

КАЗИТУ

**Казахстанский инженерно-
технологический университет**



**Темирбеков Н.М.
Байгереев Д.Р.
Темирбеков А.Н.
Жолшиева К.А.**

**ИНТЕГРИРОВАННАЯ
РАСПРЕДЕЛЕННАЯ
ИНФОРМАЦИОННАЯ СИСТЕМА
АКАДЕМГОРОДКА ГОРОДА АЛМАТЫ**

Учебное пособие

Алматы, 2019

КАЗАХСТАНСКИЙ ИНЖЕНЕРНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

Н. М. Темирбеков, Д. Р. Байгереев, А. Н. Темирбеков,
К.А.Жолшиева

Интегрированная распределенная
информационная система Академгородка
города Алматы

Учебное пособие

Алматы, 2019

УДК 004.75 (075.8)
ББК 32.923.26.я73
И 73

Рецензенты:

Кандидат физико-математических наук, доцент, директор ТОО «Академсеть»

Т. М. Такабаев

Кандидат физико-математических наук, доцент,

декан факультета информационных технологий КазНУ им. аль-Фараби

Б. А. Урмашев

Кандидат технических наук, начальник методического отдела Департамента по академическим вопросам Международного университета информационных технологий

Ф. Н. Абдолдина

Рекомендовано к изданию Ученым советом КазИТУ
Протокол № 2 от 25.09.2019 г.

ISBN 978-601-332-452-4

И 73 Интегрированная распределенная информационная система
Академгородка города Алматы / Н. М. Темирбеков, Д. Р. Байгереев,
А. Н. Темирбеков, К.А. Жолшиева– Алматы: КазИТУ, 2019. – 73 с.

В учебном пособии описан процесс проектирования, разработки, внедрения, адаптации и сопровождения интегрированной распределенной информационной системы Академгородка города Алматы, приведены инструкции по наполнению контентом и рекомендации по эффективной эксплуатации информационной системы.

Учебное пособие предназначено для бакалавров и магистрантов IT-специальностей, профессорско-преподавательского состава, широкого круга специалистов, занимающихся вопросами разработки, внедрения и эксплуатации интегрированных распределенных информационных систем.

© КазИТУ, 2019

ПРЕДИСЛОВИЕ

В данном учебном пособии изложен процесс проектирования, реализации, наполнения и использования интегрированной распределенной библиотечной информационной системы Академгородка города Алматы.

Достаточно подробно описана архитектура информационной системы, приведено описание ее подсистем, а также программного обеспечения, положенного в их основу. При выборе программного обеспечения проанализированы литературные источники. В частности, проанализирован опыт российских ученых по созданию научно-образовательной информационной системы и облака «Электронной библиотеки Сибирского отделения РАН», которая сегодня называется Порталом СО РАН (<http://www.sbras.ru/>); опыт ученых Канады, Новой Зеландии, Чехии и Шри-Ланки по внедрению и эксплуатации систем, используемых в качестве репозитория цифровых объектов.

Создаваемая информационная система ориентирована на обеспечение потребностей участников научно-образовательного кластера, основана на передовых информационных технологиях и будет реализована на основе свободно распространяемого программного обеспечения и уникальных программных модулях, разработанных в процессе выполнения Проекта.

В учебном пособии приведено описание следующих составляющих:

1. Анализ возможностей и потребностей всех участников научно-образовательного кластера для построения оптимальной архитектуры распределенной информационной системы (РИС).

2. Определение оптимальной архитектуры РИС, функциональности и номенклатуры всех ее компонентов.

3. Отбор и адаптация к выбранной архитектуре необходимого свободно распространяемого программного обеспечения.

4. Создание необходимых уникальных программных компонент, необходимых для функционирования РИС в выбранной архитектуре.

5. Создание полнофункционального прототипа РИС на базе ограниченного числа участников.

6. Оцифровка наиболее важных научных трудов указанных научно-исследовательских институтов, накопленных десятилетиями

(за всю историю существования).

7. Наполнение РИС актуальной информацией и информацией, созданной в процессе оцифровки документальных источников.

8. Опытная эксплуатация прототипа РИС.

9. Создание полного комплекта документации по РИС и отдельным ее компонентам.

10. Обеспечение научных исследований современными компьютерными ресурсами через Центр обработки данных помощью радиотехнологий WiMax и WLL.

В первом разделе приведены предпосылки и цели создания информационной системы, приведен обзор аналогичных ресурсов, функционирующих в Казахстане и зарубежом, сделан анализ возможностей и потребностей участников научно-образовательного кластера для построения оптимальной архитектуры распределенной информационной системы. Данный раздел пособия может быть полезен разработчикам библиотечных информационных систем, институциональных репозиториев, а также инженерам, проектирующим распределенные информационные системы.

Во втором разделе описан процесс проектирования, реализации и интеграции информационной системы. Приведено описание аппаратной части системы, определена оптимальная архитектура системы, описание функциональности и номенклатуры всех ее компонент. Особое внимание уделено аспектам внутренней и внешней интеграции системы. Данный раздел может быть полезен инженерам распределенных информационных систем, поскольку охватывает широкий круг вопросов от постановки требования к системе до ее реализации. Раздел сопровождается диаграммами, схемами данных и другими иллюстрациями.

Третий раздел может быть полезен для сотрудников научно-исследовательских институтов, занимающихся наполнением системы цифровым контентом. В разделе подробно приведена инструкция по заполнению метаданных ресурсов и загрузке цифровых объектов. Приведено описание административной части системы.

В последнем разделе приведена инструкция для пользователей по использованию созданной информационной системы: ученых научно-исследовательских институтов, профессорско-преподавательского состава и обучающихся высших учебных заведений и других исследователей, интересующихся проблемами агропромышленной отрасли.

Созданная информационная система призвана сохранить и популяризовать научные труды казахстанских ученых в области аграрной науки. Система в полной мере обеспечивает необходимыми вычислительными ресурсами проводимые научные исследования и образовательные процессы, упрощая перспективу дальнейшего ее развития. Информационная система позволяет построить передовую IT-инфраструктуру, управляющую интеллектуальным капиталом, электронной библиотекой, в которой будут храниться все книги и научные труды научно-исследовательских институтов.

1 НАЗНАЧЕНИЕ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ, АНАЛИЗ ВОЗМОЖНОСТЕЙ И ПОТРЕБНОСТЕЙ УЧАСТНИКОВ НАУЧНО-ОБРАЗОВАТЕЛЬНОГО КЛАСТЕРА ДЛЯ ПОСТРОЕНИЯ ЕЕ ОПТИМАЛЬНОЙ АРХИТЕКТУРЫ

1.1 Актуальность создания информационной системы для агропромышленной отрасли

В эпоху информационных технологий информация оказывает существенное влияние на направление развития в научно-технической, экономической, социально-культурной и других сферах жизни любого сообщества, государства или организации. В настоящее время информация является одним из наиболее значимых ресурсов, сохранение, рациональное использование и развитие которой является одним из стратегических направлений.

На протяжении нескольких десятилетий в Академгородке города Алматы функционируют несколько научно-исследовательских институтов, которые проводят научные исследования в различных областях агропромышленного направления: институт физиологии человека и животных, Казахский научно-исследовательский институт перерабатывающей и пищевой промышленности, Казахский научно-исследовательский институт плодоводства и виноградарства, Казахский научно-исследовательский институт почвоведения и агрохимии имени У. Успанова, Казахстанский инженерно-технологический университет, институт зоологии, а также научно-исследовательский институт микробиологии и вирусологии, институт общей генетики и цитологии, институт сейсмологии и другие. Значительные объемы информации, полученной в результате исследований данных институтов, их непрерывное увеличение и разнородный по многим признакам характер хранения и распространения, отсутствие унифицированного доступа к ним создают существенные проблемы их эффективного использования. Данные проблемы приводят к необходимости поиска новых подходов и решений проблем создания хранилищ информационных ресурсов, их организации, средств и способов доступа к ним пользователей. Такие подходы сегодня называют «цифровыми» или «электронными» библиотеками [1, 2].

Одним из актуальных направлений является создание и применение распределенных компьютерных систем для интенсивной

работы с данными, что актуально как для решения новых научных задач, использование больших объемов научных данных, генерируемых современными измерительными средствами, так и экономических и социальных проблем на основе больших данных и технологии их обработки и анализа. Одним из основных результатов научной деятельности является создание и накопление опыта предыдущих поколений.

Существует достаточно много технологических разработок информационных систем для электронных библиотек. Наиболее известными в России являются, например, Euro CRIS (eurocris.org), eLibrary (elibrary.ru), Информика (informika.ru), MathNET (mathnet.ru). В Казахстане функционирует Республиканская межвузовская электронная библиотека (РМЭБ) на базе университетских образовательных ресурсов по единой технологии, предложенной Ассоциацией вузов Республики Казахстан и научно-образовательной компьютерной сетью Казахстана KAZRENA. Одним из пионеров, и как следствие, одним из мировых лидеров в развитии научно-образовательной информационной системы является «Электронная библиотека Сибирского отделения РАН», который сегодня называется Порталом СО РАН (sbras.ru). Данный ресурс по данным Webometrics занимает первое место в России среди научных организаций и 45-е место в мире. Однако основной недостаток большинства систем – ограниченность интеграции ресурсов как внутри каждой из них, так и с внешними системами. Серьезной проблемой является идентификация информационных ресурсов, позволяющая получать библиографические сведения, а также устанавливать связи определенного ресурса с другими фактами и объектами.

Использование информационных технологий позволило агропромышленному комплексу выйти на новый уровень, повысить эффективность и решить стратегически важные вопросы. Во-первых, перспективными становятся технологии для дистанционного зондирования земли, точного земледелия, технологии визуализации космических данных на карте, оперативного доступа к высокодетальным снимкам для точного картографирования границ полей и севооборотов, зон плодородия, а также мониторинга состояния вегетации [1]. При этом используются данные со спутников, датчиков и других систем.

Во-вторых, актуальны системы онлайн-контроля элементов

производственного цикла и управления агробизнесом: техники, полей и складов. Данные системы обрабатывают данные о погоде, индексе вегетации, состоянии почвы и севообороте. Например, российский проект «Агросигнал», к системе которого подключено свыше 150 хозяйств, обрабатывающих более 2 млн гектаров земли, сосредоточен в основном на контроле за посевами.

В-третьих, немаловажную роль играют специализированные информационные сервисы, содержащие каталоги и базы знаний, а также данные наблюдений. Данные сервисы нередко используются для публикации новостей и аналитических статей, позволяют работникам агропромышленного направления обсудить острые вопросы и поделиться опытом.

Каждое из перечисленных направлений, как правило, приводит к необходимости эффективного хранения полученного массива данных.

В Российской Федерации существует несколько десятков информационных сервисов, посвященных агропромышленному комплексу. Например, существуют порталы agroyug.ru, agrofoodinfo.com, webpticeprom.ru, agbz.ru, latifundist.com, agroserver.ru, rynok-apk.ru, agrobook.ru, agro.ru, ya-fermer.ru, agro2b.ru, agroxxi.ru, предназначенные для лиц, чья деятельность связана с сельским хозяйством и сопутствующими товарами и услугами. Данные порталы распространяются на клиентов из России, Беларуси, Молдовы, Украины, Китая, Европы и дальнего зарубежья. Сайты содержат каталоги компаний, производителей, товаров и услуг, прайс-листы, мероприятия и статьи. Таким образом, данные информационные сервисы обеспечивают население качественным контентом в сфере агропромышленного комплекса, что помогает решить множество проблем производителей и потребителей агропромышленного комплекса.

В Республике Казахстан информационных сервисов, посвященных IT-решениям в агропромышленном комплексе крайне мало. Сайт информационно-аграрной газеты «АгроИнфо» agroinfo.kz предоставляет информацию по агрономии, животноводству, о новинках сельхозтехники и оборудования, разъяснения законодательства, обзор зернового рынка, ответы на вопросы читателей.

Портал agrobiz.kz направлен на повышение эффективности системного взаимодействия фермерского сообщества с

государственными органами, организациями, финансовыми институтами, сектором рынка и населением. Идея портала – предоставить фермерам из отдаленных регионов прямой доступ к открытому рынку, разносторонне увеличить спрос на их продукцию, преградить путь многочисленным посредникам. Портал призван стать единой базой данных всех крестьянских хозяйств и сельских округов, сотен ученых-аграриев и экспертов, тысяч агропредприятий страны. Существует портал agro.kz для публикации объявлений в аграрном секторе.

Признавая необходимость создания единого научно-образовательного кластера, общим решением руководств указанных научно-исследовательских институтов и Казахстанского инженерно-технологического университета была поставлена цель создать интегрированную распределенную библиотечную информационную систему Академгородка города Алматы, позволяющую сохранить результат интеллектуальной деятельности приведенных выше научно-исследовательских институтов в актуальном виде и предоставлять к ним доступ на основе Web-технологий.

Для проведения совместных научных исследований в данном направлении заключен Меморандум об информационном сотрудничестве между Казахстанским инженерно-технологическим университетом, научно-исследовательскими институтами, расположенными в Академгородке г. Алматы (НИИ почвоведения и агрохимии имени У. Успанова, НИИ перерабатывающей и пищевой промышленности, НИИ плодоводства и виноградарства, НИИ микробиологии и вирусологии, НИИ физиологии человека и животных, НИИ общей генетики и цитологии) и ТОО «Академсеть». Заключение Меморандума нацелено на разработку и введение в действие распределенной информационной системы в сфере практико-ориентированных образовательных услуг и научных разработок путем активного партнерского взаимодействия, в том числе на базе цифровых технологий путем объединения информационно-коммуникационных технологий инфраструктуры Партнеров в единую сеть.

В настоящем учебном пособии поставлена цель описать архитектуру, реализацию и инструкции по использованию интегрированной распределенной библиотечной информационной системы Академгородка города Алматы, позволяющей сохранить результат интеллектуальной деятельности научно-исследовательских

институтов МОН РК, МСХ РК и КазИТУ, расположенных в Академгородке г. Алматы. Создаваемая информационная система ориентирована на обеспечение потребностей участников научно-образовательного кластера, основана на передовых информационных технологиях и будет реализована на основе свободно распространяемого программного обеспечения и уникальных программных модулей.

Стоит отметить, что исследования соответствуют целям, поставленным в рамках Государственной программы «Цифровой Казахстан» на 2017-2020 годы и нацелена на выполнение следующих прикладных задач:

- развитие креативного общества (развитие человеческого капитала путем повышения цифровой грамотности населения, повышение квалификации специалистов в области инфокоммуникационных технологий, развитие креативного мышления и т.д.);

- цифровые преобразования в отраслях экономики (развитие цифровой индустрии путем, автоматизации транспортно-логистической системы страны; внедрения цифровых технологий в сфере сельского хозяйства; пищевой промышленности).

1.2 Определение потребностей участников научно-образовательного кластера

Для создания базовой концепции и предварительной архитектуры создаваемой интегрированной распределенной информационной системы директорам институтов были направлены письма за следующей информацией:

- наличие информационных ресурсов;
- структуры этих ресурсов и технологий доступа к ним;
- наличие серверного и коммутационного оборудования;
- наличие серверных платформ;
- создание цифрового репозитория (полные тексты, сканирование и прочее).

К создаваемой информационной системе выдвинуты следующие требования:

1. Обеспечение надежного долговременного хранения следующих типов информационных ресурсов с сохранением всех

смысловых и функциональных характеристик исходных документов:

- полнотекстовые материалы, в том числе результаты научной деятельности институтов, не имеющие цифровых копий;

- географические материалы (карты, спутниковые снимки, данные полевых наблюдений и т. п.), а также соответствующие базы метаданных;

- аудио и видеозаписи, электронные презентации и др.;

- библиографические базы данных и электронные каталоги;

- авторитетные базы данных (словари, справочники и т.д.).

