



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 4 за 2013 г.



Кирсанов С.В., [Коваленко В.Н.](#),
[Коваленко Е.И.](#), [Куликов А.Ю.](#),
Шифрин М.А.

Программный комплекс
интеграции архивов
медицинских изображений

Рекомендуемая форма библиографической ссылки: Программный комплекс интеграции архивов медицинских изображений / С.В.Кирсанов [и др.] // Препринты ИПМ им. М.В.Келдыша. 2013. № 4. 24 с. URL: <http://library.keldysh.ru/preprint.asp?id=2013-4>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В. Келдыша
Российской академии наук**

**С.В. Кирсанов, В.Н. Коваленко, Е.И. Коваленко,
А.Ю. Куликов, М.А. Шифрин**

**Программный комплекс интеграции
архивов медицинских изображений**

Москва — 2013

**С.В. Кирсанов, В.Н. Коваленко, Е.И. Коваленко, А.Ю. Куликов,
М.А. Шифрин**

Программный комплекс интеграции архивов медицинских изображений

В работе рассматривается программный комплекс, обеспечивающий удаленный доступ к множеству медицинских изображений, которые хранятся в стандартном виде на множестве серверов. Назначение комплекса - поддержка медицинских специалистов в получении исследований, проведенных на диагностических установках различных подразделений и учреждений. В работе описывается структура комплекса, метод интеграции информации, способ подключения внешних приложений визуализации.

Ключевые слова: интеграция данных, интеграция приложений, PACS серверы, медицинские изображения, информационный грид

S.V. Kirsanov, V.N. Kovalenko, E.I. Kovalenko, A.Y. Kulikov, M.A. Shifrin

Program complex for integration of medical image archives

The paper presents the program complex providing remote access to a set of medical images, stored in standard representation on multiple image servers. The aim of the complex - support of medical experts in receiving studies which have been carried out on diagnostic installations of various divisions and institutions. The paper contains the description of the complex structure, the information integration method, the way of linking with external visualization applications.

Key words: data integration, application integration, PACS, medical image, informational Grid

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 11-07-00147-а.

1. Введение

В работе рассматривается задача создания программного комплекса, обеспечивающего удалённый доступ к медицинским изображениям, которые производятся и хранятся в стандартной цифровой форме на различных серверах в одной или нескольких независимых организациях. В результате интеграции множество серверов образуют "виртуальный архив" с единой точкой доступа. Прикладное назначение программного комплекса – поддержка профессиональной деятельности медицинских работников в поиске исследований, проведённых на различных установках, и визуализация полученных в них изображений на специализированных рабочих местах.

Рассматриваемая задача относится к тематике интеграции распределённой информации. Применение технологий интеграции способствует решению задач управления, планирования, контроля в различных областях социальной жизни и производства, а также исследовательских задач, которые направлены на разработку новых методов. Оценивая потребности медицины, мы исходим из того, что качество диагностики и лечения зависит от полноты информации о состоянии пациента, в том числе в ретроспективном плане. Важная часть этой информации представлена в виде диагностических изображений различных модальностей: томографических, рентгеновских, ультразвуковых и др.

Настоящая работа имеет две основные цели. Первая – обеспечить врачам разных специальностей доступ к исследованиям, выполненным в рамках одного учреждения. Такая потребность возникает из-за того, что данные исследований порождаются на специализированных установках различных подразделений, но для принятия решения специалисту могут потребоваться результаты любого из проведенных исследований. Вторая цель – выход за границы отдельных учреждений и обеспечение возможности получения данных из инфраструктур, в том числе крупномасштабных, в которые интегрированы автономные системы хранения независимых учреждений. Подобная цель представляет интерес как для обычной практики, так и для исследовательского направления медицины.

Медицинская тематика имеет свою специфику, однако в работе показывается, что многие задачи интеграции могут быть решены с помощью общих технологий и методов, не привязанных к конкретной прикладной области. К их числу относятся технологии информационного грида, которые опираются на стандартизованные формы удалённого доступа к распределённым базам данных. Соответствующие стандарты, заданные в спецификации WS-DAI (Web Service Data Access and Integration) [1], основаны на современных принципах построения распределённых систем – архитектуре грида OGSA (Open Grid Services Architecture) и Web-служб. Стандарты удалённого доступа к базам данных имеют реализацию в комплексе OGSA-DAI/DQP [2], который содержит программные интерфейсы для дистанционного выполнения информационных запросов к СУБД реляционного и XML типов.

Язык запросов позволяет выполнять операции над данными, полученными из разных баз, а также операции передачи и преобразования данных.

Кроме того, в работе используются методы интеграции данных, развитые в многочисленных исследованиях. Направленность этих методов – представление совокупности распределённых гетерогенных баз в виде единого информационного пространства с общей схемой данных и обеспечение к ним доступа, который не зависит от их физического расположения и способа хранения [3].

Работа исходит из следующего представления об организации данных медицинских диагностических исследований и базовых способах управления этими данными. Современные диагностические установки в рамках одного исследования порождают один или серию файлов цифровых изображений. Установки снабжены программными системами управления, с помощью которых врач аннотирует полученные данные дескриптивной информацией – сведениями об обследуемом пациенте, условиях проведения исследования и т.д. Для постоянного размещения полученные результаты передаются на сервер хранения. С точки зрения архитектуры такие серверы представляют собой архивы изображений – PACS (Picture Archiving and Communication System), в которых сочетается файловая система для хранения файлов изображений и база данных, содержащая дескриптивную информацию об исследованиях совместно с указателями на соответствующие файлы изображений.

Хотя PACS-серверы могут быть реализованы разными способами, широко используется стандарт DICOM (Digital Imaging and Communications in Medicine) [4], принятый Международной организацией по стандартизации как ISO standard 12052:2006 "Health informatics – Digital imaging and communication in medicine (DICOM) including workflow and data management". Этот стандарт, в частности, унифицирует удалённый доступ к PACS-серверам, определяя информационную модель объектов исследований и ряд сетевых служб. Служба DICOM Store поддерживает запись данных исследований в архив, для поиска/получения данных из архива предназначена служба DICOM Query/Retrieve, а также WADO (Web Access to DICOM Persistent Objects), основанная на протоколе HTTP/HTTPS.

Стандартизация на основе DICOM даёт возможность создавать приложения, способные работать с любыми PACS-серверами независимо от способа их реализации. Однако основная часть используемых на практике приложений предназначена для работы с единственным PACS-сервером. Достигнутый уровень стандартизации медицинской аппаратуры, развитие сетевых коммуникаций и технологий распределённых систем позволяет ставить вопрос о создании программных средств, которые дают возможность проводить поиск исследований по множеству распределённых PACS-серверов, получать файлы изображений и визуализировать их с помощью используемых на практике аналитических программ. Далее описываются возможности, архитектура и способ реализации разработанного комплекса для интеграции

архивов медицинских изображений. В разделе 6 приводится обзор исследований и разработок по этой теме.

