метаданных.

Всё это вызывает необходимость время от времени пересматривать методики работы с сервисом, проводить обновление как сервисного программного обеспечения, так и операционной системы, и операционной среды, проводить аудит контента, аудит процессов агрегирования и их результатов. Данная статья будет посвящена вопросам реализации поисковых оптимизаций (SEO) - комплекса мероприятий по внутренней и внешней оптимизации сервиса для поднятия позиций в результатах выдачи поисковых систем по определённым запросам пользователей, с целью увеличения видимости контента [2].

Несмотря на наличие отличной документации по системе управления контентом (DSpace¹) на английском языке и очень дружного интернационального сообщества пользователей, порой трудно реализовывать данные задачи полностью самостоятельно и в рабочем порядке, да еще и с минимальными простоями сервиса или вовсе без них. Данная статья описывает опыт реализации описанных выше задач, который позволил сотрудникам, поддерживающим репозиторий, приобрести необходимые навыки, обнаружить проблемы, нуждающиеся в исправлении, уточнить методику, лучше понять и документировать сервис и его место в сети.

Описание процесса получения опыта, разделено на части, каждая из которых отражает определённый этап развития институционального репозитория в контексте поисковых оптимизаций. Описание происходит на примере Электронного архива Российского государственного профессионально-педагогического университета (РГППУ)².

Часть первая – вводная

РГППУ³ – небольшой региональный отраслевой ВУЗ, основанный в 1979 году. В вузе обучается около четырех тысяч студентов [3]. Репозиторий запущен в 2014-м году, динамика наполнения (количество новых заглавий в год) следующая:

Таблица 1. Динамика наполнения

Годы	2014	2015	2016	2017	2018	2019	2020	ИТОГО
Заглавия	1 307	5 498	9 012	6 421	4 532	3 041	2 930	32 741

Как можно заметить, репозиторий небольшого вуза наполнялся достаточно активно. Спад активности после 2018 года во многом связан с исчерпанием контента. Архивы сборников конференций, журналов, а также авторефераты и пр. «источники заглавий» были обработаны и опубликованы, для публикации выпускных квалификационных работ был реализован отдельный сервис, а статистика за последние несколько лет отражает лишь актуальную

¹ <https://wiki.lyrasis.org/display/DSPACE/>

² <https://elar.rsvpu.ru/>

³ <https://www.rsvpu.ru/about/>

публикационную активность вуза. Т.е. «ретроввод» фактически завершен пять лет назад.

Все штатные средства для SEO, предоставляемые DSpace, использовались и продолжают использоваться в полной мере. Сервис имеет актуальную карту сайта (sitemap), актуальный файл с инструкциями для поисковых машин (robots.txt), описание полей метаданных для сервисов Google (google-metadata.properties). В DSpace введены собственные реестры метаданных, позволяющие лучше описать контент, сервис также интегрирован с системами аналитики и инструментов вебмастера ведущих поисковых систем (Google, Yandex, Bing).

