

АРХИТЕКТУРА СИСТЕМЫ ХРАНЕНИЯ И ОБРАБОТКИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

В.А. Сапрыкин

Нижегородский госуниверситет им. Н.И. Лобачевского

Введение

В работе ставится задача проектирования системы хранения и обработки медицинских изображений (томограмм, рентгенограмм, снимков УЗИ и др.) и предлагается вариант ее решения. Подобные системы помогают в решении множества проблем, среди которых следующие.

- Обеспечение врачу централизованного доступа к медицинским изображениям пациента, полученным, возможно, в разных местах и в различное время. Это полезно для оценки динамики течения заболевания.
- Обеспечение исследователю и обучающемуся централизованного доступа к большой совокупности обезличенных медицинских изображений.
- Поиск изображений по различным параметрам (пациент, диагноз, часть тела, дата снимка и др.), в том числе, поиск изображений, похожих на данное изображение во всей системе хранения.
- Сбор обобщенных данных для обучения алгоритмов предварительной диагностики.
- Предоставление врачу возможности комментирования изображений, в том числе, выделения на них областей интереса.
- Обеспечение конфиденциальности медицинских и персональных данных.

Постановка задачи

Для создания системы хранения медицинских изображений, способной решать описанные выше задачи, необходимо, кроме самих изображений, хранить дополнительную информацию по каждому изображению и информацию о взаимосвязи отдельных изображений. Эти данные должны включать как минимум информацию о пациенте, дату снимка, тип патологии или признак ее отсутствия и часть тела на изображении. Имеет смысл построить иерархию частей тела для быстрого поиска всех изображений, содержащих интересующую область. Кроме того, для обеспечения поиска по содержанию необходимо для изображений хранить дескрипторы, ускоряющие поиск.

Ввиду большого объема медицинских изображений (в частности томограмм) требуется способность системы работать в условиях, когда файлы изображений распределены по различным серверам в медицинских учреждениях и специально выделенным серверам. Жестким требованием является поддержка формата DICOM – промышленного стандарта создания, хранения, передачи и визуализации медицинских изображений и документов обследованных пациентов.

Для обеспечения защищенности информации нужно предоставлять доступ к системе только уполномоченным пользователям. Поэтому перед началом работы с хранилищем должно запрашиваться имя пользователя и пароль. Поскольку система распределенная, имеет смысл для обмена данными между узлами пользоваться протоколами с шифрованием. Отдельно стоит вопрос обеспечения конфиденциальности персональных данных пациентов.

Архитектура системы

Будем отталкиваться от классической двухзвенной архитектуры клиент-сервер. Для хранения метаданных о медицинских изображениях воспользуемся реляционной базой данных. Сами медицинские изображения есть файлы, поэтому для их хранения воспользуемся специально созданным для этого инструментом – файловой системой диска. В базе данных в таблице Storage будем хранить адреса серверов, которые непосредственно хранят файлы изображений. Для каждого изображения создадим пару полей: одно будет определять сервер хранения (поле StorageID), а другое – путь к изображению на этом сервере (поле FilePath). Для доступа клиента к файлам на серверах хранения будем использовать FTP сервер (FTPS при необходимости шифрования передаваемых данных).

Для ускорения работы при медленном сетевом соединении операции над изображениями, не требующие передачи всего изображения на клиентский компьютер, необходимо производить на самом сервере. В качестве примера таких операций можно указать выборку изображения отдельного среза томограммы и поиск изображения по образцу в рамках данного сервера хранения. Многие операции (в частности над томограммами) могут потребовать больших вычислительных ресурсов, в том числе расчетов на GPU. Это обстоятельство определяет требования к вычислительной мощности серверов хранения. Ответственным за выполнение операций будет CGI приложение, принимающее параметры операции и возвращающее результат по протоколу HTTP (HTTPS при необходимости шифрования передаваемых данных).

Чтобы при повторном доступе к недавно использовавшемуся изображению не скачивать его заново, создадим на каждом клиентском компьютере кэш изображений. Для хранения персональных данных есть два способа. Хранить их в общей базе в зашифрованном виде с предоставлением ключа шифрования только медицинскому учреждению, которому принадлежат персональные данные. Либо же хранить их в отдельной базе данных, размещенной на одном из компьютеров ЛВС медицинского учреждения. Останемся на втором варианте.

Получившаяся архитектура системы показана на рис. 1.

База данных

Для ряда сущностей предметной области заведем в базе данных отдельные таблицы с необходимыми полями: Patient для хранения информации о пациентах (отношение «один ко многим» с мед. изображениями), Pathology для хранения списка патологий (отношение «многие ко многим» с мед. изображениями), BodyPart для хранения списка частей тела (отношение «один ко многим» с мед. изображениями). Для простоты ограничимся только следующими типами изображений: томограммами, рентгеновскими снимками, снимками УЗИ. Соответствующие таблицы назовем Tomogram, Radiogram и UltraSound. Поскольку за один прием может быть сделано несколько снимков, не будем связывать сущности «Изображение» и «Пациент» непосредственно. Вместо этого добавим сущность «Исследование» (таблица Examination), у которой будет атрибут ExamDate, хранящий дату получения снимка.