2. Состав комплекса и назначение компонентов

Разработанный комплекс содержит следующие составляющие:

- управляющее приложение (M-PACS);
- система интеграции реляционных баз данных MQ-DAI;
- внешние программы визуализации изображений.

M-PACS поддерживает графический интерфейс для задания параметров поиска исследований, вывода результатов поиска и отбора исследований для визуализации. В M-PACS также реализованы операции доставки файлов изображений и механизмы вызова внешних программ их визуализации – DICOM-браузеров. Отметим, что вызов браузеров осуществляется программно, и в этом плане M-PACS выполняет функцию интеграции приложений. M-PACS является Web-приложением, то есть устанавливается и функционирует на некотором Web-сервере, а доступ пользователей к M-PACS осуществляется через Web-браузер, в среде которого генерируется описываемый далее интерфейс. Система интеграции баз данных MQ-DAI берёт на себя наиболее ресурсоёмкую часть обработки – поиск исследований по всему множеству подключённых к M-PACS архивов. Кроме того, MQ-DAI содержит набор административных средств для подключения PACS-серверов и образования из них информационной инфраструктуры. Эти средства дают возможность администраторам серверов самостоятельно подключать свои серверы к инфраструктуре, также как и отключать их от неё.

3. Сценарий работы с приложением M-PACS

Приложение M-PACS может использоваться как независимо, так и в составе какой-либо лабораторной или клинической информационной системы (КИС). Примером может служить система электронной карты e-MED [5,6], используемая в НИИ нейрохирургии имени Н.Н.Бурденко. В этом варианте M-PACS вызывается из медицинской системы для получения исследований обслуживаемого врачом пациента. При вызове M-PACS с помощью специального элемента интерфейса автоматически формируется и передаётся набор значений параметров, по которым осуществляется поиск исследований. Если приложение используется независимо, значения параметров вводятся вручную (рис. 1).

Поиск исследований: DCM | RAD

Фамилия жукова	Пол F	Дата (От)	Время (От)	поиск	очистить
Модальность	Дата рождения	Дата (До)	Время (До)		

Рис. 1. Панель ввода поисковых параметров.

В число параметров поиска входят:

- Фамилия, пол и дата рождения пациента.
- Модальность исследования.
- Нижние и верхние границы даты и времени проведения исследования.

По полученным одним из двух способов значениям параметров приложение M-PACS строит поисковый запрос на языке, являющимся расширением стандарта SQL-92, и передаёт его серверу системы MQ-DAI. Запрос является массовым, то есть он позволяет получить данные об исследованиях от всех PACS-серверов, подключённых к серверу MQ-DAI. Найденные исследования отображаются на Web-странице вместе с панелью ввода поисковых параметров (рис. 2).

Поиск исследований: DCM | RAD

Фамилия жукова	Пол F	Дата (От)	Время (От)	поиск	очистить
Модальность	Дата рождения	Дата (До)	Время (До)		

Результаты поиска

просмотр исследований в

Сервер	Фамилия	Дата рождения	Пол	Время	Описание исследования	Описание серии	Дата	Модальность
<input checked="" type="checkbox"/>	DCM	ЛУКОВА I.L.	F	122444.648	6+OTD.		20111129 CT	
<input checked="" type="checkbox"/>	DCM	ZHUKOVA I.L.	F	104724.768	6+OTD.		20111020 CT	
<input type="checkbox"/>	DCM	ZHUKOVA V.P.	F	18991230	114048	chest	thorax+bucky+PA	20111103 CR

Рис. 2. Результаты поиска исследований.

При поиске решается проблема транслитерации. Она заключается в том, что в медицинских системах именные данные обычно представляются в соответствии с отечественными правилами (фамилия, имя, отчество) на русском языке, в то время как актуальная версия стандарта DICOM не предусматривает использования кириллицы (а так же латиницы с диакритическими знаками). Это приводит к тому, что каждый специалист,

работающий на установке, регистрирует пациентов, вручную транслитерируя их именные данные в латинский алфавит. В результате даже на одном PACS-сервере именные данные одного и того же человека могут различаться¹.

Способ поиска учитывает это обстоятельство. При подготовке запроса программа M-PACS выполняет однозначную транслитерацию именных данных, переводя их в латинский алфавит. При поиске исследований на PACS-серверах отбираются те из них, которые “похожи” на латинскую транслитерацию. Найденные исследования ранжируются по степени близости значений именных данных, и в результате пользователь получает список, из которого он может выбрать требуемые исследования. Эту ситуацию иллюстрируют строки 1 и 2 на рис. 2.

В результатах поиска приводится дополнительная информация об исследованиях, которая имеется на PACS-серверах, но может быть неизвестна пользователю. Кроме того, в ходе поиска извлекаются служебные данные – идентификаторы исследований и серий, которые служат далее для извлечения файлов исследований.

Следующий шаг сценария – визуализация изображений. Пользователь отмечает несколько исследований, которые он хочет просмотреть, и нажимает одну из двух кнопок визуализации Workstation/Radscaper (рис. 2). M-PACS производит доставку файлов выбранных исследований и в отдельном окне запускает соответствующую программу визуализации.

Одна из подключённых к M-PACS программ визуализации – Windows-приложение ClearCanvas Workstation [7]. Это приложение свободно распространяется, широко используется специалистами и обладает разнообразными возможностями анализа, в том числе:

- визуализация изображений с группировкой их по сериям;
- обработка оптических характеристик изображения (изменения яркости, контраста, инвертирование цветов);
- поворот изображений, увеличение отдельных частей;
- вывод детальной технической информации об исследовании;
- замер расстояний и площадей на изображении.

Окно ClearCanvas Workstation показано на рис. 3.

¹ Следует заметить, что проблемы транслитерации можно избежать, если на диагностической установке имеется опция Modality Worklist, позволяющая принимать заявки на исследования из других информационных систем, которые могут производить автоматическую транслитерацию. Но даже при наличии этой опции проблема может оставаться для старых исследований, в которых транслитерация производилась вручную.