2. Обеспечение «прозрачного» поиска и доступа пользователей к документам, как для ознакомления, так и для анализа содержащихся в них фактов.

3. Организация сбора информации по удаленным цифровым репозиториям и из реферативных баз данных.

Кроме того, при анализе потребностей участников научно-образовательного кластера было выяснено о необходимости разработки единой системы для проведения конференций, обладающей гибкими возможностями, включающими в себя сбор и обработку материалов конференции, регистрацию документов, подтверждающих оплату организационного взноса, рассылку материалов рецензентам и регистрацию их заключений, бронирование гостиниц и др.

1.3 Анализ возможностей участников научно-образовательного кластера

В соответствии с полученной от руководств научно-исследовательских институтов информацией, подготовлен реестр информационных ресурсов участников, линий связи и вычислительных мощностей, на основании которого сделано следующее заключение:

- четыре из шести научно-исследовательских институтов и КазИТУ имеют веб-сайты в сети Интернет, содержащие общую информацию об организации, новости о последних достижениях и др.;

- отсутствует единое хранилище публикаций сотрудников научно-исследовательских институтов и других типов информационных ресурсов; большая часть результатов научных

исследований не оцифрована;

- инфраструктура ИКТ большинства участников научно-образовательного кластера изношена;

- сети внутренних коммуникаций и шлюз выхода в Интернет не защищены в соответствии с требованиями к сетям государственных органов;

- в некоторых из названных НИИ серверная инфраструктура отсутствует, что делает невозможным покупку и эксплуатацию серверного оборудования, а там где она есть, ее характеристики не соответствуют отраслевым стандартам;

- компьютерный парк, несмотря на относительную работоспособность и возможность выполнения простейших задач, требует модернизации или замены;

- во всех научно-исследовательских институтах используется разнородное программное обеспечение и системы управления базами данных, что затрудняет их унификацию, необходимую для оптимизации затрат;

- современное программное обеспечение не используется в работе научно-исследовательских институтов по причине неподготовленности собственного, либо привлекаемого ИТ-персонала. Привлечение специалистов более высокой квалификации на данный момент представляется затруднительным в силу ограниченности бюджета. По этой же причине затруднена покупка и эксплуатация более современных аппаратных ИКТ комплексов.

Сделанные выводы учтены при создании архитектуры распределенной информационной системы.

Одним из участников научно-образовательного кластера является ТОО "Академсеть", успешно работающее на Казахстанском рынке коммуникаций более 15 лет. Это один из немногих Интернет-провайдеров г. Алматы, у которого есть собственный дата-центр. Дата-центр, оборудован по стандарту TIER-2, имеет шесть гермозон площадью 200 м², мощность энергоузла дата-центра составляет 630 кВА, 450 кВт. Дата-центр оснащен:

- 4 источниками бесперебойного и резервного энергоснабжения, каждый по 160 кВА/128 кВт, схема резервирования 2N + 1;

- системой гарантированного энергоснабжения: дизель-генератор 825 кВА/660 кВт, подогрев топливного бака, подогрев топливной системы, подогрев антифриза;

- системой охлаждения;

- системой контроля доступа;
- системой противопожарной и информационной безопасности.

ТОО "Академсеть" предоставляет научно-исследовательским институтам и КазИТУ современные компьютерные ресурсы через Центр обработки данных с помощью радиотехнологий WiMax и WLL. Использование радиотехнологии WLL позволяет оперативно создать планируемый сегмент Облака научно-образовательного кластера, который по стоимости намного дешевле по сравнению с большим объемом волоконно-оптической линий связи (ВОЛС) между научно-исследовательскими институтами. Однако в будущем при запланированном развитии описываемого научно-образовательного кластера возможен безболезненный перевод всей сети на ВОЛС технологии с подключением к этому кластеру других образовательных и научных учреждений г. Алматы и Республики Казахстан.

В 2018 году программное обеспечение, положенное в основу подсистем информационной системы, было установлено на сервере, расположенном в дата-центре ahost компании «Академсеть». Дата-центр оборудован по стандарту TIER-2, имеет шесть гермозон площадью 200 м², мощность энергоузла дата-центра составляет 630 кВА, 450 кВт. Дата-центр оснащен:

- 4 источниками бесперебойного и резервного энергоснабжения, каждый по 160 кВА/128 кВт, схема резервирования 2N + 1;
- системой гарантированного энергопитания: дизель-генератор 825 кВА/660 кВт, подогрев топливного бака, подогрев топливной системы, подогрев антифриза;
- системой охлаждения;
- системой контроля доступа;
- системой противопожарной и информационной безопасности.

В рамках выполнения задач научно-исследовательской работы, в 2019 году закуплен сервер со следующими техническими характеристиками, обеспечивающий хранение постоянно пополняющейся базы данных и бесперебойную работу фронт-интерфейса, публичного доступа к базе данных: Сервер 2x - процессорный стоечный Supermicro CPU 2xE5-2650V4 2.2 GHz / Fan 2x1U / DDR4 4x8GB 2666MHz / MB 8xDDR4 - KVM - 10xSATA - RAID SW 0,1,5,10 / Контроллер RAID 8-int / Корпус стоечный 1U 4xHS 2PSU / SSD 2x480GB / Опции для HDD/SSD 2xMCP / Net 2xSFP 10Gb.

Данный сервер также размещен в дата-центре ahost, где на безвозмездной основе в защищенных условиях обеспечен доступ к сети Интернет и технологическим микроклиматом.

Характеристики сервера обеспечивают его актуальность на достаточно длительную перспективу даже с учетом возможной смены поколений программного обеспечения. Кроме того, структура портала позволяет масштабировать и совершенствовать аппаратные средства что даст возможность при необходимости существенно расширить вычислительные мощности портала.

Фронт-интерфейс ограниченного публичного доступа к базе данных построен на базе портала (сайта) с доменным именем asagor.kz.

Учитывая значимость оцифрованной информации и её возможный уровень секретности, доступ к накопленным данным строго регламентируется владельцами информации, то есть администрацией НИИ. В то же время, внутренняя, служебная часть портала обеспечивает связь рабочих мест в НИИ с сервером посредством VPN канала.

На основании проведенного обследования ИКТ-инфраструктуры научно-исследовательских институтов был сделан вывод об отсутствии возможностей реализации задач с использованием только их мощностей. Даже при вливании значительных средств на полную модернизацию ИКТ-инфраструктуры названных научно-исследовательских институтов в соответствии с современными требованиями, потребуются значительные средства на ее содержание и развитие. В связи чем, при выборе путей реализации задач было принято решение о централизации ИКТ инфраструктуры предполагаемого научно образовательно кластера с ориентацией на создание в дальнейшем единой облачной IT-платформы дислоцированных в центре Алматы НИИ и ВУЗов МОН РК, объединяемых физической сетью на базе ВОЛС.

Облако в полной мере обеспечивает необходимыми вычислительными ресурсами проводимые ими научные исследования и образовательные процессы, упрощая перспективу дальнейшего ее развития.

Ключевые этапы – 1) построение единой сети, 2) централизация серверов 3) виртуализация общей платформы - позволяют построить передовую масштабируемую, отказоустойчивую IT-инфраструктуру с

К программному обеспечению, положенному в основу подсистемы "Репозиторий цифровых объектов", были выдвинуты следующие основные требования:

1. Возможность работы с произвольными документами, географическими картами, аудио- и видеоматериалами;

2. Гибкая организация хранения ресурсов (разбиение на коллекции, подколлекции и пр.);

3. Гибкая система прав пользователей: создание групп пользователей; возможность задания доступа пользователей заданной группы к заданному множеству объектов по требуемому способу доступа (загрузка, просмотр, редактирование, удаление, изменение атрибутов); идентификация, аутентификация и авторизация пользователей;

4. Возможность интеграции распределенных информационных ресурсов на основе стандартных протоколов (Z39.50/SRU/SRW);

5. Наличие программного интерфейса (API) для осуществления интеграции с внутренними ресурсами;

6. Распознавание текста в оцифрованных материалах для организации полнотекстового поиска;

7. Сбор статистики и предоставление различной отчетности.

При выборе подсистемы репозитория цифровых объектов было изучено следующие институциональные репозитории и системы для создания электронных библиотек: Ambra, Digital Commons, DSpace, ePrints, Evergreen ILS, Greenstone, Fedora Commons, Invenio, RODA и VuFind, а также изучен опыт их использования организациями Канады, Новой Зеландии, Чехии и Шри-Ланки [2-7].

Анализируя преимущества и недостатки перечисленных систем, были отобраны системы DSpace, ePrints и Greenstone, как наиболее удовлетворяющие установленным требованиям.

К сильным сторонам Greenstone можно отнести иерархическую структуризацию каждого документа, автоматическое извлечение метаданных из документа при его загрузке. Однако, данная система поддерживает лишь ограниченное количество форматов: MS Word, Excel, RTF, HTML, Plain, PDF, ZIP, MP3. Хранение географических карт, а также других результатов научной деятельности, имеющих более сложную структуру не предусмотрено. Система предоставляет широкие возможности поисковых запросов: помимо булевых операций и группировки слов с помощью скобок, поддерживается поиск слов в исходной форме.

К сильным сторонам DSpace можно отнести более совершенную систему прав пользователя по сравнению с рассмотренными системами: различные научно-исследовательские институты могут иметь свои области в пределах системы. В каждом институте могут быть назначены определенные сотрудники, ответственные за премодерацию, т.е. пользователи, имеющие возможность просматривать и редактировать материалы до того, как они будут включены в репозиторий. DSpace, как и другие рассмотренные системы, предоставляет интерфейсы интеграции с другими подсистемами на основе открытых международных стандартов. DSpace поддерживает более 70 форматов информационных ресурсов. Материалы, имеющиеся в DSpace индексируются в Google Scholar. Существует большое количество плагинов к системе DSpace, расширяющие его возможности, в том числе и система управления научными исследованиями DSpace-CRIS.

Система ePrints поддерживает больше форматов метаданных, однако не поддерживает расширенное дублинское ядро. В системе можно назначить следующие роли пользователей: пользователи, редакторы, администраторы, причем редакторы имеют доступ к материалам всех научно-исследовательских институтов, хранящихся в репозитории.

В результате проведенного анализа был выбран репозиторий цифровых объектов DSpace версии 6.2. Для корректной работы данного программного обеспечения, в системе дополнительно установлены следующие компоненты:

1. Комплект разработчика приложений на языке Java – Open JDK;
2. Apache Maven, Apache Ant для компиляции DSpace из исходных кодов;
3. Свободный веб-сервер Apache Tomcat для взаимодействия системы с пользователем посредством Web-технологий;
4. Открытая платформа полнотекстового поиска Apache SOLR.
5. Свободная СУБД PostgreSQL для хранения метаданных о цифровых объектах и другой информации.

Репозиторий доступен по адресу <https://acagor.kz/resource/repository>, главная страница которого показана на рисунке 4. При сборке программного обеспечения DSpace 6.2, внесены изменения в ее конфигурацию с целью адаптации к условиям, установленным в Республике Казахстан.

Стандартная схема метаданных DSpace, основанная на схеме DCMI, расширена следующими полями: «Журнал в списке ВАК», «Номер названия, тома, выпуска журнала», «Полная библиографическая ссылка по ГОСТ», «Первая и последняя страница статьи» и др.



Рисунок 4. Главная страница репозитория цифровых объектов
<https://acagor.kz/resource/repository>

Список типов ресурсов дополнен позициями «Статья в материалах конференций», «Статья в газете», «Отчет по НИР», «Патент», «Технический отчет», «Музейный объект», «Рабочие материалы». Кроме того, добавлена возможность работы с географической информацией согласно работе [8].

Для поддержки процесса наполнения полнотекстовых баз данных, созданные профили метаданных зарегистрированы в системе DSpace, и в соответствии с ними настроены рабочие процессы, а также пользовательский интерфейс системы.

Внутренняя организация хранения ресурсов в системе DSpace построена следующим образом: в репозитории созданы 7 сообществ, соответствующие научно-исследовательским институтам. Каждое сообщество, в свою очередь, состоит из нескольких коллекций, соответствующих типу ресурса (статьи, монографии, отчеты НИР и

пр.)

На рисунке 5 приведен пример страницы репозитория с информацией о публикации.



Рисунок 5. Пример страницы репозитория с информацией о публикации

Ввод данных в подсистему осуществляется:

- в интерактивном режиме через встроенные WEB-интерфейсы;
- заимствованием данных из других систем (DOI, PubMed, ArXiv, CiNii, CrossRef и др.);
- в пакетном режиме - импорт данных в форматах DIM, MEKOF, MARC21, DC и др.;
- синхронизацией данных по OAI-PMH с внешними источниками.

Доступ к данным репозитория возможен не только через WEB-интерфейсы DSpace, но и по протоколам OAI-PMH, SOLR, SRW/SRU, Z39.50. Последнее обеспечивается связью DSpace с системой ZooSPACE [9].

2.3.2 Подсистема управления текущей научно-исследовательской информацией

Подсистема управления текущей научно-исследовательской информацией предназначена для хранения информации о НИИ, их сотрудниках, а также информации об их научной деятельности (участие в финансируемых проектах, конференциях, прохождение стажировок и др.).

В качестве системы управления научными исследованиями [10, 11] было выбрано свободное расширение системы DSpace – DSpace-CRIS. Система позволяет хранить следующую информацию:

- информацию о научно-исследовательских организациях;
- информацию о сотруднике научно-исследовательских организаций, различные написания его имени, в том числе на разных языках;
- ссылки на профили в различных базах данных (Scopus, Researcher ID, ORCID);
- информацию о научной деятельности (участие в финансируемых проектах, конференциях, прохождение стажировок и др.).

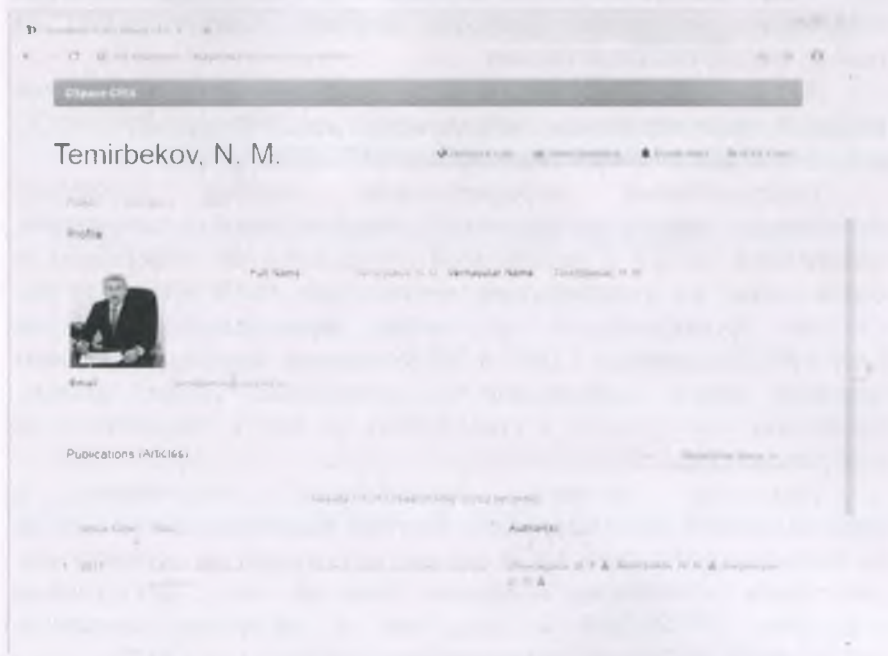


Рисунок 6. Пример профайла ученого в системе CRIS

Система ведет статистику публикаций по каждому ученому. Система интегрирована с репозиторием цифровых объектов, что позволяет просматривать публикации ученых. CRIS-система позволяет экспортировать информацию о публикациях ученого в популярные форматы. Система допускает отслеживание изменений на странице с помощью технологий RSS.