Для выделения и комментирования прямоугольных областей (произвольной ориентации) на плоских изображениях и областей в виде прямоугольного параллелепипеда (с ребрами, параллельными осям координат) на объемных изображениях добавим таблицы InterestAreaTom, InterestAreaRad, InterestAreaUS. Они связаны в отношении «многие к одному» с соответствующими таблицами, описывающими медицинские изображения и с таблицей Pathology, разрешая тем самым связь «многие ко многим» между сущностями «Изображение» и «Патология».

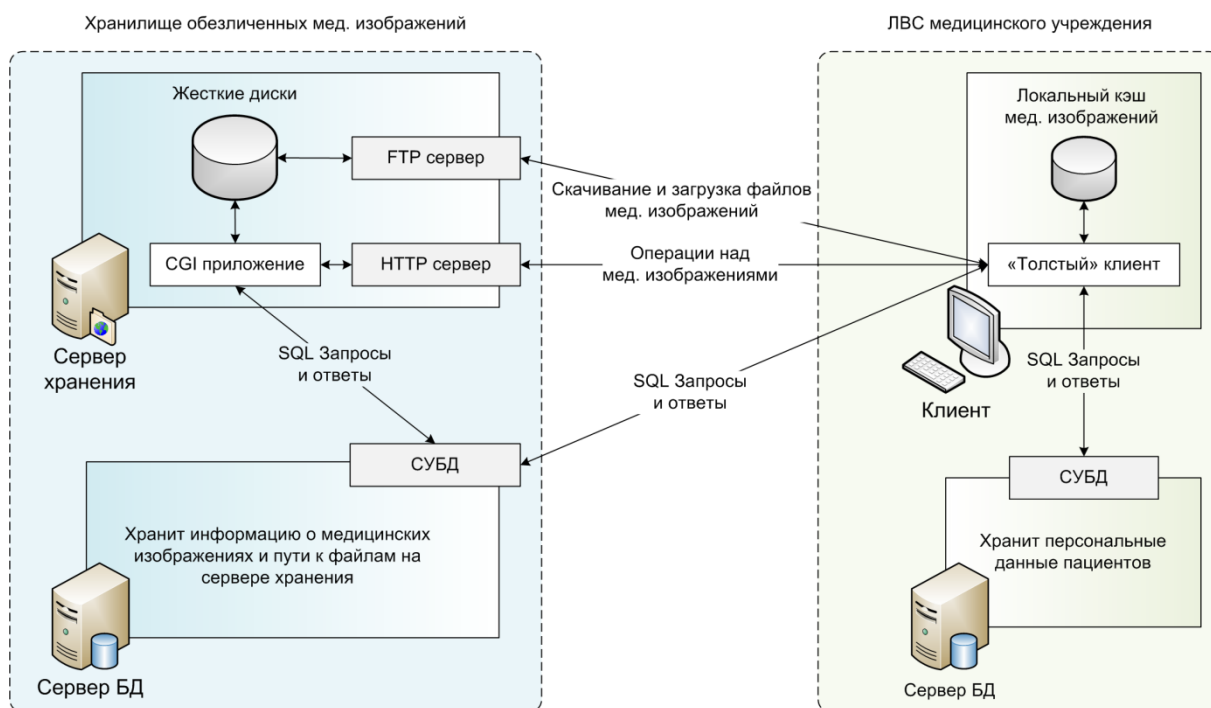


Рис. 1. Архитектура системы хранения медицинских изображений

Чтобы ускорить поиск по содержимому, для каждого изображения (в нашем случае ограничимся рентгеновскими снимками) будем хранить серию его квадратных уменьшенных копий. Пусть самая маленькая из них имеет размер 32×32 пикселя, а каждая следующая вдвое большее число пикселей по стороне квадрата, чем предыдущая. Уменьшенные сохраним на сервере хранения, добавив к имени основного файла расширение «.сору32», «.сору64», «.сору128», «.сору256» и т.д. Поскольку многие СУБД поддерживают векторный тип данных, можно поместить самую маленькую копию непосредственно в поле `Сору32` таблицы `Radiogram`. Таким образом, при поиске большая часть изображений может быть быстро отсеяна без анализа крупных копий, расположенных на сервере хранения.

Иерархия частей тела

При добавлении медицинского изображения в коллекцию необходимо указывать, какая часть тела на нем изображена. Использовать для этого словесное описание неудобно потому, что на снимке могут присутствовать несколько областей, тогда как целью снимка была лишь одна из них, указанная в словесном описании. В результате остальные области оказываются недоступны для поиска по названию части тела.

Чтобы справиться с этим, формализуем понятие части тела. Возьмем изображение тела человека с нанесенной на него равномерной сеткой. Получившиеся квадраты пронумеруем с помощью пары индексов (x, y) . Множество координат квадратов, пересекающихся с фигурой человека, примем за универс U . При добавлении каждого нового изображения попросим пользователя, выделяя мышкой квадраты, указать область на теле человека, которая изображена на снимке (см. рис. 3). Получим множество A_i , где i – индекс изображения.