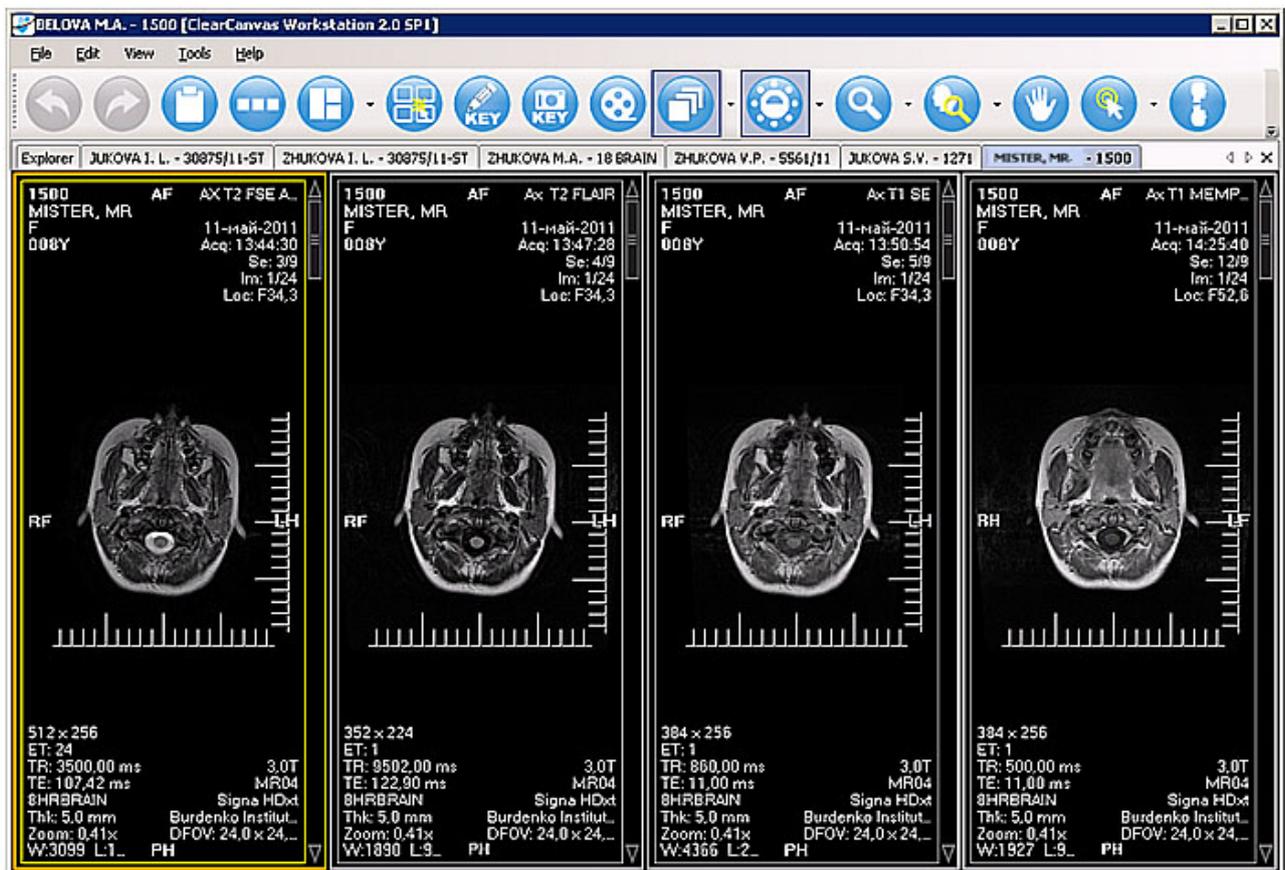


Рис. 3. Окно визуализации изображений браузера ClearCanvas Workstation.

Аналогичными, но несколько меньшими возможностями обладает второй используемый DICOM-браузер – Radscaper [8]. В сравнении с ClearCanvas Workstation этот браузер имеет два преимущества. Во-первых, он может работать с PACS-серверами, отличными от ClearCanvas Image Server, на который ориентирован браузер ClearCanvas Workstation, так как взаимодействует с ними только по стандартам DICOM. Из-за этого, однако, увеличивается время доставки файлов изображений (в ClearCanvas Workstation ускорение достигается за счёт использования нестандартной службы, установленной на Image Server). Во-вторых, Radscaper реализован в виде апплета, то есть не требует установки на компьютере пользователя дополнительного приложения.

При визуализации с помощью программы Radscaper для каждого исследования открывается новое окно непосредственно в Web-браузере (рис. 4).

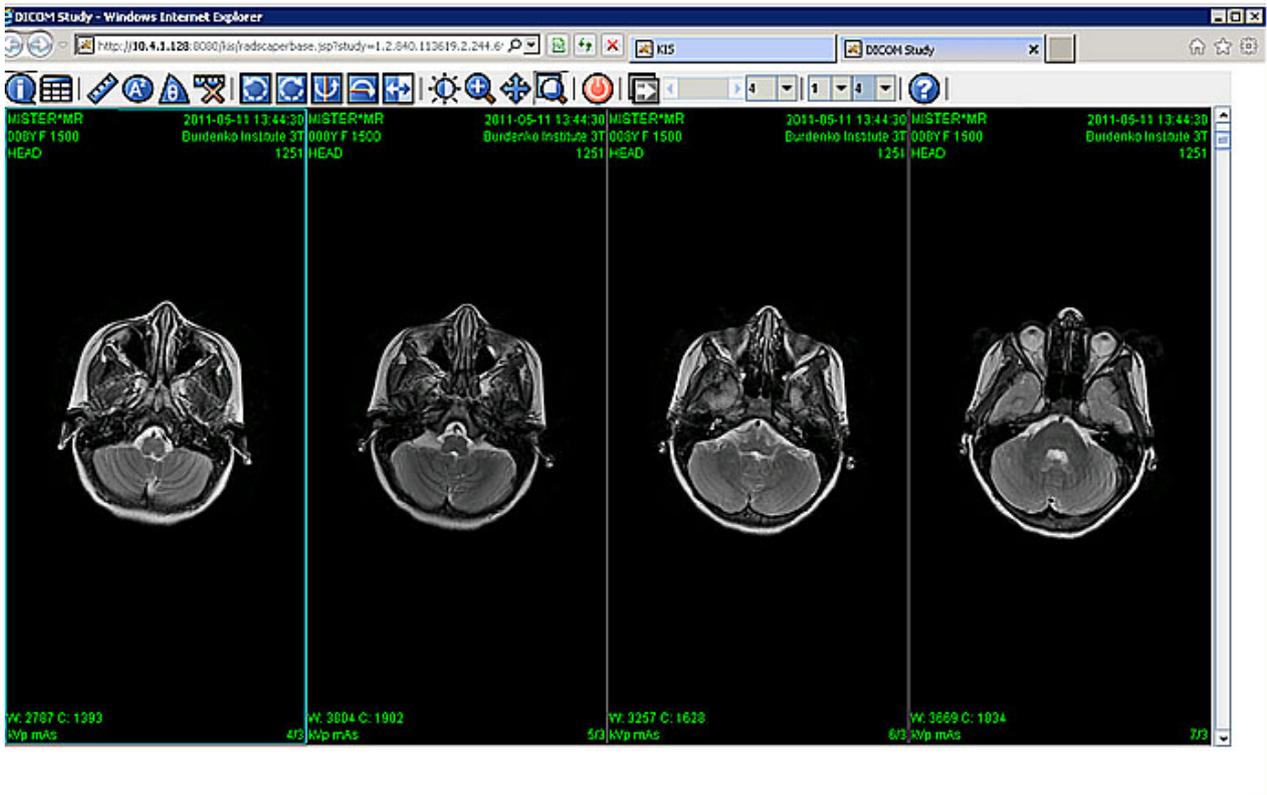


Рис 4. Окно визуализации изображений браузера Radscaper.