Система доступна по адресу <https://acagor.kz/resource/cris>. Пример профайла ученого в системе DSpace-CRIS приведен на рисунке 6.

2.3.3 Подсистема интеграции распределенных информационных ресурсов

Подсистема интеграции распределенных информационных ресурсов предназначена для импорта метаданных из внешних источников (реферативных баз данных), а также для предоставления метаданных внутренним ресурсам (порталу Академгородка) на основе стандартных протоколов.

В качестве интегрирующего программного обеспечения выбрана распределенная информационная система ZooSPACE, разработанная научными сотрудниками ИБТ СО РАН [12-14].

Распределенная информационная система ZooSPACE интегрирует данные из различных информационных источников, обеспечивая доступ к разнородной распределенной информации в соответствии со стандартными протоколами (SRW/SRU, Z39.50). Система функционирует на основе оригинальных серверов ZooPARK-ZS, серверов LDAP и WEB-серверов Apache, обеспечивая сквозной поиск информации в разнородных базах данных, извлечение информации в стандартных схемах и форматах и ее отображение [9].

Адаптация данного программного обеспечения к распределенной информационной системе Академгородка находится на завершающей стадии, после отладки которого будет осуществлена интеграция репозитория цифровых объектов как с внутренними ресурсами (ИРБИС-64 и др.), так и ресурсами сторонних организаций, доступных по протоколам SRW/SRU или Z39.50.

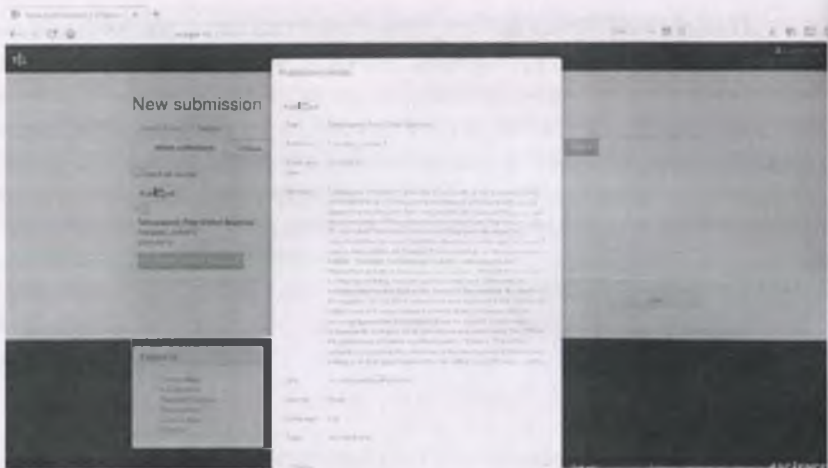


Рисунок 8. Результат поиска во внешней базе данных

2.3.4 Подсистема доступа к распределенным информационным ресурсам на основе Web-технологий

Подсистема доступа к распределенным информационным ресурсам на основе Web-технологий предназначена для предоставления стандартизированного единого пользовательского интерфейса для всех функций и модулей, входящих в состав распределенной информационной системы. Портальная часть должна обеспечивать удаленный доступ к информационным ресурсам и услугам. Подсистема объединяет процессы предоставления доступа к информационным ресурсам всем заинтересованным субъектам взаимодействия.

Разработано техническое задание для разработки программной части распределенной библиотечной информационной системы, в котором определены требования к отдельным компонентам системы, к ее архитектуре, описаны роли пользователей, требования к эргономике и технической эстетике, требования к поддержанию логической целостности данных и обеспечению обратимости изменений, требования к надежности и безопасности, требования к режимам функционирования, требования к защите информации от несанкционированного доступа, требования к эксплуатации,

техническому обслуживанию и хранению компонентов системы и требования к патентной чистоте.

Данная подсистема распределенной информационной системы предназначена для предоставления стандартизированного единого пользовательского интерфейса для всех функций и модулей, входящих в состав распределенной информационной системы. Портальная часть должна обеспечивать удаленный доступ к информационным ресурсам и услугам. Подсистема объединяет процессы предоставления доступа к информационным ресурсам всем заинтересованным субъектам взаимодействия. Представление услуг предполагает:

- навигацию по сервисам;
- доступ к нормативным правовым и методическим материалам;
- идентификацию, аутентификацию и авторизацию пользователей;
- предоставление основной информации о научно-исследовательских институтах, их сотрудниках;
- предоставление информации о последних достижениях научно-исследовательских институтов, о предстоящих событиях, конференциях;
- наличие фото- и видеогалереи;
- личный кабинет, анализ времени работы на сайте, хранение истории запросов;
- управление запросами по поиску авторов, названию публикаций;
- управление запросами по полнотекстовому поиску по ресурсам;
- экспорт данных в различные форматы;
- предоставление различных отчетностей.

К порталу выдвинуты следующие основные требования:

– Серверная часть портала должна использовать шаблон проектирования Model-View-Controller (MVC) – схему разделения данных приложения, пользовательского интерфейса и управляющей логики на три отдельных компонента: модель, представление и контроллер таким образом, что модификация каждого компонента может осуществляться независимо.

– Клиентская часть портала должна использовать адаптивную верстку для удобства просмотра сайта как на настольных компьютерах, так и на мобильных устройствах.

Портал Академгородка, главная страница которого приведена на Рисунке 9, доступен по адресу <https://acagor.kz>.



Рисунок 9. Главная страница портала Академгородка

Сайт разработан с использованием веб-фреймворка Django 2.0.5 (на языке Python 3.6). Выбор данного веб-фреймворка определен возможностями быстрого прототипирования, автоматического создания миграций для деплоймента. Схема взаимодействия портала Академгородка с остальными составляющими распределенной информационной системы представлена на Рисунке 10.

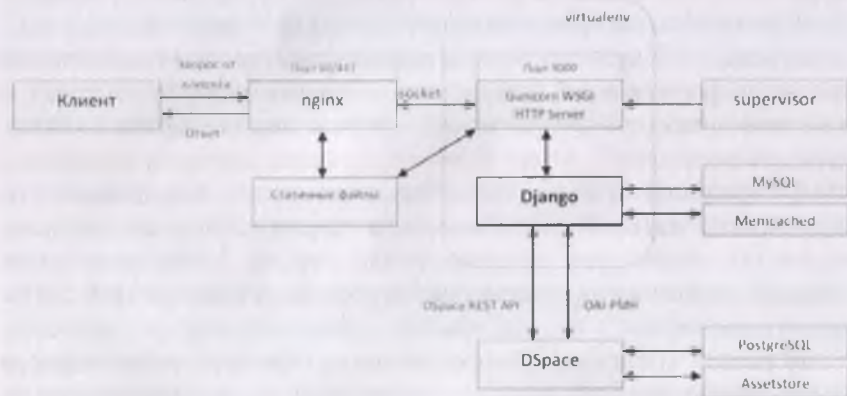


Рисунок 10. Схема взаимодействия портала Академгородка с составляющими распределенной информационной системы

Сайт функционирует на WSGI HTTP-сервере Gunicorn 19.8.1 с HTTP-сервером nginx 1.14, установленным в качестве обратного прокси-сервера.

Сайт использует программное обеспечение Memcached 1.5.6 для реализации временного хранения (кэширования) данных в оперативной памяти с целью разгрузки базы данных. Для хранения данных портала используется система управления базами данных MySQL 5.7.23.

Django установлен в виртуальном окружении Python с целью избежания возможных конфликтов версий пакетов, возможности обновления необходимого вспомогательного программного обеспечения независимо от пакетов, установленных в операционной системе. Кроме бинарных пакетов python, pip и wheel, в виртуальном

окружении установлен веб-сервер Gunicorn. В Django установлены следующие дополнительные пакеты: пакет mysqlclient для работы с базой данных, пакет memcached для работы с сервисом кэширования, пакет PIL для работы с изображениями.

Для запуска и мониторинга Gunicorn используется система контроля процессами supervisor, которая запускает сокет UNIX, прослушивает его и автоматически запускает Gunicorn в ответ на запрос.

Установлен SSL-сертификат от центра сертификации Let's Encrypt.

Портал Академгородка включает следующие разделы:

а) раздел «О проекте» предоставляет информацию о назначении системы, информацию об используемом программном обеспечении, а также информацию о нормативных документах, на основе которых она разработана;

б) раздел «Институты» предоставляет информацию о направлениях научной деятельности, истории, основных научных результатах институтов Академгородка города Алматы, а также содержит список сотрудников институтов и ссылки на веб-сайты институтов;

в) раздел «Персоналии» предоставляет краткую информацию о каждом сотруднике института, его научный вклад, а также ссылки на труды, загруженные в репозиторий; информация формируется на основе данных системы управления научной информацией;

г) раздел «Новости» содержит информацию о последних достижениях в жизни институтов, проведенных мероприятиях, а также информацию о предстоящих событиях;

д) раздел «Ресурсы» содержит ссылки на внутренние и сторонние ресурсы, в том числе на цифровой репозиторий объектов DSpace и систему управления научной информацией DSpace-CRIS, а также форму поиска ресурсов данных систем по метаданным;

е) раздел «Галерея» содержит фото- и видеоматериалы научно-исследовательских институтов и проведенных мероприятий.

Материалы на портале доступны на трех языках.

2.3.5 Централизованное хранилище учетных записей пользователей

В целях централизованной аутентификации пользователей использована свободная реализация протокола LDAP – OpenLDAP. Данным способом осуществляется аутентификация пользователей репозитория цифровых объектов DSpace, портала Академгородка (проект Django) и почтового сервера. LDAP-сервер доступен по адресу `ldap://acagor.kz:389`.

Модифицирована стандартная схема аутентификации проекта Django. По умолчанию, данная схема предполагает аутентификацию пользователя с использованием логина и пароля. В модифицированной схеме аутентификация производится по адресу электронной почты и паролю. Для реализации данной модификации переопределен класс `UserManager` (потомок класса `BaseUserManager`) и класс `User` (потомок класса `AbstractUser`). Кроме того, создан класс `Profile`, определяющий разрешения конкретному пользователю на изменение разделов административной части. Схема данных модуля аутентификации представлена на Рисунке 11. В рамках данной схемы реализована возможность регистрации пользователей на портале Академгородка.

LDAP-аутентификация в DSpace является одним из стандартных способов аутентификации, добавление и настройка которой осуществляется в конфигурационном файле.

Запланирован и подготовлен переход на технологию единого входа (SSO), при которой пользователь переходит из одного раздела портала в другой без повторной аутентификации. Для этих целей модель приложения `UserAuth` проекта Django, отвечающей за авторизацию, изменена, как показано на Рисунке 12. В этом случае DSpace будет переведен на Shibboleth-аутентификацию, которая также является стандартной.

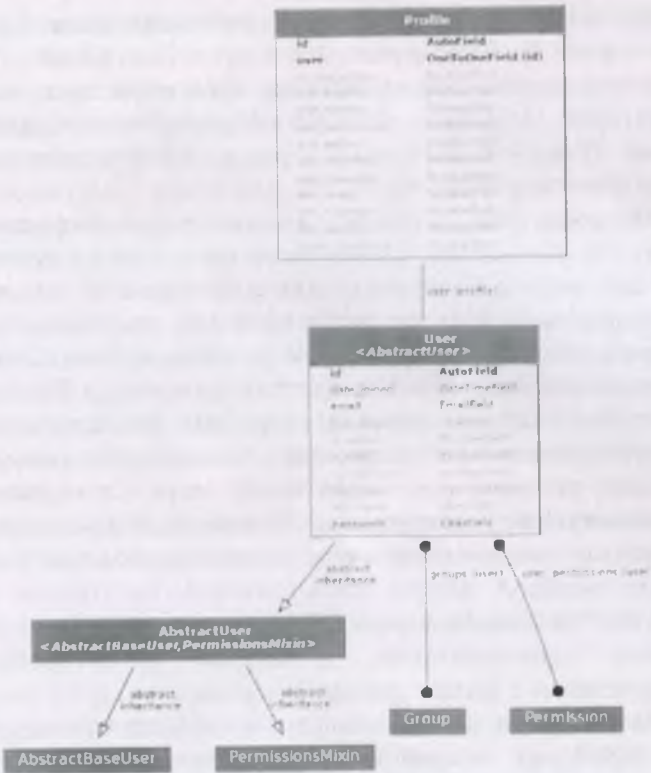


Рисунок 11. Схема данных модуля авторизации

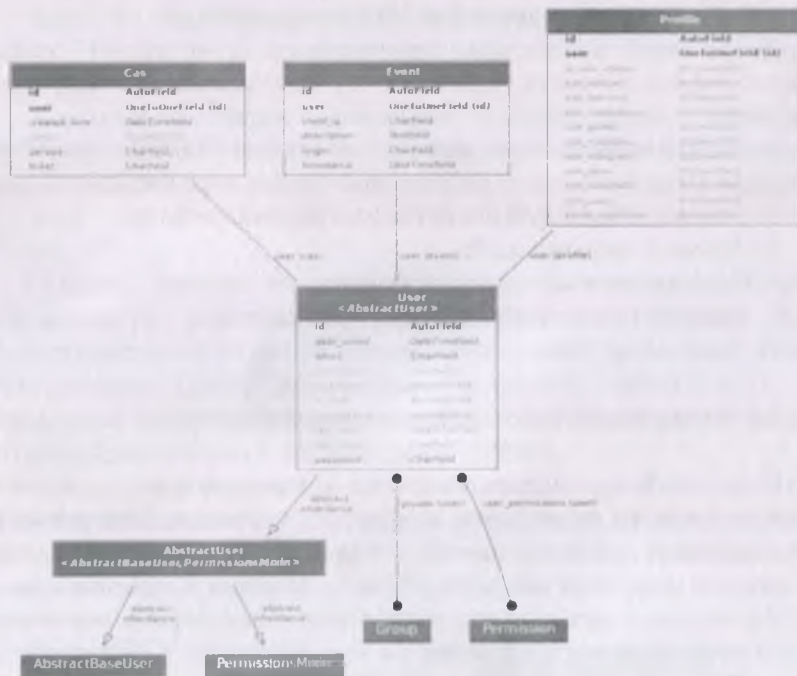


Рисунок 12. Схема данных модуля авторизации

2.3.6 Механизм автоматического деплоя портала

Разработан механизм автоматического деплоя портала Академгородка. При разработке используется распределенная система управления версиями GIT.

На сервере развернуты две копии сайта: версия для разработчиков (далее – dev-версия) и рабочая версия (далее – prod-версия).

На рабочих станциях разработчиков создан скрипт на языке Bash, который осуществляет подготовку деплоя в следующем порядке:

1. Создание миграций.

2. Синхронизация исходных файлов проекта на рабочих станциях разработчиков с файлами dev-версии портала на сервере

(кроме файла конфигурации и файлов в папке media).

Деплоймент на сервере производится в несколько этапов, которые осуществляются скриптом на языке Bash в следующем порядке:

1. Создание дампа базы данных с помощью утилиты mysqldump и создание резервной копии файлов prod-версии портала;
2. Синхронизация файлов dev и prod версий портала;
3. Применение миграций;
4. Перезагрузка веб-сервера Gunicorn;
5. Принудительный сброс кэша в Memcached.

2.4 Интеграция DSpace и портала Академгородка

В целях предоставления пользователю единого пользовательского интерфейса ко всем функциям интегрированной распределенной информационной системы, осуществлена интеграция репозитория цифровых объектов DSpace и Портала Академгородка.

Рассмотрено два подхода интеграции. Программно реализован первый подход, который основан на использовании API-интерфейса REST DSpace, обеспечивающий программный интерфейс для сообществ, коллекций, метаданных элементов и файлов. Для этого развернуто веб-приложение REST. При каждом запросе пользователя портала Академгородка осуществляется запрос в DSpace REST с помощью утилиты curl. В DSpace создан непривилегированный пользователь, от имени которого производится поиск по репозиторию. Ответы от DSpace REST возвращаются в формате JSON, парсинг которых производится в разработанном программном модуле Integrator проекта Django.