В случае поиска по атрибутам или по содержимому, если известен список координат клеток для образца (A), можно сразу исключить из поиска большую часть изображений, для которых выполнено условие $(A \cap A_i = \emptyset)$. Для оставшихся изображений степень близости изображенной части тела с частью тела на образце можно оценить величиной $|A \cap A_i|/|A|$.

Для хранения списка координат квадратов, которые покрывает каждое конкретное изображение, создадим таблицы BodyPartCoordTom, BodyPartCoordRad, BodyPartCoordUS, связанные в отношении «многие к одному» с соответствующими таблицами, описывающими медицинские изображения.

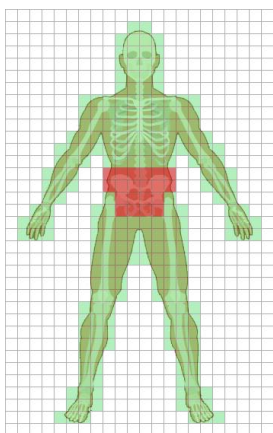


Рис. 2. Разбиение изображения человека на квадраты. Выделена область тазовой кости.

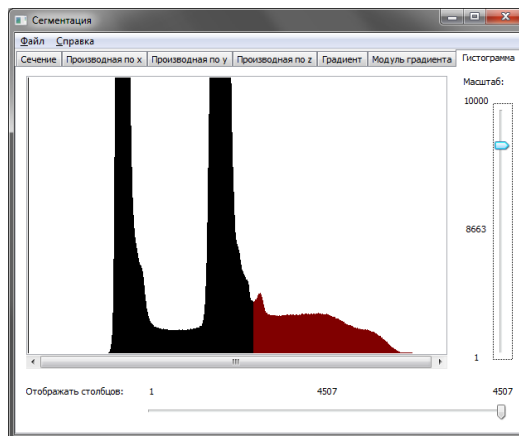


Рис. 3. Результаты сегментации костной ткани на основе гистограммы интенсивности

Предварительная диагностика

Точное указание части тела на изображении и ракурса снимка совместно с информацией о наличии и локализации патологии, а также возможность выполнять сегментацию медицинских изображений позволит собирать обобщенную информацию об органах и тканях. Эта информация, будучи сопоставленной со снимком конкретного пациента, позволит указать врачу на подозрительные области на снимке и, возможно, классифицировать их. На рис. 4 представлена компьютерная томограмма головы и результат сегментации костной ткани на основе гистограммы интенсивности. Первый пик на гистограмме соответствует пустому пространству на снимке, второй – мягким тканям. Более плотная ткань – костная.

Заключение

Мы спроектировали базовую архитектуру системы хранения медицинских изображений, от которой будем отталкиваться в дальнейшем. На момент написания этой работы частично реализован прототип системы. Ведутся работы по реализации подсистем формального указания части тела, выполнения операций над изображениями и реализации поиска изображений по содержимому. Планируется реализовать возможность выполнения предварительной диагностики, в том числе, и на основе сопоставления нового медицинского изображения с уже имеющимися снимками.

За рамками статьи остались возможные усовершенствования архитектуры. Выделим среди них следующие направления работы.

- Реализация сервера приложений как промежуточного звена между СУБД и клиентом. Это позволит увеличить масштабируемость системы.
- Создание гибкой системы управления правами пользователей в зависимости от группы, к которой принадлежит пользователь.
- Рассмотрение возможности слияния таблиц, описывающих двумерные медицинские изображения (Radigram и UltraSound).

- Создание более совершенной системы описания части тела, показанной на изображении, на основе приведения всех хранимых изображений к единой системе координат, связанной с трехмерной моделью человеческого тела.
- Интеграция системы хранения медицинских изображений с рядом других систем. Например, это могут быть системы медицинских карт пациентов, документооборота и системы, поставляемые совместно с медицинским оборудованием для получения снимков.
- При большом количестве снимков целесообразно специализировать серверы хранения с возможностью переноса снимков между ними. Одна часть серверов будет ответственна за долговременное хранение (интерфейс только по протоколу FTP), а другая за оперативное хранение и обработку (полная архитектура сервера хранения).

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России», госконтракт № 02.740.11.0839.

Литература

1. Breton V., Dean K., Solomonides T. Healthgrid The Healthgrid White Paper From Grid to Healthgrid: Proceedings of Healthgrid 2005: P. 249-321, IOS Press, Studies in Health Technology and Infjrmatics, Vol. 112, 2005.
2. Pianykh O.S. Digital Imaging and Communication in Medicine (DICOM). – Berlin, Germany: Springer-Verlag, 2008. – 383 p.
3. Грабер М. Введение в SQL. – М.: ЛОРИ, 1996. – 479 с.
4. Дейт К. Дж. Введение в системы баз данных. – 8-е изд. – М.: «Вильямс», 2006. – 1328 с.
5. Таненбаум Э. Компьютерные сети. 4-е изд. - СПб.: Питер, 2003. – 992 с.