4. Распределённая программная среда интеграции PACS-серверов

Программное обеспечение, реализующее описанный выше сценарий, решает следующие основные задачи:

- поиск исследований на множестве PACS-серверов и доставка данных исследований;
- подключение программ анализа изображений;
- создание и управление инфраструктурой PACS-серверов.

Эти задачи решаются на основе распределённой программной архитектуры, включающей клиентскую и серверную компоненты.

Клиентская часть – это приложение M-PACS, которое реализует пользовательский интерфейс и поддерживает связь между поиском, доставкой изображений и приложениями визуализации. В качестве серверной составляющей используется разработанная в ИПМ РАН система MQ-DAI [9], с помощью которой выполняется поиск, образуется инфраструктура PACS-серверов, осуществляется управление инфраструктурой.

Схема взаимодействия компонентов между собой и с внешними компонентами, представлена на рис. 5.

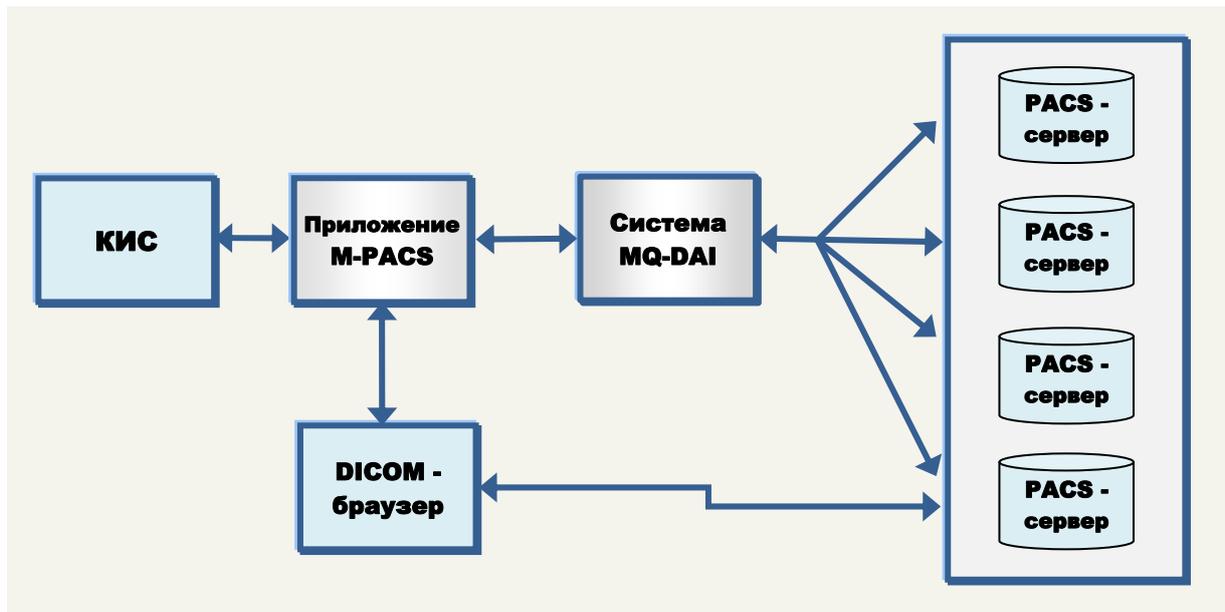


Рис. 5. Взаимодействие компонентов комплекса интеграции.

Компоненты, показанные на этой схеме, распределены по компьютерам следующим образом.

Так как М-PACS является Web-приложением, минимально на рабочем месте пользователя должны быть установлены Web-браузер и Java-машина. Само приложение устанавливается в контейнере Apache Tomcat на каком-либо Web-сервере.

Если в качестве DICOM-браузера используется программа ClearCanvas Workstation, являющаяся приложением Windows, она должна быть установлена на пользовательском компьютере и зарегистрирована на всех PACS-серверах, на которых осуществляется поиск изображений. Браузер Radscafer реализован в форме Java-апплета, который подключён к одной из страниц JSP (Java Server Pages) интерфейса М-PACS и вызывается после выбора исследований для визуализации. В этом случае дополнительных установок на пользовательском компьютере не требуется.

Второй компонентой программной архитектуры является система массовой интеграции MQ-DAI. Она устанавливается на отдельном сервере и обслуживает всех пользователей М-PACS. Система MQ-DAI выполняет следующие функции:

- принимает запросы на поиск исследований от клиентов (в данном случае от установок М-PACS);
- обрабатывает их, получая данные от множества подключенных PACS-серверов и формируя общий результат запроса;
- передаёт (асинхронно, по отдельному запросу) результат пользовательскому клиенту.

В дополнение к этой основной функциональности, система MQ-DAI поддерживает административные функции управления распределённым

множеством PACS-серверов. Функции управления позволяют включать отдельные PACS-серверы в инфраструктуру или исключать их без остановки её функционирования в целом. Функции управления возлагаются на администраторов PACS-серверов, которые могут дистанционно выполнять запросы управления с помощью специальных интерфейсов. Как поисковые, так и запросы управления реализованы на основе стандартов Web- и грид-служб.

Третья компонента распределённой архитектуры – множество PACS-серверов. Включение PACS-сервера в инфраструктуру не требует установки на нём дополнительного программного обеспечения или какого-либо его конфигурирования. При выполнении массового поискового запроса система MQ-DAI в конечном счёте преобразует его в набор SQL-запросов к каждому подключённому PACS-серверу. Полученные от них частичные результаты объединяются и делаются доступными выдавшему запрос приложению M-PACS. Взаимодействие с СУБД PACS-серверов осуществляется по стандартному дистанционному протоколу JDBC (Java Database Connectivity).