Первое обращение в DSpace REST осуществляется следующим запросом:

```
curl -v -X POST --data «email=acagormanager@yandex.kz &password=****» https://acagor.kz:8080/rest/login.
```

Ответ возвращается в следующем виде:

```
HTTP/1.1 200 OK
```

```
Set-Cookie: JSESSIONID=***; Path=/rest/; Secure; HttpOnly
```

Следующая последовательность запросов возвращает список коллекций (научно-исследовательских институтов):

```
curl -v «http://acagor.kz:8080/rest/communities» --cookie «***»
```

curl -v «http://acagor.kz:8080/rest/communities/ i/collections» – cookie «***», где i – идентификатор (handle) сообщества, полученного в предыдущем запросе.

Следующий запрос возвращает список трудов Казахского научно-исследовательского института перерабатывающей и пищевой промышленности:

curl -v «http://acagor.kz:8080/rest/collections/7/items» –cookie «***».

Анализ ответов показал, что в зависимости от текущей загруженности репозитория, каждый запрос в DSpace REST осуществляется от 10 до 30 секунд. Это создает временный отказ в обслуживании (DoS) репозитория цифровых объектов, а при выполнении запросов из нескольких рабочих станций может вызвать распределенный отказ в обслуживании (DDoS).

Программно реализован второй подход, который заключается в выгрузке метаданных и кэшировании результатов во временной базе данных в течение некоторого времени. Создана временная база данных со схемой, представленной на Рисунке 13.

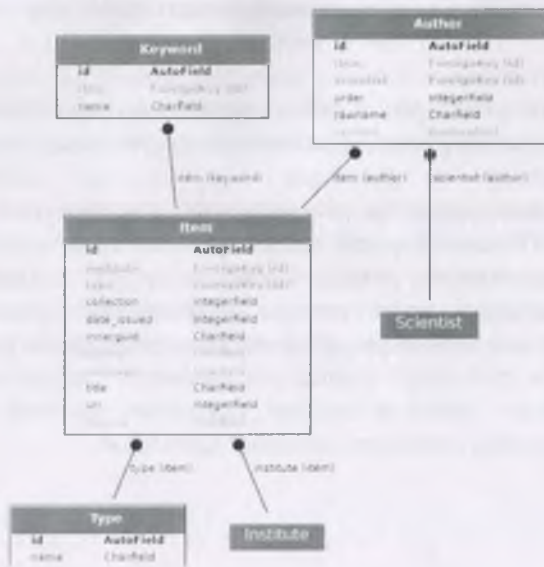


Рисунок 13. Схема базы данных

Данный подход позволил существенно сократить время отклика. При использовании закешированных данных время отклика составило в пределах 0,01-0,08 секунд (Рисунок 14).

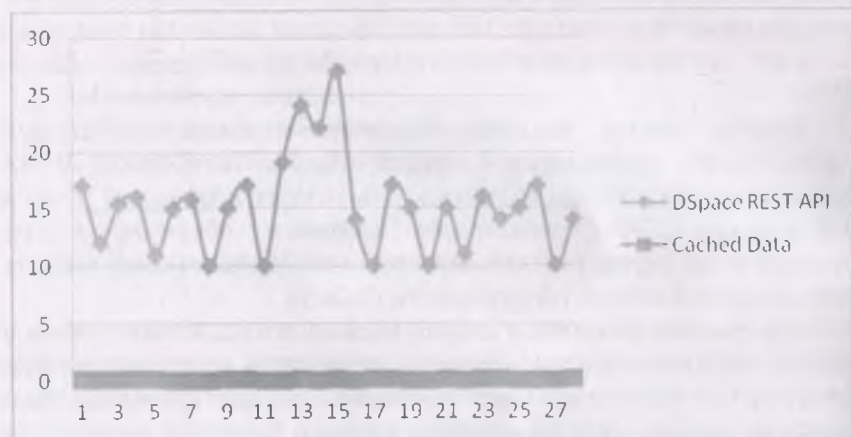


Рисунок 14. Среднее время отклика при использовании DSpace REST API и закешированных данных

Результат интеграции репозитория цифровых объектов DSpace и портала Академгородка прослеживается на следующих страницах веб-портала.

а) Поиск ресурсов по названию ресурса или его авторов (Ресурсы > Поиск). В левой части страницы отображается статистика найденных ресурсов в разрезе научно-исследовательских институтов, даты публикации, типа ресурса и ключевых слов. Реализована фильтрация найденных ресурсов по указанным фильтрам. На Рисунке 15 приведен результат поиска по названию ресурсов, содержащих слово «почва». Числа в скобках указывают количество найденных ресурсов, соответствующих данному критерию.



Рисунок 15. Блок фильтров в разрезе научно-исследовательских институтов, даты публикации, типа ресурса и ключевых слов

При нажатии на ссылку с именем ученого в результатах поиска открывается его профиль (Рисунок 16, 17), который содержит краткую информацию об его научном вкладе, ссылки на профили в различных базах данных (при наличии), а также перечень публикаций, загруженных в репозиторий цифровых объектов (Рисунок 18). При наличии ссылок на реферативные базы данных (ORCID, Researcher ID), указывается данная информация. Список всех ученых, чьи труды загружены в репозиторий цифровых объектов отображается в разделе «Персоналии» (Рисунок 19).

Темирбеков Нурлан Муханович

Ученый сотрудник, доктор технических наук, профессор

Внешний вид

Основной вклад

Ученый Н.М. Темирбеков внес вклад в развитие науки и техники в области теории автоматического управления и теории систем управления. Его научные исследования в области теории систем управления и теории систем управления в целом для системы управления объектами с запаздыванием, в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием. Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием.

Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием. Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием.

Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием. Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием.

Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием. Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием.

Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием. Ученый внес вклад в развитие науки и техники в области теории систем управления объектами с запаздыванием и в области теории систем управления объектами с запаздыванием.



Рисунок 16. Профиль ученого

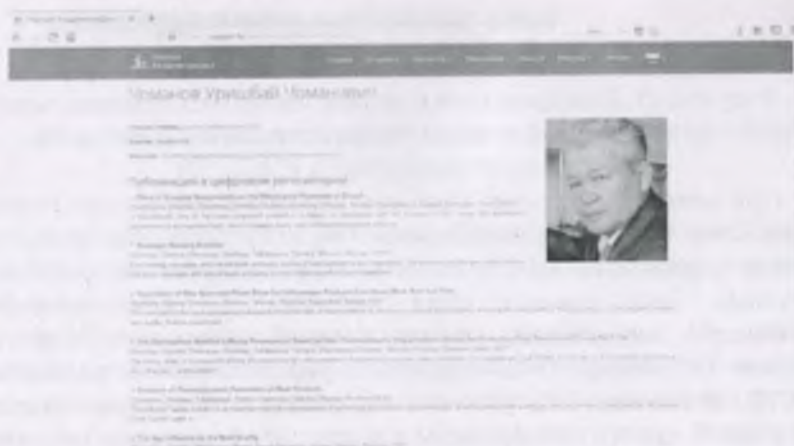


Рисунок 17. Результаты запроса, сделанного в репозиторий цифровых объектов

считаются разными учеными. При поиске ресурсов по автору учитывается различное написание его имени, в том числе на разных языках. Пример результатов поиска по автору приведен на рисунках 20, 21. Из приведенного примера видно, что указанные ресурсы попали в выборку, хотя имя автора имеет разное написание.

б) Отображение списка научно-исследовательских институтов, описание научных направлений и достижений, списка их сотрудников, а также метаданных публикаций, имеющихся в репозитории цифровых объектов (Рисунок 22).

Numerical solution of the problem of optimizing the process of oil displacement by water

Мухоморова, Г.М., Вязовкина, Т.В.

AIP Publishing (2016)

Probabilistic Statistical Modeling of Air Pollution from Vehicles

Алиханова, И., Мухоморова, В.А., Мадыева, М.Н., Тиммаева, Г.А.

AIP Conference Proceedings 1880 (2017)

Study of Stability of the Difference Scheme for the Mixed Problem of the Ginzburg-Itskii

Тиммаева, Г.А., Тиммаев, А.

International Conference "Functional Analysis in Interdisciplinary Applications" (2017)

Распределение электро-образовательной аппаратуры на общенациональном уровне

Тиммаева, Г.А., Байкенов, Д.Р., Тиммаев, А.И., Мухоморова, Г.А., Сидорова, Е.С.

(2018)

Рисунок 20. Результат поиска по автору

наполнением системами дистанционного обучения, электронного документооборота, ERP, управления интеллектуальным капиталом, электронной библиотекой и другими. Наличие которых, в свою очередь, порождает экономический стимул для коммерческого подключения в создаваемый научно-образовательный кластер новых участников.

Экономический эффект от такого подхода определяется возможностями облачной инфраструктуры позволяющей пользователям избавиться от непрофильных IT-затрат, получая в то же время современные ресурсы «по требованию». При этом, в случае необходимости IT-инфраструктура клиентов может быть оперативно модернизирована до уровня суперкомпьютера, ускоряя расчеты и эксперименты, необходимые в научных исследованиях любого масштаба. Кроме того, централизованный серверный узел может обслуживаться существенно меньшим количеством сотрудников, позволяя за счет экономии общего ФЭП привлекать специалистов высшей по отрасли и по ведомству квалификации. Открывая тем самым возможность внедрения в перспективе современных IT систем, недоступных доселе МОН по причине кадровой необеспеченности и содействуя сокращению сложившегося отставания Казахстанской науки от мировых тенденций.

При существенном ограничении бюджета на создание ВОЛС сети, объединение ресурсов участников НОК возможно посредством организации радиоканалов связи, поскольку практически все НИИ, как показало обследование, по удачному стечению обстоятельств находятся в пределах прямой видимости. В пределах прямой видимости производительность таких каналов достигает 100 мбит/сек и позволяет использовать клиент-серверные технологии с использованием удаленных серверов. Виртуализация серверной инфраструктуры участников НОК позволит создавать ресурсы без привязки к технической спецификации используемых серверов, опираясь лишь на необходимое количество ядер и дискового пространства. Кроме того, учитывая, что на первоначальном этапе не все серверные ресурсы будут задействованы в полной мере, нет необходимости выделения большого пула серверных мощностей.

считаются разными учеными. При поиске ресурсов по автору учитывается различное написание его имени, в том числе на разных языках. Пример результатов поиска по автору приведен на рисунках 20, 21. Из приведенного примера видно, что указанные ресурсы попали в выборку, хотя имя автора имеет разное написание.

б) Отображение списка научно-исследовательских институтов, описание научных направлений и достижений, списка их сотрудников, а также метаданных публикаций, имеющих в репозитории цифровых объектов (Рисунок 22).

Normalized solution of the problem of sampling the process of oil displacement by water

Tokomarov, E. M., Shapovalov, N. B.
AIP Publishing (2018)

Probabilistic Statistical Modeling of Air Pollution from Vehicles

Adzhimov, S., Muzumdar, R. A., Madhavi, M. H., Tokomarov, E. M.
AIP Conference Proceedings 1880 (2017)

Stability of Stability of the Difference Scheme for the Model Problem of the Control Process

Tokomarov, E. M., Tolmachev, A. A.
International Conference "Functional Analysis in Interdisciplinary Applications" (2017)

Безразмерные функции Брэдшоу в теории нелинейных колебаний в электрических цепях с нелинейными элементами

Токмаров, Е.М., Вайсман, Д.Р., Геллерштейн, А.Н., Уайсман, Е.М., Токмарова, Е.С.
(2018)

Рисунок 20. Результат поиска по автору

наполнением системами дистанционного обучения, электронного документооборота, ERP, управления интеллектуальным капиталом, электронной библиотекой и другими. Наличие которых, в свою очередь, порождает экономический стимул для коммерческого подключения в создаваемый научно-образовательный кластер новых участников.

Экономический эффект от такого подхода определяется возможностями облачной инфраструктуры позволяющей пользователям избавиться от непрофильных ИТ-затрат, получая в то же время современные ресурсы «по требованию». При этом, в случае необходимости ИТ-инфраструктура клиентов может быть оперативно модернизирована до уровня суперкомпьютера, ускоряя расчеты и эксперименты, необходимые в научных исследованиях любого масштаба. Кроме того, централизованный серверный узел может обслуживаться существенно меньшим количеством сотрудников, позволяя за счет экономии общего ФЗП привлекать специалистов высшей по отрасли и по ведомству квалификации. Открывая тем самым возможность внедрения в перспективе современных ИТ систем, недоступных доселе МОН по причине кадровой необеспеченности и содействуя сокращению сложившегося отставания Казахстанской науки от мировых тенденций.

При существенном ограничении бюджета на создание ВОЛС сети, объединение ресурсов участников НОК возможно посредством организации радиоканалов связи, поскольку практически все НИИ, как показало обследование, по удачному стечению обстоятельств находятся в пределах прямой видимости. В пределах прямой видимости производительность таких каналов достигает 100 мбит/сек и позволяет использовать клиент-серверные технологии с использованием удаленных серверов. Виртуализация серверной инфраструктуры участников НОК позволит создавать ресурсы без привязки к технической спецификации используемых серверов, опираясь лишь на необходимое количество ядер и дискового пространства. Кроме того, учитывая, что на первоначальном этапе не все серверные ресурсы будут задействованы в полной мере, нет необходимости выделения большого пула серверных мощностей.

2 РАЗРАБОТКА ИНТЕГРИРОВАННОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ АКАДЕМГОРОДКА ГОРОДА АЛМАТЫ

2.1 Аппаратная часть распределенной информационной системы

Проанализировав факты, приведенные в подразделе 1.3, сделан единственно возможный вывод о целесообразности и необходимости выноса систем хранения данных за пределы научно-исследовательских институтов (участников научно-образовательного кластера) в Центр обработки данных ТОО «Академсеть», оставив рабочие места для первичной обработки информации непосредственно в научно-исследовательских институтах. Такая распределенная централизованная архитектура систем хранения данных позволяет:

- унифицировать инструментарий хранения и обработки данных;

- использовать современное, в достаточной мере производительное аппаратное обеспечение, в том числе серверы и системы хранения данных, обеспечив, при необходимости, их модернизацию и масштабирование;

- привлечь к разработке создаваемых систем IT-персонал достаточно высокой квалификации, разделив в итоге затраты по его содержанию. По суммарным затратам это, возможно, будет не дешевле существующих затрат, но такому персоналу будут доступны задачи более высокого современного уровня;

- обеспечить создаваемую информационную систему необходимыми средствами защиты от атак и взлома. В том числе и в соответствии с требованиями к сетям государственных органов. Такое требование, кажущееся на первый взгляд избыточным к научно-исследовательским организациям, в результате может оказаться необходимым, поскольку информация, размещаемая в информационной системе может попадать под различные уровни секретности.

В дальнейшем эта информация будет передаваться внешним пользователям только при наличии соответствующей авторизации через единый шлюз доступа в Интернет. Шлюз также должен получить развитие как по увеличению функционала защиты от атак и взлома, так и по разграничению прав доступа и защите

подключаемых каналов связи.

Потенциальная возможность присутствия информации с ограниченными категориями доступа требует соответствующей защищенности каналов связи от рабочих мест НИИ до Центра обработки данных (ЦОД). На этапе отладки допустимо использование общих Интернет-каналов, но весь комплекс должен опираться только на защищенные каналы связи НИИ-ЦОД. Более надежными могут считаться закрытые каналы связи типа "точка-точка", которые намного надежнее открытых каналов даже при использовании в последних различных алгоритмов шифрования.

Каналы связи от рабочих мест НИИ до ЦОД должны иметь возможность существенного увеличения пропускной способности. В этом случае у НИИ появится возможность использования других сервисов ЦОД. В частности, технологий виртуализации, сервисов IaaS, DaaS, SaaS, PaaS, CCTVaas, SecaaS и других. Это позволит существенно расширить возможности IT-парка всех НИИ на базе облачных мощностей используемого ЦОД и объединить их в итоге в единый научно-образовательный кластер.