Таблица 2. Перечень полей дополнительного реестра метаданных

local.conference.date	local.identifier.rsi
local.conference.name	local.identifier.temp
local.contributor.department	local.identifier.wos
local.contributor.subdepartment	local.subject.oksvnk
local.description.firstpage	local.identifier.source
local.description.lastpage	local.thesis.degree
local.description.order	local.thesis.discipline
local.fund.rffi	local.thesis.level
local.fund.rgnf	local.thesis.speciality
local.identifier.affiliation	local.thesis.specialitycode
local.identifier.doi	local.thesis.specialityname
local.identifier.eid	local.issue
local.identifier.other	local.volume

```
<meta name="citation_keywords" content="STUDENTS WITH DISABILITIES AND SPECIAL EDUCATION NEEDS; INCLUSIVE EDUCATION;" />
<meta name="citation_title" content="Профессионально-педагогический консилиум в вузе" />
<meta name="citation_publisher" content="РГПУ" />
<meta name="citation_journal_title" content="Образование и наука" />
<meta name="citation_author" content="Malyarchuk, N. N." />
<meta name="citation_author" content="Krinitzyna, G. M." />
<meta name="citation_author" content="Pashchenko, E. V." />
<meta name="citation_author" content="Pivnenko, V. V." />
<meta name="citation_author" content="Малыарчук, Н. Н." />
<meta name="citation_author" content="Криницына, Г. М." />
<meta name="citation_author" content="Пашенко, Е. В." />
<meta name="citation_author" content="Пивненко, В. В." />
<meta name="citation_date" content="2019-04" />
<meta name="citation_abstract_html_url" content="https://elar.rsvpu.ru/handle/123456789/29620" />
<meta name="citation_issue" content="3" />
<meta name="citation_firstpage" content="175" />
<meta name="citation_issn" content="2310-5828" />
<meta name="citation_doi" content="10.17853/1994-5639-2019-3-175-193" />
<meta name="citation_language" content="ru" />
<meta name="citation_pdf_url" content="https://elar.rsvpu.ru/bitstream/123456789/29620/1/edscience_2019_3_010.pdf" />
<meta name="citation_lastpage" content="193" />
```

Рис. 1. Маппирование метаданных google-metadata.properties

Казалось бы, такие базовые вещи, о которых не стоило и говорить отдельно, должны быть реализованы в каждом институциональном репозитории, но к сожалению, это далеко не так.

Также репозиторий вуза отлично отражался во всех отраслевых агрегаторах, т.к. был зарегистрирован там практически с момента старта, а доступность и описание контента были на отличном уровне.

Часть вторая - динамическая

В результате последовательных и продуманных действий по настройке сервиса и наполнению его контентом к середине 2017 года репозиторий включал в себя более 20 тысяч заглавий, около 18 тысяч из которых индексировались в Академии Google. Результат очень достойный.

К сожалению, в середине 2017 года в информационной инфраструктуре вуза происходила реформация и сервис мог быть недоступен для поисковых систем в течение длительного времени. Результатом недоступности сервиса извне, и наличия копий контента на стороне агрегаторов стала ситуация, отраженная на графике:

Верхняя линия отражает зависимость количества заглавий в кеше Google Search от года, нижняя – в кеше Google Scholar в зависимости от года. Т.е. мы видим, что на 2018 год вхождение заглавий в Google Scholar отсутствовало. Для институционального репозитория это очень нехорошая практика. Нижняя линия продолжается и после 2018 года и видна тенденция к исправлению положения, об этом следующая часть.