5. Реализация основных механизмов

5.1. Поиск исследований

Первым шагом пользовательского сценария является поиск исследований по набору идентифицирующих параметров. При использовании приложения M-PACS в составе КИС часть параметров определена в контексте этой системы и передаётся M-PACS при вызове Web-страницы интерфейса:

```
http://10.4.1.128:8080/kis/index.jsp?patientsName=%D0%B8%D0%B2%D0%B0%D0%BD%D0%BE%D0%B2&patientsSex=M&studyDateFrom=&studyTimeFrom=&modality=&patientsBirthDate=&studyDateTo=&studyTimeTo=&mode=0&user=
```

Если запрос на поиск автоматически формируется КИС, интерфейс M-PACS позволяет ввести дополнительные параметры, например дату исследования, а если M-PACS используется независимо, то все параметры вводятся вручную.

Поиск исследований производится по множеству PACS-серверов, причём предполагается, что базы данных дескриптивной информации всех PACS-серверов соответствуют реляционной модели и поддерживают язык запросов SQL-92.

Для поиска исследований M-PACS строит массовый поисковый запрос в формате, который определяется системой MQ-DAI. Массовые запросы, поддерживаемые системой MQ-DAI, представляют собой расширение стандартного оператора SELECT SQL-92 и отличаются от него в двух аспектах. Во-первых, в массовом запросе элементы данных (имена таблиц и столбцов) адресуются в терминах глобальной схемы [3], которая унифицирует представления семантически эквивалентных данных в разных базах.

Особенностью способа конструирования глобальной схемы в системе

MQ-DAI является то, что элемент глобальной схемы представляет собой объединение эквивалентных данных из всех интегрированных в инфраструктуру баз. Благодаря этому, запрос, сформулированный в терминах глобальной схемы, позволяет производить поиск по всему множеству данных инфраструктуры независимо от того, в какой базе эти данные расположены. Кроме того, поисковое пространство можно ограничить, задав поиск не по всей инфраструктуре, а по некоторому подмножеству содержащихся в ней баз. Это достигается за счёт модифицированного способа адресации элементов данных в запросе SELECT: вместо конструкции *Имя_Таблицы.Имя_Столбца* используется *Имя_Группы.Имя_Таблицы.Имя_Столбца*. Под Группой понимается программный объект, содержащий список баз данных, по которым производится поиск. Группа может содержать как все базы данных, подключённые к серверу MQ-DAI, так и какое-то их подмножество. Детальное описание языка массовых запросов дано в работе [10]. Здесь мы ограничимся примером запроса, который соответствует поиску, показанному на рис. 1, 2:

SELECT

*IServers.serverIp, s.AeTitle, s.Port, s.StudyInstanceUid, s.StudyId, s.PatientId,
IServers.serverName, s.PatientsName, s.PatientsBirthDate,
s.PatientsSex, s.StudyDate, s.StudyDescription, s.SeriesDescription,
s.StudyTime, s.Modality,
heuristic_cmp(s.PatientsName, 'JUKOVA') AS heuristic*

FROM IServers.Study_Series_ServerPartition AS s

WHERE s.PatientsSex='M' AND heuristic > 0.7 ORDER BY heuristic DESC;

Формируемый приложением M-PACS запрос рассчитан на глобальную схему, которая состоит из одной таблицы *Study_Series_ServerPartition* с полями:

- *AeTitle, Port, StudyInstanceUid, StudyId, PatientId* – используются для уникальной идентификации исследования и получения изображений.
- *PatientsName, PatientsBirthDate, PatientsSex, StudyDate, StudyDescription, SeriesDescription, StudyTime, Modality* – значения этих полей отображаются пользователю как результат поискового запроса.

В результатах запроса содержатся также характеристики баз данных PACS-серверов: *serverIp* – сетевой адрес, который используется для получения файлов изображений, *serverName* – название медицинского учреждения-владельца сервера или самого сервера, которое служит пользователю как дополнительный описатель исследования.

Поясним две особенности приведённого выше запроса: реализацию эвристического поиска фамилий и способ отображения глобальной схемы в схемы баз данных реальных PACS-серверов.

Перед выполнением запроса M-PACS приводит фамилию, заданную в качестве параметра поиска, к верхнему регистру, и, если её ввели русскими

буквами, транслитерирует её с помощью набора правил, заменяющих одну или несколько русских букв одной или несколькими английскими. Список правил приведен в приложении 1. Правила просматриваются последовательно, применяется первое подходящее.

Функция эвристического сравнения `heuristic_cmp`, которая используется в поисковом запросе, определяет близость («похожесть») пары строк, записанных прописными латинскими буквами, возвращая вещественное число (показатель близости) тем большее, чем ближе друг к другу находятся строки (согласно алгоритму в приложении 2). Такой поиск может найти исследование, даже если в записи фамилии пациента были допущены ошибки.

Условие в поисковом запросе `heuristic > 0.7` отсеивает строки, которые алгоритм счёл мало похожими, после чего результат сортируется по мере убывания показателя близости строк. Число 0.7 – параметр поиска, определяющий порог для отсеивания результатов. Его значение подобрано экспериментально так, чтобы в результате поиска не попадали в слишком большом числе фамилии, мало похожие на искомую.

Стандарт DICOM определяет перечисленные в запросе поля как обязательные атрибуты исследований, которые должны быть представлены в базах данных PACS-серверов. Однако имена, состав таблиц и форматы представления данных в реальных PACS-серверах могут отличаться от определённых в глобальной схеме. В этом случае требуется определить отображение глобальной схемы на схему PACS-сервера. Отображение задаётся запросами `SELECT`, которые порождают перечисленные выше таблицы глобальной схемы (аналогично пользовательским представлениям `VIEW`). Важно, что отображения определяются в системе MQ-DAI, так что в базы данных PACS-серверов никаких изменений не вносится. Ввод определений производится администратором PACS-сервера дистанционно с помощью интерфейса управления инфраструктурой системы MQ-DAI.

В качестве примера приведём отображение, которое используется для подключения сервера `ClearCanvas Image Server`. В этом варианте требуемые данные содержатся в трех таблицах PACS-сервера:

- `Study` (`st` в приведённом ниже запросе) содержит данные пациента и исследования.
- `ServerPartition` (`sp`) содержит ссылочные данные для доступа к изображениям.
- `Series` (`sr`) содержит описание составных частей исследования – серий.

Отображение глобальной таблицы на таблицы PACS-сервера производится запросом:

SELECT

*st.StudyInstanceUid, st.StudyDate, st.StudyId,
st.StudyDescription, st.PatientsName,
st.PatientId, st.StudyTime, st.PatientsBirthDate, st.PatientsSex,*

```

    sr.SeriesDescription, sr.Modality, sp.AeTitle, sp.Port
FROM Study st
LEFT JOIN ServerPartition sp ON st.ServerPartitionGUID = sp.GUID
LEFT JOIN Series sr ON sr.ServerPartitionGUID = sp.GUID
AND sr.StudyGUID = st.GUID

```

5.2. Выполнение запросов поиска исследований

Система MQ-DAI, выполняющая массовые запросы, ориентирована на интеграцию баз данных реляционного типа, и это сужает применимость комплекса в целом. Хотя в большинстве современных PACS-серверов используются реляционные базы данных, стандарт DICOM не предписывает этого. Информационная модель DICOM – Query/Retrieve Information Model допускает и иные способы хранения данных, а также частные протоколы их поиска и извлечения.