В качестве архитектуры централизованной системы хранения данных при запланированном относительно небольшом объеме данных, выбрана архитектура типа NAS. Этот вывод сделан в силу того, что типовая процедура обработки данных строится по схеме, когда клиентский запрос серверу обрабатывается одним или несколькими серверами обработки данных, которые получают необходимые данные, производят их обработку и возвращает результат этой обработки пользователю. При архитектуре системы хранения данных типа NAS обработка запросов ведется в рамках одной сети. При этом по этой сети идут как сами запросы, так и все необходимые для обработки этого запроса данные. Поэтому для того, чтобы сеть, объединяющая НИИ и ЦОД, обладала необходимой пропускной способностью, она должна обеспечивать необходимую полосу пропускания для больших объемов данных практически по всей сети. Это не всегда возможно и всегда достаточно дорого в реализации. Обработка и предварительная подготовка данных по определенным стандартам будет производиться на рабочих местах в НИИ, без загрузки сети, и только впоследствии она будет передаваться на хранение в библиотеки на централизованную систему хранения данных, размещаемую в ЦОД.

В последующем, при расширении объемов данных от НИИ и

увеличении количества потенциальных участников, возможен переход к топологии системы хранения данных типа SAN, позволяющей разделять используемый инструментарий и опции СУБД в зависимости от НИИ.

Вследствие относительно небольшого объема данных все программное обеспечение было установлено на один виртуальный сервер, расположенный в Центре обработки данных ТОО «Академсеть» и имеющий следующие технические характеристики: Intel Xeon E5 2V Core, 8 GB DDR3 ECC, 500 GB HDD RAID 5.

Экспоненциальный рост хранимой информации привел к пересмотру архитектуры хранилища данных.

Для организации хранения данных было решено использовать распределенные файловые системы. Выбор распределенной файловой системы производится исходя из соответствия следующим критериям: высокая надежность хранения; доступность данных; отказоустойчивость; децентрализация; масштабируемость; низкая удельная стоимость хранения; простота развертывания и эксплуатации.

Рассмотрены системы хранения данных Ceph и GlusterFS. Обе системы обеспечивают высокую производительность и масштабируемость. Однако эти системы архитектурно противоположны.

GlusterFS [11] работает в пространстве пользователя с использованием технологии FUSE, поэтому не требует поддержки со стороны ядра операционной системы и работает поверх существующих файловых систем. Сильная сторона GlusterFS - более простое развертывание. В отличие от Ceph, GlusterFS не требует отдельного сервера для хранения метаданных, он хранится вместе с данными в расширенных атрибутах файла. Из-за отсутствия привязки к централизованному серверу метаданных файловая система обеспечивает практически неограниченную масштабируемость. Однако GlusterFS имеет меньше возможностей и менее гибок по сравнению с Ceph.

Преимущество Ceph [12] перед GlusterFS заключается в отсутствии единых точек отказа и практически нулевых затрат на обслуживание операций восстановления. В Ceph массив данных автоматически перебалансируется при добавлении или удалении новых узлов. Это происходит практически незаметно для клиентов и обеспечивает высокую живучесть системы.

Серп является легко масштабируемым петабайтным хранилищем, которое может объединять тысячи узлов. Встроенные механизмы продублированной репликации данных обеспечивают высокую живучесть системы, при добавлении или удалении новых узлов массив данных автоматически перебалансируется с учётом изменений.

```

node2.acagor.kz|[WARNIN] command: Running command: /usr/bin/ceph conf --filesystems=ceph --name=mon0 --lookup-oid=mon0
node2.acagor.kz|[WARNIN] command: Running command: /usr/bin/ceph conf --cluster=ceph --name=osd. --lookup-oid=fsid
node2.acagor.kz|[WARNIN] get dm uuid: get dm uuid /dev/sda3 uuid path is /sys/dev/block/8:3/dm/uuid
node2.acagor.kz|[WARNIN] set data partition: OSD data device /dev/sda3 is a partition
node2.acagor.kz|[WARNIN] get dm uuid: get dm uuid /dev/sda3 uuid path is /sys/dev/block/8:3/dm/uuid
node2.acagor.kz|[WARNIN] command: Running command: /sbin/blkid -o udev -p /dev/sda3
node2.acagor.kz|[WARNIN] set data partition: incorrect partition UUID: 8f6c3daf-8483-4772-8e79-3d69d8477de4, expected: 41bd7e29-9d25-41ba-a108-5e6dcecf185d, 41bd7e29-9d25-41ba-a108-62c2acef185d, 41bd7e29-8ae8-49e2-b19d-5d88667a1568, 41bd7e29-9d25-41ba-af08-35865c0ff85d
node2.acagor.kz|[WARNIN] populate data path device: Creating xfs fs on /dev/sda3
node2.acagor.kz|[WARNIN] command: Running command: /sbin/mkfs -t xfs -f -i size=2048 -- /dev/sda3
node2.acagor.kz|[WARNIN] meta data /dev/sda3          isize=2048    account=4, aqsize=136948672 blks
node2.acagor.kz|[WARNIN]                               =                sectsz=512   attr=2, prproj32bit=1
node2.acagor.kz|[WARNIN]                               =                crc=1       finobt=1, sparse=0, nmapbt=0, reflink=0
node2.acagor.kz|[WARNIN]                               =                data        bsize=4096  blocks=52376288, unmapct=5
node2.acagor.kz|[WARNIN]                               =                sunit=0     swidth=0 blks
node2.acagor.kz|[WARNIN]                               =                naming      version 2   bsize=4096  ascii-cl=8  ftype=1
node2.acagor.kz|[WARNIN]                               =                log         internal log  blocks=255743, version=2
node2.acagor.kz|[WARNIN]                               =                sectsz=512  sunit=0 blks, lazy-count=1
node2.acagor.kz|[WARNIN]                               =                realtime    none        extsz=4096  blocks=0, rtextents=0
node2.acagor.kz|[WARNIN] mount: Mounting /dev/sda3 on /var/lib/ceph/tmp/ant.W1kGB5 with options noatime,(node64
node2.acagor.kz|[WARNIN] command: Running command: /bin/mount -t xfs -o noatime,(node64 -- /dev/sda3 /
node2.acagor.kz|[WARNIN] populate data path: Preparing osd data dir /var/lib/ceph/tmp/ant.W1kGB5
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5/ceph/fsid
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5/fsid.5251
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5/magic.525
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5/type.5251
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5
node2.acagor.kz|[WARNIN] command: Running command: /bin/chown -R ceph:ceph /var/lib/ceph/tmp/ant.W1kGB5
node2.acagor.kz|[WARNIN] command check call: Running command: /bin/umount -- /var/lib/ceph/tmp/ant.W1kGB5
node2.acagor.kz|[WARNIN] get dm uuid: get dm uuid /dev/sda3 uuid path is /sys/dev/block/8:3/dm/uuid
node2.acagor.kz|[INFO] Running command: sudo systemctl enable ceph.target
node2.acagor.kz|[INFO] checking OSD status.
node2.acagor.kz|[INFO] find the location of an executable
node2.acagor.kz|[INFO] Running command: sudo /usr/bin/ceph --cluster=ceph osd stat --format=json
ceph_deploy.osd|[INFO] Host node2.acagor.kz is now ready for osd use.
  
```

Рисунок 1. Процесс установки хранилища данных Серп

По утверждению разработчиков, текущая версия CephFS нестабильна для производственного использования, но тест, развернутый на основе трех физических узлов и одной виртуальной машины, не выявил проблем. В тестовом хранилище данных использовались четыре тестовые рабочие станции: административный узел (hub.acagor.kz) и три узла с данными (node1.acagor.kz, node2.acagor.kz, node3.acagor.kz).

Тестирование проводилось в течение месяца с аварийным отключением одной из рабочих станций. В этом случае перебалансировка кластера происходит без второго простоя и

прозрачна для клиентов (Рисунок 1).

В результате многочисленных исследований предпочтение отдается сети хранения объектов Ceph. Ceph показал более приемлемые результаты с точки зрения производительности по сравнению с GlusterFS при работе с цифровым репозиторием DSpace.

2.2 Модель функционирования распределенной информационной системы

На рисунке 2 описана модель функционирования распределенной информационной системы, согласно которой:

– программная часть (Web-сервер, хранилище данных и базы данных) расположены в ЦОД ТОО «Академсеть»;

– контент-менеджеры (сотрудники НИИ и КазИТУ) наполняют контент посредством Web-интерфейса на страницах репозитория и системы управления научными исследованиями;



Рисунок 2. Модель функционирования распределенной информационной системы

– внешние пользователи со своих рабочих станций посещают Web-портал Академгородка и просматривают информацию об

институтах, производят поиск по репозиторию и системе управления научными исследованиями на странице портала.

Построенная модель распределенной информационной системы полностью соответствует потребностям участников научно-образовательного кластера как по информационному содержанию, так и по поддержке отраслевой и языковой специфике, поскольку решает основные задачи, предъявляемые к этим системам: обеспечение системы надежного долговременного хранения цифровых (электронных) документов с сохранением всех смысловых и функциональных характеристик исходных документов; обеспечение «прозрачного» поиска и доступа пользователей к документам, как для ознакомления, так и для анализа содержащихся в них фактов; организация сбора информации по удаленным цифровым репозиториям, поддерживающим протоколы OAI-PMH, SRW/SRU, Z39.50. Система в полной мере обеспечивает необходимыми вычислительными ресурсами проводимые научные исследования и образовательные процессы, упрощая перспективу дальнейшего ее развития, и позволяет построить передовую IT-инфраструктуру управления интеллектуальным капиталом, электронной библиотекой, в котором будут храниться все книги и научные труды Казахстанского инженерно-технологического университета и научно-исследовательских институтов Академгородка города Алматы.

2.3 Программная часть распределенной информационной системы

При проектировании и разработке программной части распределенной информационной системы придерживались следующих нормативных документов:

- Закон РК "О персональных данных и их защите" № 94-V (от 21.05.2013 г. с изменениями от 28.12.2017 г.);
- Закон РК "О связи" №567-II (от 05.07.2014 г. с изменениями от 09.01.2018 г.).

Программная часть распределенной информационной системы состоит из следующих подсистем, представленных на рисунке 3:

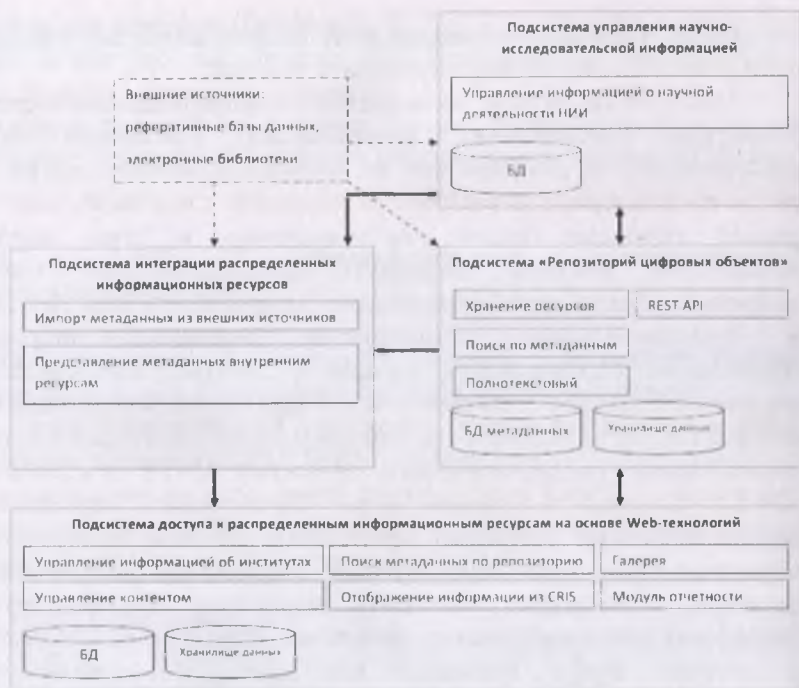


Рисунок 3. Взаимосвязь подсистем распределенной информационной системы

- Репозиторий цифровых объектов;
- Подсистема управления текущей научно-исследовательской информацией;
- Подсистема интеграции распределенных информационных ресурсов;
- Подсистема доступа к распределенным информационным ресурсам на основе Web-технологий (Web-портал).

2.3.1 Подсистема "Репозиторий цифровых объектов"

Подсистема "Репозиторий цифровых объектов" предназначена для долгосрочного хранения результатов научной деятельности НИИ с возможностью поиска информационных ресурсов по метаданным, полнотекстового поиска, ведения статистики.

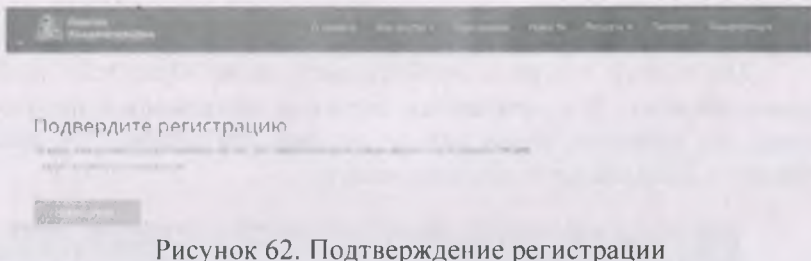


Рисунок 62. Подтверждение регистрации

При успешной регистрации новый пользователь автоматически авторизуется на портале Академгородка. На рисунке 63 показан пример страницы, на которой отображается имя текущего пользователя.



Рисунок 63. Пример страницы портал

После авторизации на портале пользователь получает доступ ко следующим функциям:

1. Поиск ресурсов по репозиторию цифровых объектов;

2.

Для поиска ресурсов необходимо в меню «Ресурсы» выбрать пункт «Поиск». В открывшейся странице предлагается произвести поиск по названию и/или автору ресурса. На рисунке 64 показан пример поиска ресурсов по слову «почва».



Рисунок 64. Поиск ресурсов.

Результаты поиска отображаются на странице, показанной на рисунке 65. В левой части страницы выводится информация о количестве найденных ресурсов (равном 55) и панель фильтров, на правой – список найденных ресурсов.



Рисунок 65. Результаты поиска

Панель фильтров включает в себя:

1. Фильтр по институтам;
2. Фильтр по дате;
3. Фильтр по типам ресурсов;
4. Фильтр по ключевым словам.

Фильтр по институтам содержит список научно-исследовательских институтов, учеными которого создан ресурс, содержащий в названии искомое слово. Числа в скобках указывают количество таких ресурсов. В вышеприведенном примере при поиске слова «почва» найденные 55 ресурсов распределились в разрезе научно-исследовательских институтов следующим образом (Рисунок 66, а):

1. Казахстанский инженерно-технологический университет – 32 ресурса;
2. КазНИИ почвоведения и агрохимии имени У.У.Успанова – 13 ресурсов;
3. НИИ микробиологии и вирусологии – 9 ресурсов;

4. КазНИИ плодоовощеводства – I ресурс.

Аналогичным образом выводится статистика и по другим филтрам (Рисунок бб, б-г).