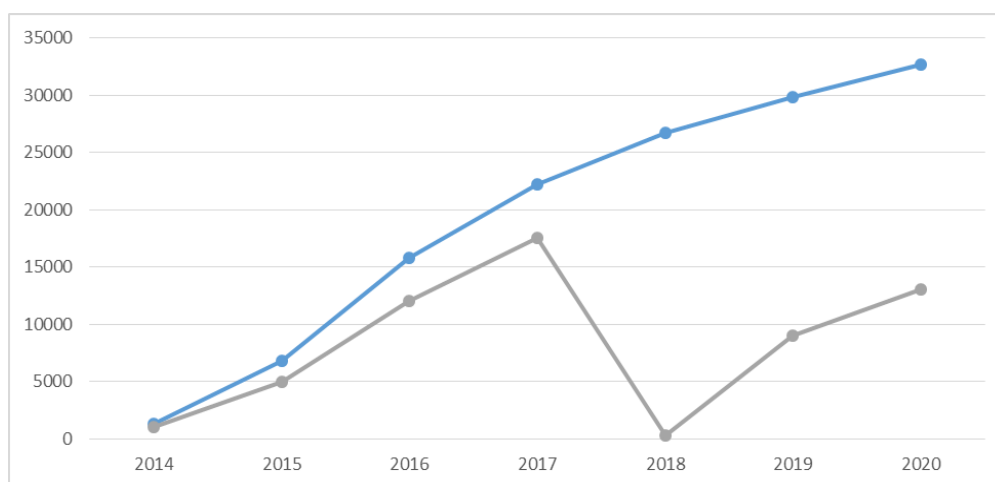


Рис. 2. Динамика вхождения контента в поисковые системы

Часть третья – формализация опыта

Академия Google не имеет явного API, не предоставляет поддержки пользователям и организациям, но тем не менее, доступны контакты, по которым можно задать некоторые вопросы. После анализа ситуации и общения с коллегами – представителями Академии Google и отраслевых агрегаторов выяснилось следующее – в то время как институциональный репозиторий был регулярно и в течение длительного времени недоступен для индексирования поисковыми машинами, интенсивность обхода домена вуза была уменьшена, и приоритет как в индексировании, так и в поисковой выдаче отдавался альтернативным источникам. В нашем случае таким источником был

агрегатор CORE¹. Фактически 100% оригинальных заглавий, ранее доступных в домене rsvpu.ru перешли в домен core.ac.uk.

Представители Академии Google предложили исправить ошибки на стороне сервиса, обеспечить его высокую доступность и ждать, пока ситуация исправится сама по себе. Представители CORE вошли в наше положение, убрали из публичного доступа в своём домене полные тексты, полученные из институционального репозитория, и провели еще ряд поисковых оптимизаций на своей стороне, которые показывали поисковым системам, где находится оригинальный контент. Хотелось бы отметить очень оперативную работу, участие и глубокое погружение в проблему представителей CORE, в особенности Ненси Понтика (Nancy Pontika)². Благодаря этому команда CORE приобрела новый опыт во взаимодействии с сообществом, получила иллюстрацию возможной проблемы и разработала пути решения, а вуз, институциональный репозиторий которого потерял позиции в поисковых системах, смог оперативно восстановить позиции в Google Search, но не в Google Scholar.

В результате данного инцидента в сообществе эксплуатантов институциональных репозиториях прошло обсуждение, в результате которого многие участники проекта CORE попросили прекратить публикацию полных текстов в домене core.ac.uk, но позволили, как и прежде агрегировать полные тексты для нужд сервисов CORE без публичного к ним доступа. Агрегация и публикация метаданных при этом производятся, как и прежде. На стороне «пострадавшего» институционального репозитория РГППУ был проведен переход на HTTPS, дописаны новыми метаданными некоторые ранее загруженные материалы, а также были проведены иные работы, которые объективно обновляли страницы с метаданными и требовали обновления карты сайта (sitemap).

Результатом данных действий стало полное восстановление результатов в Google Search в течение двух недель, а динамика восстановления результатов в Google Scholar куда менее динамична.

В завершении хотелось бы поблагодарить коллег, вникнувших в проблему и сделавших всё для скорейшего устранения последствий, и посоветовать сообществу учиться на чужом опыте!

СПИСОК ЛИТЕРАТУРЫ

1. Поисковая оптимизация [Электронный ресурс] // Википедия, свободная энциклопедия. – Сан-Франциско, 2021. – Режим доступа: <https://ru.wikipedia.org/?curid=186839&oldid=113951204>. – Дата доступа: 02.05.2021.

2. Институциональный репозиторий [Электронный ресурс] // Википедия, свободная энциклопедия. – Сан-Франциско, 2021. – Режим доступа: <https://ru.wikipedia.org/?curid=3825945&oldid=113801833>. – Дата доступа: 24.04.2021.

¹ <https://core.ac.uk/>

² <http://kmi.open.ac.uk/people/member/nancy-pontika>

3. Ефимов, А. А. Повышение видимости репозитория. SEO и отраслевые агрегаторы / А. А. Ефимов // Электронный научный архив УрФУ. – 2004–2021. – Режим доступа: <https://elar.urfu.ru/handle/10995/94122>. – Дата доступа: 12.04.2021.