В перспективе ограничение лишь реляционными PACS-серверами может быть снято. Система MQ-DAI в части доступа к базам данных и выполнения распределённых запросов опирается на комплекс OGSA-DAI/DQP [11]. Этот комплекс способен обеспечивать доступ к гетерогенным базам, которые в рамках реляционной модели различаются типами СУБД. Достигается это за счёт использования сменных драйверов в соответствии с типом СУБД. Благодаря этому, в рамках программной архитектуры OGSA-DAI/DQP вопрос выполнения поисковых запросов на основе стандарта DICOM может быть решён путём разработки специального типа драйвера, который преобразует реляционные запросы в запросы DICOM и реализует взаимодействие со стандартными службами PACS-серверов.

В то же время, использование таких средств общего назначения как система MQ-DAI имеет ряд преимуществ при решении задач интеграции данных.

- MQ-DAI даёт возможность интегрировать базы с содержательно разными данными. Информационная инфраструктура, получаемая в результате интеграции, может включать не только PACS-серверы, но также и другие базы данных. Это позволяет создавать приложения, выполняющие многоаспектную обработку данных, получаемых из разнообразных источников.

- В MQ-DAI реализованы механизмы, которые оптимизируют выполнение массовых запросов большой вычислительной сложности, причём это относится не только к объединению данных (операция UNION), получаемых из разных баз, но и к операции связывания (JOIN).

- Одной из наиболее важных проблем в системе интеграции является защита данных. Для MQ-DAI разрабатываются методы контроля доступа, основанные на современных подходах.

- В составе MQ-DAI имеются средства управления инфраструктурой распределённых баз данных, которые позволяют дистанционно подключать/отключать отдельные базы без нарушения работы приложений.

5.3. Подключение программ визуализации изображений и доставка файлов

Программы визуализации изображений – ClearCanvas Workstation и Radscaper – получают результаты поиска исследований, доставляют файлы изображений и визуализируют их.

DICOM-браузер ClearCanvas Workstation представляет собой интерактивное Windows-приложение, реализованное на платформе Microsoft .NET и языке C#. В силу того, что архитектура ClearCanvas Workstation ориентирована на интерактивный режим, её включение как составной части в другие программы вызывает определённые трудности (хотя реализация основана на архитектуре плагинов и допускает подключение к ней внешних модулей). Для вызова ClearCanvas Workstation из приложения M-PACS разработан промежуточный модуль, который инициализирует исполнительную среду браузера и запускает его, сразу визуализируя окно с найденными исследованиями.

При запуске ClearCanvas Workstation получает от M-PACS атрибуты исследований, необходимые для доставки файлов изображений, доставка выполняется собственными механизмами браузера. При передаче файлов используется потоковый протокол (streaming), что ускоряет эту операцию. Однако протокол является нестандартным, что ограничивает круг PACS-серверов, для которых может применяться ClearCanvas Workstation. Кроме того, в конфигурации этого браузера должны быть объявлены все доступные ему PACS-серверы. В целом, использование ClearCanvas Workstation в качестве программы визуализации изображений можно рекомендовать для инфраструктур с небольшим количеством PACS-серверов, которые поддерживают потоковую передачу файлов.

Второй вариант программы визуализации – Radscaper может использоваться при работе с любыми PACS-серверами. Radscaper представляет собой Java-апплет, который визуализирует заданные в формате DICOM файлы изображений, внутри страницы Web-браузера. Визуализируемые файлы должны размещаться на том же сервере, на котором выполняется Radscaper, либо Radscaper может их доставить с некоторого Web-сервера, используя для этого протокол HTTP(S). Однако доставку файлов изображений непосредственно с PACS-серверов Radscaper не поддерживает, и эту часть работы берёт на себя приложение M-PACS, используя для этого службу WADO.

Удалённый доступ WADO (Web Access to DICOM Persistent Objects) [12] является одной из составляющих стандарта DICOM. Согласно стандарту для получения файла в запросе WADO требуется указать набор атрибутов, который определяет DICOM-изображение, содержащееся в некотором исследовании. К таким атрибутам относятся идентификаторы серий в исследовании и идентификаторы изображений в сериях. Запрос на передачу файла к службе

WADO реализуется операцией GET протокола HTTP(S):

http://host:1000/wado/sae?requestType=WADO&contentType=application/dicom&studyUID=<studyUID>&seriesUID=<seriesUID>&objectUID=<objectUID>&transferSyntax=1.2.840.10008.1.2.1

Помимо атрибута StudyUID, полученного на стадии поиска исследований, необходимы также атрибуты SeriesUID и ObjectUID. Для получения этих двух недостающих параметров используется стандартная служба Query/Retrieve PACS-сервера, обращение к которой строится с помощью открытой библиотеки dcm4che [13]. Приложение M-PACS обращается к службе Query/Retrieve определённого PACS-сервера, выполняя операцию C-FIND, результаты которой содержат идентификаторы серий и изображений.

Таким образом, образуется полный набор параметров для выполнения запроса на передачу файлов, которые доставляются на сервер, где размещается M-PACS. Далее M-PACS динамически конфигурирует Radscapec, формируя список локальных адресов файлов, и запускает его.

6. Работы по интеграции PACS-серверов

По-видимому, впервые цель создания инфраструктуры из распределённых PACS-серверов была поставлена в проекте MammoGrid [14]. Предложенный в нём способ основан на применении технологий грида и характеризуется двухуровневой организацией инфраструктуры. Глобальный уровень состоит из множества серверов (грид-серверов), каждый из которых играет роль шлюза для доступа к информации нескольких PACS-серверов, находящихся с ним в одной локальной сетевой среде. Непосредственный доступ к PACS-серверам не предполагается, вместо этого информация локальных PACS-серверов дублируется на их грид-сервере в форме, позволяющей обеспечить поиск исследований и доступ к файлам изображений. Для этого на шлюзовых серверах поддерживается файловая система, в которую помещаются файлы изображений с PACS-серверов, и создаётся база данных, которая содержит дескриптивную информацию – атрибуты исследований и адреса соответствующих файлов.