Фильтр по институтам

- Казахстанский инженерно-технологический университет (32)
- КазНИИ почвоведения и агрохимии имени У. Чкалова (13)
- НИИ микробиологии и вирусологии (9)
- КазНИИ плодоовощеводства (1)

Фильтр по дате

- 1999 (1)
- 1996 (0)
- 2003 (5)
- 2017 (4)
- 2018 (4)
- 1977 (4)
- 2007 (2)
- 1960 (2)
- 2016 (2)
- 1958 (1)

а) Фильтр по институтам

б) Фильтр по дате

Фильтр по типам ресурсов

- Статьи (33)
- Бюль (10)
- ОРИ (1)

в) Фильтр по типам ресурсов

Фильтр по ключевым словам

- Статьи по теме (7)
- ТЕКСТОВЫЙ МАТЕРИАЛ
ОБЩИНА ЧАСТЕНАКУ (1)
- ОТВЕТЫ НА ВОПРОСЫ
ЧАСТЕНАКУ ВОПРОСЫ И
ПЕРВОМУ ЧАСТЕНАКУ
ЧАСТЕНАКУ (1)
- ВОПРОСЫ И ОТВЕТЫ
НА ВОПРОСЫ ЧАСТЕНАКУ (1)
- роль повышения порядка в об-
ществе (1)
- Порядок хранения и / или тип
порядка (1)
- Статьи (статейный список)
порядка (1)
- в результате которого
предоставляются персональные
данные (1)
- Точность сфер деятельности
лица (1)

г) Фильтр по ключевым словам

Рисунок 66. Панель фильтров

При нажатии на название найденного ресурса открывается страница репозитория (Рисунок 67). На данной странице указывается краткая аннотация ресурса, год издания, перечень авторов, название научно-исследовательского института, в котором создан данный ресурс. Перейдя по ссылке «Показать полную информацию» (Рисунок 68), можно получить дополнительную информацию, такую как ключевые слова, дату загрузки документа, язык документа, издатель, название журнала, тип ресурса.

Щелкнув по ссылке, указанной ниже слова «Открыть», пользователь имеет возможность скачать документ данного ресурса.

При нажатии на фамилию и инициалы автора в перечне найденных ресурсов (Рисунок 65) открывается профиль ученого, показанный на рисунке 69.

Ученый: Дžансугурова, Лjаjла Бугаjевна

Дžансугурова Лjаjла Бугаjевна

Ученый: Дžансугурова, Лjаjла Бугаjевна

Должность: Профессор

Основной вклад

Дžансугурова Лjаjла Бугаjевна является автором и соавтором более 40 научных публикаций, посвященных синтезу новых полимеров и изучению их свойств. В 1985 году защитила кандидатскую диссертацию на тему «Синтез и свойства новых полимеров». В 1995 году защитила докторскую диссертацию на тему «Синтез и свойства новых полимеров». В 2005 году была избрана членом-корреспондентом Академии наук Республики Узбекистан. В настоящее время работает в должности профессора кафедры химии в Институте химии Академии наук Республики Узбекистан.

С 1985 по 1995 гг. работала в должности ассистента кафедры химии в Институте химии Академии наук Республики Узбекистан.

С 1995 по настоящее время работает в должности профессора кафедры химии в Институте химии Академии наук Республики Узбекистан. В настоящее время работает в должности профессора кафедры химии в Институте химии Академии наук Республики Узбекистан.

Джансугурова Лjаjла Бугаjевна является автором и соавтором более 40 научных публикаций, посвященных синтезу новых полимеров и изучению их свойств. В настоящее время работает в должности профессора кафедры химии в Институте химии Академии наук Республики Узбекистан.

Рисунок 69. Профиль ученого

В нижней части данной странице приводится ссылки на труды данного ученого, загруженных в репозиторий цифровых объектов (Рисунок 70).

Джансугурова Лjаjла Бугаjевна. Синтез и свойства новых полимеров. Ташкент, 1985. 112 с.

Научные публикации

- 1. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1985. № 10. С. 10-12.
- 2. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1986. № 11. С. 11-13.
- 3. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1987. № 12. С. 12-14.
- 4. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1988. № 13. С. 13-15.

Публикации в рефератах

- 1. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1985. № 10. С. 10-12.
- 2. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1986. № 11. С. 11-13.

Публикации в цифровом репозитории

- 1. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1985. № 10. С. 10-12.
- 2. Джансугурова Лjаjла Бугаjевна, Бугаjев Р.М. Синтез и свойства новых полимеров // Журнал химии. Ташкент, 1986. № 11. С. 11-13.

Рисунок 70. Публикации в цифровом репозитории ученого

Стоит отметить, что при поиске ресурсов по автору учитывается разное написание имени ученого, в том числе на разных языках.

Раздел «Персоналий» портала содержит список ученых, чьи труды загружены в репозиторий. На начальной странице раздела отображается алфавит, буквы которого соответствуют первой букве фамилии автора (Рисунок 71).



Рисунок 71. Раздел «Персоналий»

Раздел «Институты» содержит информацию о научно-исследовательских институтах, входящих в научно-образовательный кластер (Рисунок 72).

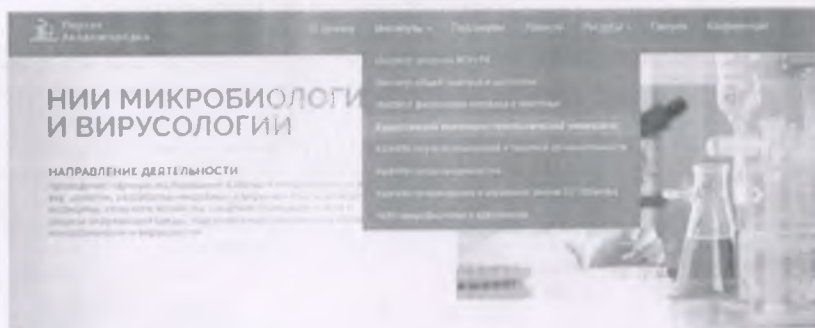




Рисунок 21. Результаты запроса, сделанного в репозиторий цифровых объектов

КазНИИ почвоведения и агрохимии имени У.У.Успанова

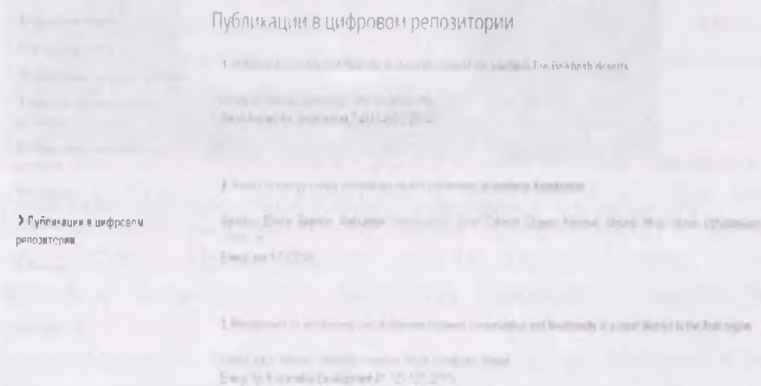


Рисунок 22. Список публикаций сотрудников научно-исследовательского института на веб-портале

Разработан механизм сопоставления автора с уже имеющимися в базе данных авторами при загрузке нового ресурса с целью

представления более релевантной информации пользователям. При появлении новых статей алгоритм перебирает все введенные имена и его вариации (например, с помощью транслитерации), и в случае нахождения, присваивает его к данной статье. Кроме того, в административной части веб-портала разработана страница для контент-менеджера (Рисунок 23), позволяющая управлять сопоставлением авторов: переназначать профиль, если алгоритм неправильно определил ученого, добавлять новые альтернативные имена и др.

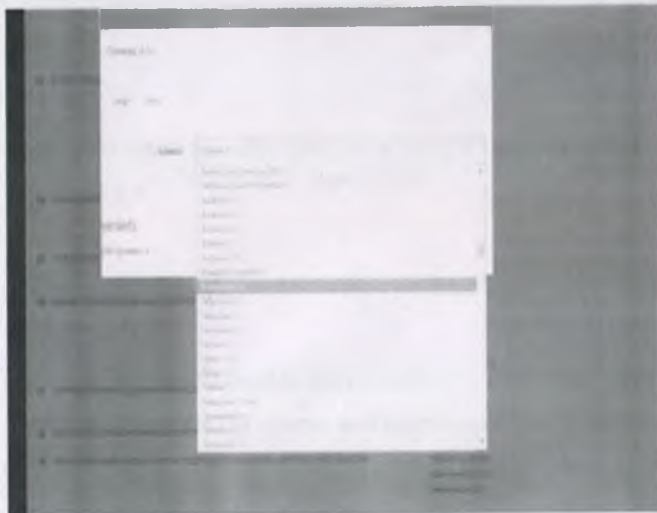


Рисунок 23. Управление процессом сопоставления авторов

2.5 Распознавание оцифрованных трудов для реализации полнотекстового поиска по репозиторию цифровых объектов

Согласно документации DSpace, при загрузке нового ресурса в репозиторий цифровых объектов и обновлении индекса, из документов формата PDF извлекается текст при условии, что текст данного документа был предварительно распознан. В этом случае DSpace предоставляет полнотекстовый поиск по данному документу. Технически, распознанный текст хранится в текстовом файле, расположенном в той же директории, что и загружаемый документ.

Поставлена задача генерировать при необходимости указанный выше текстовый документ более правильного содержания.

Наиболее приемлемое качество распознавания текста предоставляет коммерческое программное обеспечение компании Adobe. Согласно проведенному литературному обзору, из свободных технических решений наилучший результат показывает программа tesseract. Ее преимущество также заключается во встроенной поддержке сотни языков, в том числе и казахского. Однако проведенные тесты на оцифрованных трудах, показали, что поддержка казахского языка реализована достаточно слабо.

Существует коммерческий модуль «OCR & Transcription» для DSpace, разработанный организацией 4Science (www.4science.it/en/ocr-transcription), который обеспечивает интеграцию с внешней программой для распознавания текстов. Данный модуль также основан на применении программы tesseract.

Стоит отметить, что одной из особенностей tesseract является то, что она не работает с документами в формате PDF. Для этого исходный документ разбивается на файлы изображений встроенной утилитой pdf2ppm. Преимущество данной утилиты перед convert заключается в более гибкой настройке разрешения получаемых изображений.

Для документов, написанных на русском языке, tesseract может быть использован совместно с программой LanguageTool, которая проверяет грамматику, стиль и орфографию. Наряду с несколькими редакциями, существует версия для автономного использования. Системными требованиями программы являются Java 8+, Amazon Corretto 8 или IcedTea. Обзор не выявил подобных сервисов для казахского языка.

2.6 Модуль конференций

Разработан модуль конференций, позволяющий проводить организационную работу по регистрации участников конференции, сбору материалов, объявлению решений организационного комитета, загрузке квитанций по оплате и др.

Каждый участник конференции проходит обязательную регистрацию на портале Академгородка. Данные пользователи являются локальными пользователями приложения Django и не вносятся в службу каталогов.

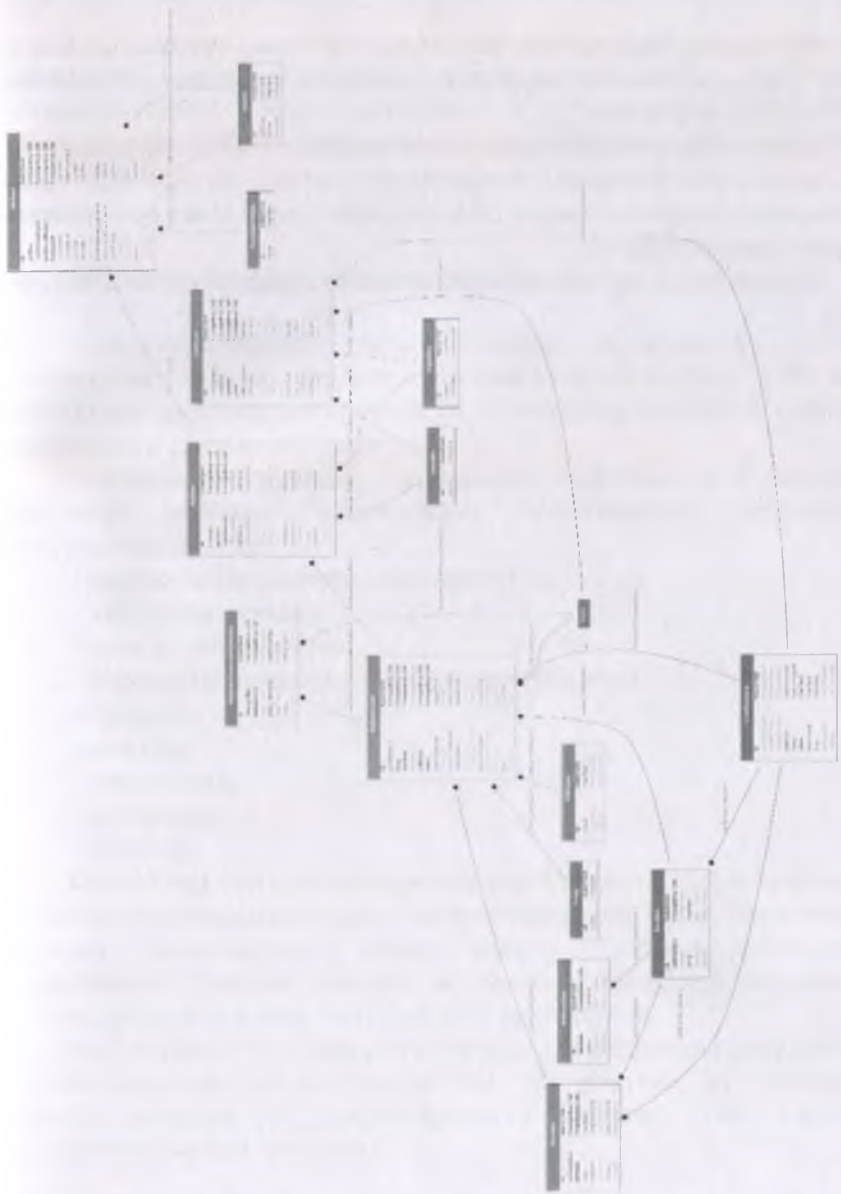


Рисунок 24. Схема данных модуля конференции

2.7 Система автоматического резервного копирования

Проведена работа по разработке системы автоматического резервного копирования данных репозитория цифровых объектов и портала Академгородка.

Помимо репликации данных на серверах ТОО «Академсеть», осуществляется резервное копирование данных на три облачных хранилища (Dropbox, Yandex.Disk и Google Drive) в автоматическом режиме каждые 8 часов.

Для работы с облачными хранилищами выбрана утилита RClone 1.4.2.

3 НАПОЛНЕНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ

На доменном имени asagor.kz действует полнофункциональный прототип распределенной информационной системы, обеспечивающий идентификацию и управление ресурсами, идентификацию, аутентификацию и авторизацию пользователей, сбор статистики, предоставление различной информации.

3.1 Административная часть портала

Административная часть портала представляет собой приложение Django, которое позволяет осуществлять действия, требующие повышенные привилегии. К текущему отчетному периоду реализованы следующие разделы:

- домашняя страница, содержащая информацию о текущем состоянии системы, залогиненных пользователей, управление пользователями и др.;

- научно-исследовательские институты;
- модуль интеграции;
- модуль конференции;
- управление баннерами на главной странице;
- галерея;
- новости;
- персонал;
- публикации;
- утилиты.

Общий вид приложения приведен на Рисунке 25. В левой части страницы отображены разделы административной части. Технически, разделы представляют собой модели Django, помеченные специальным флагом. Доступ к каждому разделу назначается индивидуально каждому пользователю приложения.

На Рисунке 25 отображена страница для редактирования списка научно-исследовательских институтов. На рисунке 26 приведена форма создания и редактирования записи для научно-исследовательского института.



Рисунок 25. Административная часть портала: список научно-исследовательских институтов



Рисунок 26. Административная часть портала: редактирование информации о научно-исследовательском институте

3.2 Загрузка оцифрованных научных трудов в репозиторий цифровых объектов

Процесс загрузки оцифрованных научных трудов в репозиторий цифровых объектов состоит из следующих шагов, проиллюстрированных на рисунках 27-55.

Для авторизации пользователя перейти на сайт asagor.kz, в правом верхнем углу нажать «Войти» (Рисунок 27).

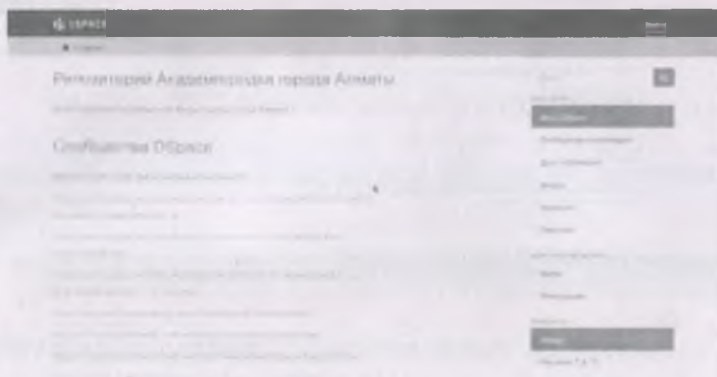


Рисунок 27. Главная страница репозитория

Ввести адрес электронной почты и пароль, нажать «Войти» (Рисунок 28).

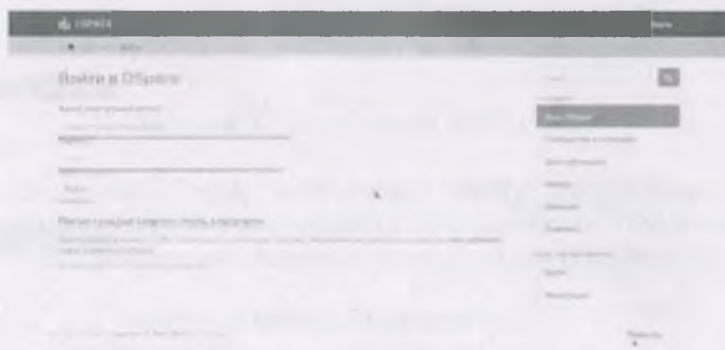


Рисунок 28. Страница аутентификации

На главной странице репозитория выбрать название института, к

которому относится данный ресурс. Например, «Казахский научно-исследовательский институт почвоведения и агрохимии имени У.У.Успанова» (Рисунок 29):

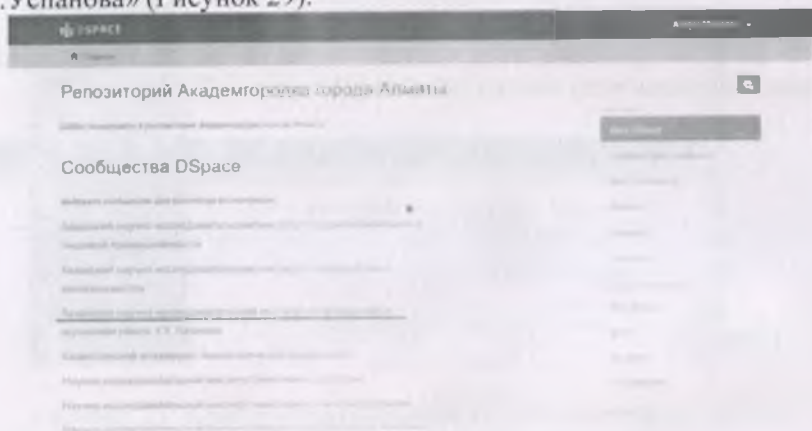


Рисунок 29. Выбор научно-исследовательского института

На появившейся странице перейти по ссылке «Труды сотрудников Института» (Рисунок 30).

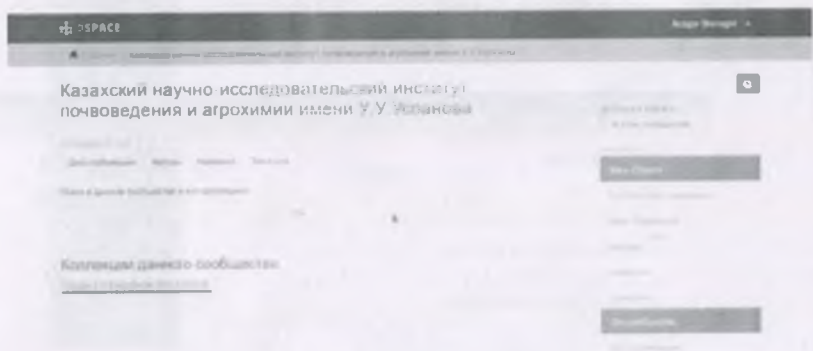


Рисунок 30. Выбор коллекции

Нажать «Добавить элемент в эту коллекцию» (Рисунок 31).

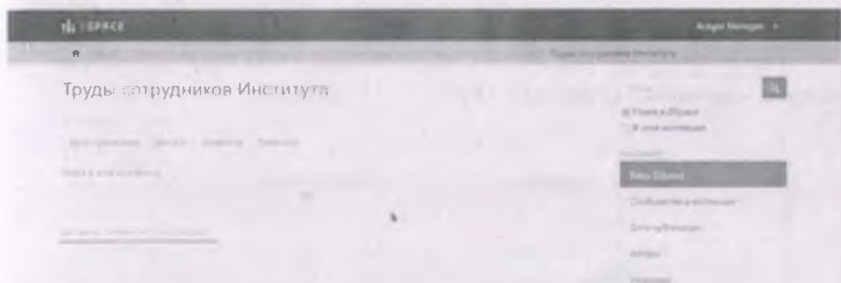


Рисунок 31. Добавление элемента в коллекцию

В результате открывается форма для заполнения метаданных ресурса (Рисунок 32).



Рисунок 32. Заполнение метаданных

На следующем этапе необходимо ввести информацию об авторах ресурса. Каждый автор вводится по отдельности. После ввода каждого автора необходимо нажимать кнопку «Add» (Рисунок 33).

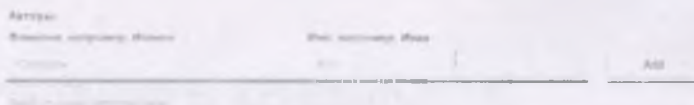


Рисунок 33. Заполнение авторов

В результате введенная информация появляется в нижней части блока «Авторы» (Рисунок 34).



Рисунок 34. Заполнение авторов

Аналогично вводится информация об остальных авторах (Рисунок 35).



Рисунок 35. Заполнение авторов

На следующем этапе вводится название ресурса (Рисунок 36). Поле «Другие названия» заполняется в случае, если данный ресурс имеет несколько альтернативных названий.



Рисунок 36. Заполнение названия ресурса

В поля «Дата публикации» вводятся год, месяц и день публикации ресурса (Рисунок 37). При этом обязательным полем

является только поле «Год».



Рисунок 37. Дата публикации ресурса

В поле «Издатель» вводится информация об издательстве (Рисунок 38).



Рисунок 38. Издательство

В поле «Наименование серии» вводится название журнала, в поле «Доклад номер» – том и номер (Рисунок 39).



Рисунок 39. Название журнала и том/номер

Информация о Международном стандартном книжном номере, Международном стандартном номере сериальных изданий вводится в поле «Идентификаторы» (Рисунок 40).

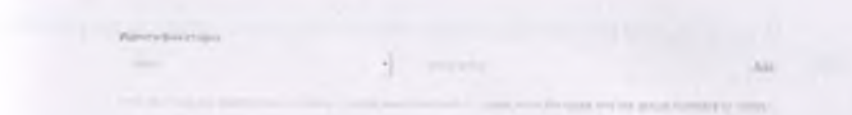


Рисунок 40. Заполнение ISSN/ISBN

Тип ресурса указывается в списке, показанном на рисунке 41.



Рисунок 41. Тип ресурса

Язык ресурса указывается в выпадающем списке, показанном на рисунке 42.



Рисунок 42. Язык ресурса

После этого шага нужно нажать кнопку «Следующий» (Рисунок 43).

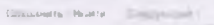


Рисунок 43. Переход к следующему этапу

Следующий этап заключается в заполнении ключевых слов загружаемого ресурса. Каждое ключевое слово вносится по отдельности. После ввода каждого ключевого слова нажимать «Add»:

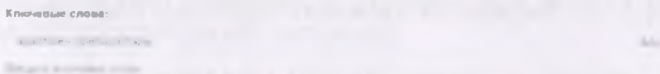


Рисунок 44. Заполнение ключевых слов

В результате оно отобразится в списке, как показано на рисунке 45.



Рисунок 45. Заполнение ключевых слов

Аналогично вводятся остальные ключевые слова (Рисунок 46).

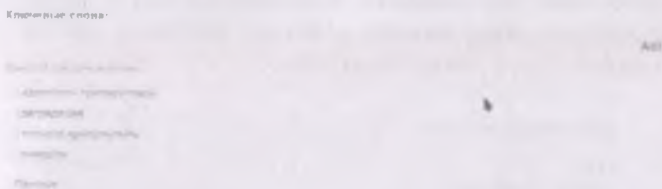


Рисунок 46. Заполнение ключевых слов

На следующем этапе заполняется краткая аннотация ресурса (Рисунок 47).

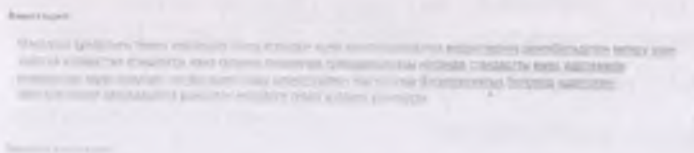


Рисунок 47. Заполнение аннотации

Поле «Спонсоры/источники финансирования» заполняется, если имеются источники финансирования (Рисунок 48). Данное поле можно оставить пустым.



Рисунок 48. Заполнение информации о спонсорах и источниках финансирования

После этого шага нажать «Следующий». Для загрузки файла с публикацией на сайт, нажать «Обзор» (Browse), найти и открыть файл в появившемся диалоговом окне.



Рисунок 49. Загрузка файла с публикацией

Если требуется загрузить больше одного файла, нажать «Загрузить документ и добавить еще один». В противном случае нажать «Следующий» (Рисунок 50).



Рисунок 50. Загрузка файла с публикацией

На следующей странице система предлагает проверить отправляемый документ (Рисунок 51).

Отправка элемента

Лицензию распространения

Остается последний шаг: выбрать лицензию. Выберите лицензию, которую вы хотите использовать для своего материала, или выберите лицензию по умолчанию.

Предоставьте стандартную лицензию распространения. Выберите лицензию, которую вы хотите использовать.

NOTE: PLACE YOUR OWN LICENSE HERE. The default license is provided for illustrative purposes only.

NON-EXCLUSIVE DISTRIBUTION LICENSE

By viewing and submitting this license, you are indicating that you have the right to grant the copyright owner the right to reproduce, translate, and otherwise use the material in any form and electronic format and in any medium, including but not limited to, web or print.

You agree that DSpace will not attempt to contact you about the use of your content in any form or medium, or to provide you with a copy of your content.

You also agree that DSpace will keep your information confidential, and will not share it with any third party.

You represent that the submission is your original work, and that you have the right to grant the copyright owner the right to use your submission in any form and electronic format, and that you have the right to grant the copyright owner the right to use your submission in any form and electronic format.

If the submission contains material for which you do not have copyright, you represent that you have obtained the authorized permission of the copyright owner to grant DSpace the right to use the material, and that you have the right to grant the copyright owner the right to use the material in any form and electronic format, and that you have the right to grant the copyright owner the right to use the material in any form and electronic format.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN DSPACE, YOU AGREE THAT YOU WILL MAINTAIN THE AGENCY OR ORGANIZATION'S REQUIREMENTS FOR PUBLIC DOMAIN STATUS IN ANY MEDIUM.

Рисунок 53. Лицензия

На данной странице необходимо прокрутить страницу вниз, отметить галочку «Предоставить лицензию» и нажать «Завершить отправку» (Рисунок 54).

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN DSPACE, YOU AGREE THAT YOU WILL MAINTAIN THE AGENCY OR ORGANIZATION'S REQUIREMENTS FOR PUBLIC DOMAIN STATUS IN ANY MEDIUM.

Если у Вас возникли какие-либо вопросы, пожалуйста, свяжитесь с нами по адресу: help@dspace.org или help@dspace.org.

Лицензия распространения

Предоставить лицензию

[Назад](#) [Отмена](#) [Завершить отправку](#)

Рисунок 54. Завершение отправки

В результате выходит сообщение об успешном добавлении ресурса в репозиторий (Рисунок 55).

Отправка завершена

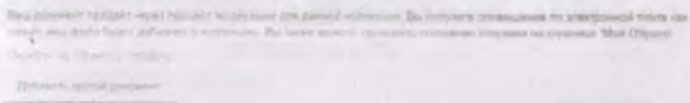


Рисунок 55. Информация об успешной загрузке документа

Если требуется загрузить еще один ресурс, который относится к этому же институту, то нажать «Добавить другой документ». В противном случае необходимо нажать «DSPACE» в левом верхнем углу и повторить шаги сначала.

3.4 Модуль конференций

В административной части модуля конференции (Рисунок 56) имеется возможность установить один из статусов заявки (Принята, Отклонена, Возвращена на доработку), просмотреть историю изменения статусов, а также текст отправленных электронных сообщений (Рисунок 57), просмотреть статистику в разрезе секций, скачать загруженные материалы конференции в zip-архиве (Рисунок 58).

Кроме того, реализована возможность подачи заявки на бронирование гостиницы, указания времени прибытия участника. Модуль может использоваться для любого количества конференций, причем для каждой конференции может быть задана различная конфигурация (правила приема докладов, уровни доступа и др.).



Рисунок 56. Административная часть модуля конференции:
список заявок



Рисунок 57. Административная часть модуля конференции:
детали заявки



Рисунок 58. Статистика заявок

Модуль успешно прошел апробацию на Международной конференции «Наука, производство, бизнес: современное состояние и пути инновационного развития аграрного сектора на примере Агрохолдинга «Байсерке-Агро»», проведенной 4-5 апреля 2019 года в г. Алматы. Всего на конференцию поступило 347 работ, из них 299 приняты, 36 отклонены, в 22 случаях подача заявки не завершена. Кроме того, разработана веб-страница, на которой участник может зарегистрировать себя в качестве слушателя.

4 ИСПОЛЬЗОВАНИЕ ИНТЕГРИРОВАННОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

4.1 Использование портала

Для доступа ко всем функциям портала Академгородка необходимо произвести регистрацию пользователя. Для этого необходимо нажать на кнопку «Войти» в нижней части страницы.

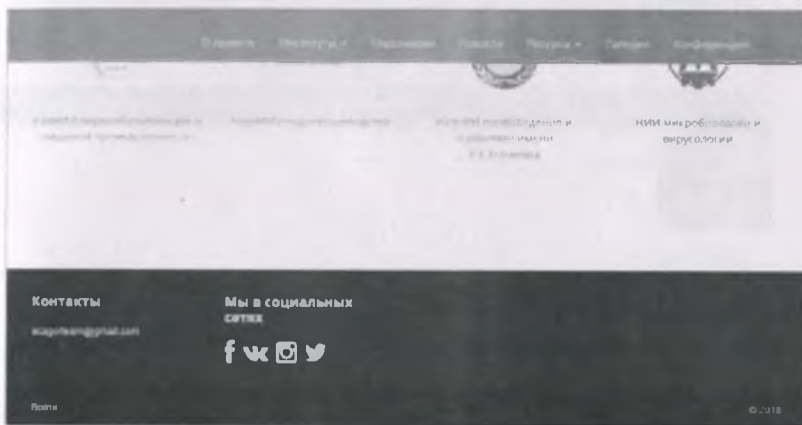


Рисунок 59. Кнопка «Войти»

В результате пользователь перенаправляется на страницу аутентификации, показанная на рисунке 60.



Рисунок 60. Страница аутентификации

Для создания новой учетной записи необходимо нажать на кнопку «Регистрация».

Форма регистрации, показанная на рисунке 61, содержит следующие обязательные поля:

- Фамилия, имя и отчество;
- Адрес электронной почты;
- Пароль;
- Подтверждение пароля.