Прикладной интерфейс проекта MammoGrid содержит операции двух типов. Первый тип операций предназначен для загрузки файлов исследования, полученного с PACS-сервера, на локальный грид-сервер. При этом из файлов выделяются и помещаются в базу данных дескриптивные данные (также может быть добавлена аннотирующая информация, которая отсутствует в стандарте DICOM). Второй тип операций - поиск на локальном или удалённом грид-сервере. Поиск ведётся по базе данных некоторого грид-сервера, в качестве результата выдаются атрибуты исследований и адреса файлов (сетевые в случае удалённого грид-сервера), которые могут быть получены клиентом с помощью отдельной операций.

Организация инфраструктуры с использованием шлюзовых грид-серверов применяется и в проекте mantisGRID [15]. Предполагается, что файлы изображений хранятся в файловой системе грид-серверов, а атрибуты – в реляционной базе данных. В отличие от MammoGrid интеграция дескриптивных данных производится с помощью комплекса OGSA-DAI/DQP. OGSA-DAI/DQP поддерживает распределённые запросы SELECT, в которых выполняются операции над таблицами нескольких распределённых баз данных. Тем самым появляется возможность производить поиск по множеству содержащихся в инфраструктуре грид-серверов. Отметим, однако, что комплекс OGSA-DAI/DQP не обладает рядом существенных возможностей системы MQ-DAI, среди которых: интеграция данных из гетерогенных баз, определение пространства поиска, формирование и управление инфраструктурой.

В проекте MEDICUS (Medical Imaging and Computing for Unified Information Sharing) [16] сделана попытка определить унифицированный способ создания инфраструктур, подобных описанным выше, на основе стандартов Web-служб и технологий грида: GridFTP, Grid Replication Service. В качестве средства управления метакаталогом атрибутов исследований и ссылок на файлы изображений используется комплекс OGSA-DAI/DQP. Совокупность этих средств обеспечивает глобальный поиск исследований, надёжную передачу, репликацию и кеширование файлов, защиту информации.

Подход, который направлен на непосредственную интеграцию PACS-серверов и не требует создания дополнительных хранилищ, реализован в программном средстве VirtualPACS [17]. VirtualPACS представляет собой набор служб, которые реализуют определённые стандартом DICOM операции управления информационными объектами (поиска, записи и получения файлов). По сути дела VirtualPACS моделирует виртуальный PACS-сервер, однако в качестве среды хранения информационных объектов используется не один, а множество PACS-серверов.

VirtualPACS выполняет роль посредника между клиентскими приложениями и множеством PACS-серверов. Клиенты взаимодействуют с презентационным слоем VirtualPACS, посылая ему запросы в формате и по коммуникационному протоколу стандарта DICOM.

Поступающие в VirtualPACS запросы преобразуются в обобщённую форму, основанную на расширенной информационной модели (NCIA DICOM Data Model), и направляются на один или несколько источников данных. Каждый источник – PACS-сервер – снабжается оболочкой, которая реализована в форме грид-службы и выполняет трансформацию внешних запросов в формат стоящего за ней PACS-сервера. Это даёт возможность включать в инфраструктуру как источники, поддерживающие стандарт DICOM, так и любые другие. Доставка результатов заключается в их преобразовании в набор атрибутов DICOM и передачи их клиентскому приложению, выдавшему запрос. Результаты поисковых запросов возвращаются по протоколу SOAP, а для

пересылки файлов изображений используется протокол GridFTP.

Благодаря тому, что клиентские интерфейсы VirtualPACS аналогичны интерфейсам обычных PACS-серверов, существующие приложения, такие как браузеры DICOM-изображений, могут без изменения использоваться для работы с распределённым множеством PACS-серверов.

В нашей работе основное внимание уделено применению новых, развитых в системе MQ-DAI методов для интеграции данных исследований и управления инфраструктурой PACS-серверов. Вопрос интеграции данных рассматривается с точки зрения разработки приложений, поэтому, в отличие от VirtualPACS, в качестве языка запросов для получения данных используется SQL. Кроме того, в клиенте M-PACS решены вопросы подключения приложений визуализации изображений.

Более подробные сведения об использовании грид-технологий в здравоохранении можно найти в книге [18].

7. Заключение

Разработка комплекса доступа к распределённым архивам медицинских изображений выполнена в сотрудничестве с НИИ нейрохирургии имени академика Н.Н.Бурденко РАМН. С помощью имеющейся версии комплекса планируется подключение нескольких PACS-серверов к системе электронной карты e-Med, которая является основным средством информационной поддержки лечебно-диагностического процесса в институте. Возможность поиска и загрузки снимков непосредственно в e-Med будет доступна, в первую очередь, сотрудникам отделения нейрорентгенологии. Общее число интегрированных серверов будет небольшим (4, включая тестовые), так что говорить о создании инфраструктуры трудно, но полученный практический опыт позволит сделать первые выводы об использовании разработанного средства в реальных условиях.

Использование в разработке универсального средства – системы MQ-DAI представляется оправданным. Помимо тех преимуществ, которые даёт опора на современные стандарты, методы интеграции данных, реализованные в MQ-DAI, обеспечивают функционирование комплекса при меняющемся составе архивов, что способствует постепенному наращиванию инфраструктуры. MQ-DAI также открывает возможность включения в инфраструктуру баз данных, отличных от PACS-серверов, и содержит программные интерфейсы для разработки использующих их приложений.

Подключение серверов возлагается на их администраторов и не требует особых навыков. При подключении в базе данных PACS-сервера создаётся учётная запись, посредством которой система MQ-DAI получает удалённый доступ по протоколу JDBC к нескольким таблицам. Кроме того, определяются отображения таблиц глобальной схемы на таблицы PACS-сервера. Это также не вызывает трудностей, так как отображения задаются в форме обычных

пользовательских представлений, которые традиционно используются на практике.

Часто применение универсальных средств приводит к плохим техническим характеристикам в конкретных приложениях. Для разработанного комплекса производительность MQ-DAI является достаточной: время выполнения сценария поиска-визуализации составляет примерно 5 секунд и тратится в основном на загрузку DICOM-браузеров и передачу файлов. Как показали проведенные на модельной инфраструктуре эксперименты, время выполнения поисковых запросов на 100 базах данных меньше 1 секунды.

Система MQ-DAI существенно упростила разработку комплекса, так что наиболее сложным оказалось подключение программ визуализации изображений – DICOM-браузеров. Обычно они разрабатываются как независимые приложения, и реализация их вызова составляет отдельную задачу. Положительным моментом является то, что число применяемых на практике браузеров невелико.

Перспективы развития комплекса мы связываем с более полным использованием стандартных служб DICOM, с разработкой средств контроля доступа, адаптированных к условиям медицинских учреждений, и с созданием графических интерфейсов управления инфраструктурой PACS-серверов.

8. Литература

- [1]. Web Services Data Access and Integration - The Core WS-DAI Specification / M. Antonioletti [etc.]. URL: <http://ogf.org/documents/GFD.74.pdf>
- [2]. OGSA-DAI, Open Grid Service Architecture – Data Access and Integration. URL: www.ogsadai.org.uk/
- [3]. Lenzerini M. Data Integration: A Theoretical Perspective // PODS 2002, pp. 233–246.
URL: <http://www.dis.uniroma1.it/~lenzerin/homepagine/talks/TutorialPODS02.pdf>
- [4]. Digital Imaging and Communications in Medicine. URL: <http://medical.nema.org/>
- [5]. Shifrin M., Kalinina E., Kalinin E. EPR Project for N.N.Burdenko Neurosurgical Institute: Goals, Technology, Results, Future Trends // British journal on healthcare computing and information management, 2003, v.20, #7.
- [6]. Shifrin M., Kalinina E., Kalinin E. Sustainability view on EPR system of N.N.Burdenko Neurosurgical Institute // Proceedings of the 12th World Congress on Health (Medical) Informatics, Part 2, IOS Press, 2007.
- [7]. ClearCanvas. URL: <http://www.clearcanvas.ca/>
- [8]. Radscaper - An online Java Web-based DICOM viewer. URL: <http://www.divinev.com/radscaper/>
- [9]. Коваленко В.Н., Коваленко Е.И., Куликов А.Ю. Система массовой интеграции баз данных: функциональные возможности и способ реализации // Распределенные вычисления и Грид-технологии в науке и образовании: Труды 5-й международной конференции «Распределенные вычисления и Грид-технологии в науке и образовании», Дубна, 2012, с. 337-342.
- [10]. Коваленко В.Н., Куликов А.Ю. Интеграция данных и язык запросов в масштабных информационных инфраструктурах // Программные продукты и системы. № 3, 2012, с. 124-130.
- [11]. The design and implementation of OGSA-DQP: A service-based distributed query processor / Lynden S. [etc.] // Future Generation Computer Systems, Volume 25, Issue 3, March 2009, pp. 224-236. URL: <http://www.ogsadai.org.uk/>
- [12]. DICOM-P18 (2004) Digital Imaging and Communications in Medicine (DICOM), Part 18: Web Access to DICOM Persistent Objects (WADO). National Electrical Manufacturers Association.
- [13]. dcm4che - Open Source Clinical Image and Object Management. URL: <http://www.dcm4che.org/>
- [14]. MammoGrid: a service oriented architecture based medical grid application / Amendolia SR. [etc.] // Lecture Notes in Computer Science, GCC 2004, LNCS 3251. Berlin: Springer, 2004, pp. 939–942.
- [15]. mantisGRID: a grid platform for DICOM medical images management in Colombia and Latin America / Garcia Ruiz M. [etc.] // J Digit Imaging. 2011 Apr 24(2):271-283.

- [16]. Grid Interface Service for Clinical and Research PACS: A Globus Toolkit Web Service for Medical Data Grids / Erberich, S.G. [etc.] // International Journal of Computer Assistant Radiology and Surgery, 1:87. 100-102. 2006.
- [17]. VirtualPACS—a federating gateway to access remote image data resources over the grid / Sharma A. [etc.] // J. Digit. Imaging 22(1):1–10, 2009.
- [18]. Healthgrid Researches, Innovation and Business Case (Proceedings of HealthGrid 2009) / Ed. T. Solomondies [etc.] // IOS Press, 2009.

Приложение 1. Правила транслитерации

Номер правила		1	2	3	4	5	6	7	8	9	10	11	12
Правило замены	Кириллица	ЬЕ	ЫЙ	А	Б	В	Г	Д	Е	Ё	Ж	З	И
	Латиница	IE	IY	A	B	V	G	D	E	E	ZH	Z	I

Номер правила		13	14	15	16	17	18	19	20	21	22	23	24
Правило замены	Кириллица	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х
	Латиница	K	L	M	N	O	P	R	S	T	U	F	KH

Номер правила		25	26	27	28	29	30	31	32	33	34		
Правило замены	Кириллица	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я		
	Латиница	TS	CH	SH	SCH	‘	Y		E	JU	YA		

Приложение 2. Алгоритм эвристического сравнения

Алгоритм эвристического сравнения следующий:

1. Каждая строка разбивается на пересекающиеся подстроки длины K (параметр алгоритма, по умолчанию 3). Например, ABCDE – [ABC, BCD, CDE].
2. Иницируются начальные переменные s_1, s_2 – массивы подстрок, $i_1=0, i_2=0, d=0, f=0$.
3. Сравниваются подстроки $s_1[i_1]$ и $s_2[i_2+d]$. Если подстроки эквивалентны, то $f:=f+1.0/(d+1)$, $i_2:=i_2+d$ и переходим на шаг 8.
4. Сравниваются подстроки $s_1[i_1+d]$ и $s_2[i_2]$. Если подстроки эквивалентны, то $f:=f+1.0/(d+1)$, $i_1:=i_1+d$ и переходим на шаг 8.
5. Увеличиваем значение переменной $d:=d+1$.
6. Если $i_1+d \geq s_1.length$ или $i_2+d \geq s_2.length$ – то на шаг 8.
7. Если значение $d < M$ (M - параметр алгоритма, определяющий максимальную глубину поиска совпадающих подстрок, по умолчанию 3), то переходим на шаг 3.
8. $i_1:=i_1+1, i_2:=i_2+1, d:=0$. Если $i_1 < s_1.length$ и $i_2 < s_2.length$, то переходим на шаг 3.
9. Конец алгоритма. Результат близости строк - f .

Пример работы алгоритма для строк “ABCDE” и “BCDE”.
 $s_1=[ABC, BCD, CDE], s_2=[BCD, CDE]$.

Шар	i1	i2	D	f
2	0	0	0	0
3(ABC – BCD)	0	0	0	0
4(ABC – BCD)	0	0	0	0
5,6,7	0	0	1	0
3(ABC – CDE)	0	0	1	0
4(BCD – BCD)	1	0	1	0.5
8	2	1	0	0.5
3(CDE – CDE)	2	1	0	1.5
8	3	2	8	1.5
9	Результат близости строк 1.5			

Оглавление

1. Введение	3
2. Состав комплекса и назначение компонентов	5
3. Сценарий работы с приложением M-PACS.....	5
4. Распределённая программная среда интеграции PACS-серверов.....	9
5. Реализация основных механизмов	11
5.1. Поиск исследований.....	11
5.2. Выполнение запросов поиска исследований	14
5.3. Подключение программ визуализации изображений и доставка файлов .	15
6. Работы по интеграции PACS-серверов	16
7. Заключение.....	18
8. Литература	20
Приложение 1. Правила транслитерации	22
Приложение 2. Алгоритм эвристического сравнения	22