The image shows a screenshot of a web registration form. At the top, there is a dark header bar with a logo on the left and navigation links on the right. Below the header, the main content area is titled 'Регистрация'. The form contains several input fields: 'Имя', 'Фамилия', 'Адрес электронной почты', 'Пароль', and 'Подтверждение пароля'. There is a 'Регистрация' button at the bottom of the form. The background is slightly blurred, showing other parts of the website.

Рисунок 61. Форма регистрации нового пользователя

К паролю выдвигаются следующие требования:

1. Пароль не должен совпадать с вашим именем или другой персональной информацией или быть слишком похожим на неё;
2. Пароль должен содержать как минимум 8 символов;
3. Пароль не может быть одним из широко распространённых паролей;
4. Пароль не может состоять только из цифр.

После ввода данных пользователя и нажатия кнопки «Регистрация» на указанную почту приходит письмо с кодом. Для завершения регистрации его нужно ввести в поле, показанное на рисунке 62 и нажать «Ввести код».

Рисунок 72. Раздел «Институты»

Данный раздел содержит информацию об истории института, его деятельности, основных направлениях научной деятельности, структуре, наиболее крупных научных результатах, списке сотрудников, а также контактах (Рисунок 73).

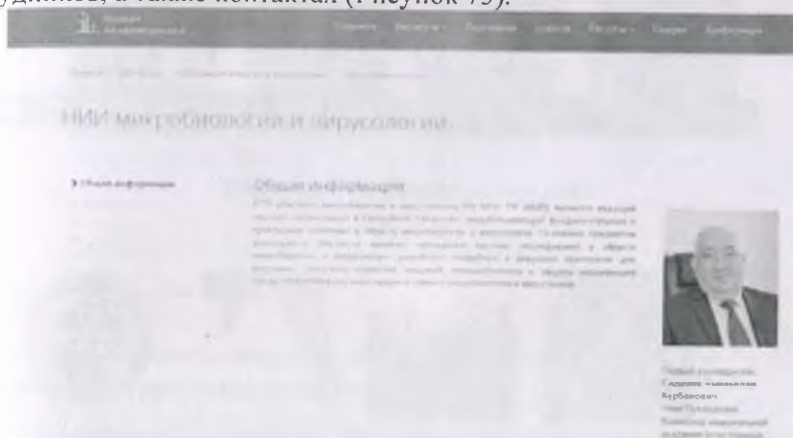


Рисунок 73. Информация об институте

Кроме того, раздел содержит перечень публикаций сотрудников института, загруженных в репозиторий цифровых объектов (Рисунок 74).

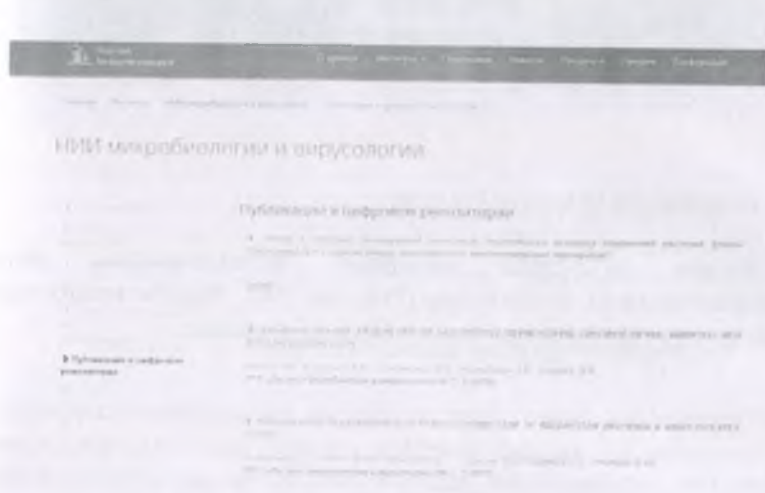


Рисунок 74. Перечень публикаций ученых института, загруженных в репозиторий

Раздел «Новости» предназначен для освещения наиболее важных событий в жизни институтов, предстоящих научных мероприятий.



Рисунок 75. Раздел «Новости»

Раздел «Галерея» содержит фотоматериалы научно-исследовательских институтов (Рисунок 76). Раздел предоставляет возможность группировки фотографий по тематике.



Рисунок 76. Раздел «Галерея»

4.2 Модуль конференций

Главная страница модуля для пользователей (участников конференции) представлена на Рисунке 77.



Рисунок 77. Главная страница модуля конференций

Процесс подачи заявки состоит из трех этапов. На первом этапе вводится название, язык доклада и форма участия. На втором этапе вводится информация об авторах доклада, место работы, контактная

информация. На третьем этапе производится загрузка файла с докладом. Уведомление о решении организационного комитета автоматически отправляется на электронную почту участника.

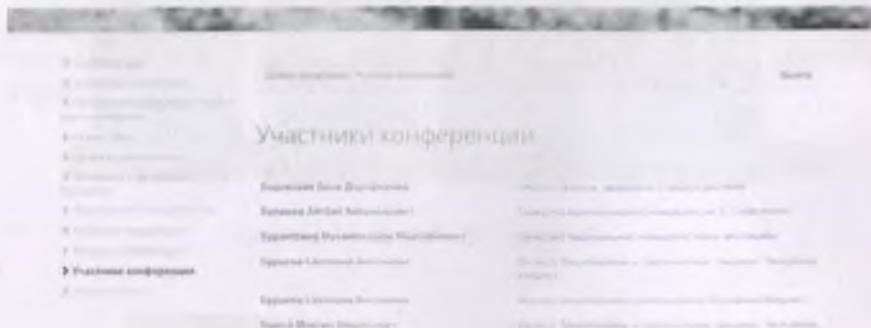


Рисунок 78. Список участников конференции

4.3 Анализ опытной эксплуатации РИС и анализ удовлетворенности пользователей

Согласно данным сайта su-pr.com, в поисковом сервисе Яндекс проиндексировано 7112 страниц, в поисковом сервисе Google – 3470 страниц портала и репозитория.

Время ответа от сервера и время до загрузки DOM страниц ресурса в разрезе стран приведено на рисунке 79. Продолжительность установки соединения для пользователей из Республики Казахстан составляет 0.210 секунд, из Российской Федерации – 0.170 секунд. Наихудший показатель (0.549 секунд) приходится на пользователей из Италии. Время до загрузки DOM для пользователей из Республики Казахстан составляет 2.625 секунд, из Российской Федерации – 2.092 секунд. Наихудший показатель (12.559 секунд) приходится на пользователей из Кипра.



Рисунок 79. Время ответа от сервера (слева), время до загрузки DOM (справа)

Согласно метрикам Yandex, среднемесячная аудитория сайта составляет 375,71 человек. Наибольший ник показателя наблюдается в марте 2019 года – 804 человек и в феврале 2019 года – 728 человек во время проведения конференции. Подавляющее большинство пользователей – из Республики Казахстан и Российской Федерации (Рисунок 80). Во время проведения конференции на сайте Академгородка с 20 января по 5 апреля 2019 года зарегистрирован 231 пользователь.

Страна браузера

1 янв — 7 окт 2019 Детализация: по дням

Просмотры

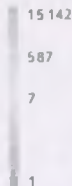


Рисунок 80. Распределение пользователей ресурса в разрезе стран

При этом 49% всех посещений являются прямыми заходами на ресурс, 24.5% – внутренними переходами между страницами сайта, доля пользователей, перешедших из поисковых систем – 20% (Рисунок 81).



Рисунок 81. Источники трафика

Среднее время (за год), проведенное пользователями ресурса на

сайте составляет 6 мин. 21 сек (Рисунок 82). Наибольший пик наблюдается в середине января – 37.5 мин.

Время на сайте

1 авг. 7 лет 2019 Детализация по дням

Время на сайте



Рисунок 82. Среднее время, проведенное пользователями на сайте

Глубина просмотра страниц ресурса приведена на Рисунке 83.

Глубина просмотра

1 авг. 7 лет 2019 Детализация по дням

Глубина просмотра



Рисунок 83. Глубина просмотра

Среднегодовой показатель отказа пользователей (отсутствие перехода на другие ссылки ресурса после первого посещения) составляет 12.9%; наибольший всплеск показателя (до 50-60%)

наблюдается в отдельных сутках летних месяцев.

Распределение пользователей по типу устройств: 83.2% – ПК, 16.6% – смартфоны, менее 1% – планшеты. Распределение возраста посетителей ресурса: 18-24 года – 34.4%, 25-34 года – 33.8%, 8.28% – 35-44 года, 12.9% – 45-54 года, 7.77% – 55 лет и старше.

Проанализирована статистика запросов в репозиторий цифровых объектов. 28.49% всех поисковых запросов приходится на поиск ресурсов по ключевым словам (из них наибольший процент – 8.94% по ключевому слову «vegetable», 3.35% – «молочнокислые микроорганизмы»); 35.75% на поиск конкретных авторов; 24.58% – на поиск ресурсов по годам.

ЗАКЛЮЧЕНИЕ

Таким образом, в данном учебном пособии освещена актуальность создания информационного ресурса для решения ряда задач в области агропромышленного комплекса. Приведены задачи создания веб-портала asagor.kz, описаны этапы его проектирования, реализации, внедрения и поддержки.

Перед разработкой архитектуры системы проведен анализ потребностей и возможностей научно-исследовательских институтов – участников научно-образовательного кластера. Определена оптимальная архитектура распределенной информационной системы, приведены описание ее подсистем и схема их взаимодействия. Разработано техническое задание для разработки программной части распределенной библиотечной информационной системы. Разработана модель функционирования распределенной информационной системы. Произведен отбор необходимого открытого и свободного программного обеспечения – компонентов подсистем распределенной информационной системы. Проведен анализ институциональных репозиториев и систем для создания электронных библиотек, изучен опыт их использования другими организациями. В результате выбрана и установлена распределенная информационная система DSpace, произведена модификация и адаптация DSpace к условиям, установленным в Республике Казахстан. Выбрано интегрирующее программное обеспечение – распределенная информационная система ZooSPACE, производится адаптация данной системы к распределенной информационной системе Академгородка.

Разработана архитектура портала Академгородка, схема его взаимодействия с остальными составляющими распределенной информационной системы, произведено описание программного обеспечения, использованного для его функционирования; произведено описание разделов портала. Осуществлена интеграция портала Академгородка с репозиторием цифровых объектов с использованием программного интерфейса DSpace REST API.

Разработанный веб-портал предназначен для поддержки научно-педагогической деятельности по обеспечению потребностей участников научно-образовательного кластера и основан на передовых информационных технологиях.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Saini O. P. The emergence of institutional repositories: a conceptual understanding of key issues through review of literature // *Library Philosophy and Practice*. – 2018. – Vol. 1774. – 19 p.

2 Franchke H., Gamalielsson J., Lundell B. Institutional repositories as infrastructures for long-term preservations // *Information Research*. – 2016. – Vol. 22 (2), No. 757. – 27 p.

3 Hippenhammer C. Comparing institutional repository software: pampering metadata uploaders // *The Christian Librarian*. – 2016. – Vol. 59, No. 1 – 6 p.

4 Baughman K., McDowell. Institutional repositories in the Czech republic // *Gleeson Library Librarians Research*. – 2016. – Vol. 10. – 29 p.

5 Ravikumar M. N., Ramanan T. Comparison of greenstone digital library and DSpace: Experiences from digital library initiatives at eastern university, Sri Lanka // *Journal of University Librarians Association of Sri Lanka*. – 2014. – Vol. 18, No. 2. – P. 76–90.

6 Castagné M. Institutional repository software comparison: DSpace, ePrints, Digital Commons, Islandora and Hydra (Report). – University of British Columbia, 2013. – 15 p.

7 Cullen R., Chawner B. Institutional repositories in New Zealand: comparing institutional strategies for digital preservation and discovery // *Proceedings of the IATUL Conference*. – 2008. – Vol. 18. – 11 p.

8 Скачков Д. М., Жижимов О. Л. Технология географического поиска информации в «негеографических» информационных системах // *Научные и организационно-технологические основы интеграции цифровых информационных ресурсов*. – 2013. – С. 74–101.

9 Шокин Ю. И., Жижимов О. Л., Федотов А. М. Информационные системы ИВТ СО РАН // *Труды XVI Всероссийской конференции DICR-2017*. – 2017. – С. 11–18.

10 Chudlarsky T., Dvorak J. A National CRIS Infrastructure as the Cornerstone of Transparency in the Research Domain. In: Jeffery, Keith G; Dvorak, Jan (eds.): *E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic)*. – 2012. - P. 9-17.

11 Guskov A.E., Zhizhimov O.L., Kikhtenko V., Skachkov D.M.,

Kosyakov D. RuCRIS: A Pilot CERIF based System to Aggregate Heterogeneous Data of Russian Research Projects // Procedia Computer Science. - 2014. - Vol.33. - P.163-167.

12 Жижимов О.Л., Федотов А.М., Федотова О.А. Построение типовой модели информационной системы для работы с документами по научному наследию // Вест. НГУ. Информационные технологии. - 2012. - Т.10, №3. - С.5-14.

13 Шокин Ю.И., Федотов А.М., Жижимов О.Л., Федотова О.А. Система управления электронными библиотеками в ИРИС СО РАН // Инфраструктура научных информационных ресурсов и систем: Сборник научных статей Четвертого Всероссийского симпозиума / Под редакцией доктора технических наук Е.Б. Кудашева, доктора физико-математических наук В.А. Серебрякова. - 2014. - Т.1. - Москва: Вычислительный центр РАН. - С.11-39.

14 Жижимов О.Л., Федотов А.М., Шокин Ю.И. Технологическая платформа массовой интеграции гетерогенных данных // Вестник Новосибирского государственного университета. Серия: Информационные технологии. - 2013. - Т.11. - № 1. - С.24-41.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	3
1 НАЗНАЧЕНИЕ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ, АНАЛИЗ ВОЗМОЖНОСТЕЙ И ПОТРЕБНОСТЕЙ УЧАСТНИКОВ НАУЧНО-ОБРАЗОВАТЕЛЬНОГО КЛАСТЕРА ДЛЯ ПОСТРОЕНИЯ ЕЕ ОПТИМАЛЬНОЙ АРХИТЕКТУРЫ.....	6
1.1 Актуальность создания информационной системы для агропромышленной отрасли	6
1.2 Определение потребностей участников научно-образовательного кластера	10
1.3 Анализ возможностей участников научно-образовательного кластера	11
2 РАЗРАБОТКА ИНТЕГРИРОВАННОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ АКАДЕМГОРОДКА ГОРОДА АЛМАТЫ	16
2.1 Аппаратная часть распределенной информационной системы.....	16
2.2 Модель функционирования распределенной информационной системы	20
2.3 Программная часть распределенной информационной системы	21
2.3.1 Подсистема "Репозиторий цифровых объектов"	22
2.3.2 Подсистема управления текущей научно-исследовательской информацией	26
2.3.3 Подсистема интеграции распределенных информационных ресурсов.....	28
2.3.4 Подсистема доступа к распределенным информационным ресурсам на основе Web-технологий	30
2.3.5 Централизованное хранилище учетных записей пользователей	34
2.3.6 Механизм автоматического деплоймента портала	37
2.4 Интеграция DSpace и портала Академгородка	38
2.5 Распознавание оцифрованных трудов для реализации полнотекстового поиска по репозиторию цифровых объектов.....	47
2.6 Модуль конференций	48
2.7 Система автоматического резервного копирования	50
3 НАПОЛНЕНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ	51
3.1 Административная часть портала	51
3.2 Загрузка оцифрованных научных трудов в репозиторий цифровых объектов	52
3.4 Модуль конференций	63

4 ИСПОЛЬЗОВАНИЕ ИНТЕГРИРОВАННОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ.....	65
4.1 Использование портала	65
4.2 Модуль конференций	77
4.3 Анализ опытной эксплуатации РИС и анализ удовлетворенности пользователей	78
ЗАКЛЮЧЕНИЕ	83
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	84