

УДК 519.68

АРХИТЕКТУРА HL-X ПОДДЕРЖКИ ДОКУМЕНТОВ В МЕДИЦИНСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Я. И. Гулиев,

канд. техн. наук, директор Исследовательского центра медицинской информатики

В. Л. Малых,

канд. техн. наук, зав. лабораторией
Институт программных систем РАН

Предлагается подход к построению архитектуры поддержки документов в корпоративных медицинских информационных системах, в основе которого лежит представление о необходимости включать в информационную систему эволюционный процесс концептуализации предметной области. Архитектура характеризуется открытостью, независимостью от средств программной реализации, ориентированностью на существующие и разрабатываемые стандарты, масштабируемостью.

Ключевые слова — медицинская информационная система, документы, стандартизация, концепты.

Введение

Концепция ориентированной на документы архитектуры в медицинских информационных системах (МИС) первоначально представлена нами в работе [1]. Описанная там архитектура была востребована и стала основой целого семейства МИС, автоматизирующих лечебно-диагностический процесс в ряде ведущих лечебных учреждений России.¹ Спустя пять лет мы с удовлетворением отмечаем правильность принятых нами решений для формирования архитектурного базиса медицинского документооборота. Вместе с тем практика применения указанной архитектуры привела к эволюции и расширению архитектуры. Поэтому мы хотим по-новому изложить в статье наше понимание проблемы автоматизации документооборота в МИС и предложить концепцию решения этой проблемы.

Проблемы построения архитектуры современной МИС

Основные проблемы, стоящие перед разработчиками медицинских информационных систем, кратко можно сформулировать следующим образом:

— необходимо вводить в МИС и анализировать все больший объем медицинской информа-

ции для оказания качественной медицинской помощи;

— в перспективе потребуется обеспечить поддержку системой единого информационного пространства, предоставлять полные данные по каждому пациенту независимо от места оказания медицинской помощи;

— необходимо актуализировать и отражать в информационной системе постоянно расширяющиеся понятийную и концептуальную базы предметной области (медицины);

— разработку и развитие МИС приходится вести в условиях недостаточной формализации (концептуализации и стандартизации) предметной области (медицины).

Медицинская информация становится все более сложной и все более важной. Известные исследователи Дипак Калра (Dipak Kalra) и Томас Биел (Thomas Beale) утверждают, что около 15 % ресурсов лечебного медицинского учреждения расходуется на сбор данных. Врачи и медсестры 25 % своего рабочего времени тратят на поиск нужной им информации. По данным US Institute of Medicine Report, около 100 тыс. граждан США ежегодно умирают по причине медицинских ошибок (восьмое место в упорядоченных по частоте причинах смерти). Одна из основных причин ошибок — недостаточная или некачественная информация о пациенте. Объем медицинской литературы, необходимой для качественного лечения, удваивается каждые 10–15 лет. Примерно

¹ Подробнее о системе см. <http://www.interin.ru>.

половина концептов (500 тыс.) из всех выработанных человечеством непосредственно относятся к медицине.

Практически все исследователи в области МИС подчеркивают разнородность, сложность и эволюционную природу медицинских данных. Отсюда проблемы стандартизации медицинской информации, проблемы обмена данными между отдельными МИС. До сих пор не преодолены сложности создания единого информационного пространства медицинских данных даже для граждан развитых западных стран. Попытки жесткой шаблонной формализации ввода и представления медицинской информации врачами отвергаются, как искажающие суть самой информации. С другой стороны, весь мировой опыт свидетельствует о необходимости стандартизации программного обеспечения. Эффективность подобного подхода в промышленности, связанной с разработкой технических изделий массового потребления, давно очевидна. И медицина не должна быть исключением из этого правила.

Многие исследования последних десятилетий связаны с поиском концептуальных решений проблемы стандартизации медицинской информации. Большие надежды возлагаются на повышение уровня абстракции предлагаемых формальных моделей. Процитируем в качестве примера подобного подхода: «Решение заключается в использовании более абстрактных моделей с немногими, но высоко выразительными абстрактными объектами. Модель данных о пациенте следует сделать проще и яснее, для чего необходимо найти унифицированный и корректный уровень абстракции... Например, такие элементы данных, как дату рождения, вес или концентрацию гемоглобина не следовало бы включать в модель в качестве именованных атрибутов. Специфические клинические сущности следовало бы представлять на основе словаря, независимого от модели данных» [2]. (Пер. авт.)

Во всех развитых странах интенсивно ведутся работы по стандартизации медицинской информации. Архитекторы современных российских МИС должны учитывать неизбежность появления в будущем национального российского стандарта.

Направления стандартизации в медицинской информатике

«К настоящему моменту практически все существующие стандарты медицинской информатики носят не обязательный, а рекомендательный характер» [3].

Практически все стандарты медицинской информатики так или иначе связаны с ведением электронной истории болезни [4]. Одни стандар-

ты описывают терминологию, которая должна быть в ней использована, другие — передачу медицинских документов и изображений в электронную историю болезни, третьи — способы организации данных в электронной истории болезни, четвертые — обеспечение доступа медицинских работников и самих пациентов к электронной истории болезни и т. д.

В сущности, разработка стандартов медицинской информатики преследует цель воссоздания универсального языка общения медицинских работников, другими словами — воскрешения латыни на самом современном уровне информационных технологий. В целом эти стандарты нужны для того, чтобы каждая запись электронной истории болезни могла быть одинаково понята представителями различных медицинских школ. При этом компьютеры должны стать как бы переводчиками с привычного естественного медицинского языка на унифицированный электронный язык и обратно. Поэтому неудивительно, что в последние десятилетия наибольшие усилия специалистов по медицинской информатике были сосредоточены в двух основных областях: стандартизации медицинской терминологии и стандартизации передачи записей в электронную историю болезни.

В настоящее время можно выделить два стандарта, разработанных в США и получивших достаточно широкое признание в других странах:

- стандарт электронного обмена текстовыми медицинскими документами Health Level Seven (HL7);

- стандарт электронного обмена изображениями лучевой диагностики Digital Imaging and Communication in Medicine (DICOM).

В последнее время в России также активизировалась работа по стандартизации электронного обмена медицинскими данными. Созданы национальные комитеты: ТК466 «Медицинские технологии» и ТК468 «Информатизация здоровья». Задача этих комитетов — унификация и гармонизация национальных стандартов медицинской информатики со стандартами, принятыми в мире. Появился первый национальный стандарт ГОСТ-Р 52636–2006 «Электронная история болезни. Общие положения». Однако состояние стандартизации в области медицинской информатики в России все еще оставляет желать лучшего.

Понимая, что разработка общероссийских государственных стандартов — это прерогатива, прежде всего, государства, и что она потребует немалых ресурсов и вряд ли окажется по силам отдельным разработчикам МИС, мы приняли стратегическое решение — разработать архитектуру, ориентированную на стандарты. Чтобы ориентироваться на то, «чего еще нет», мы сделали

попытку интуитивно предугадать направление, в котором пойдет стандартизация в медицине. По нашему мнению, стандарт представления медицинских документов должен будет заключать в себе как жестко формализованные концепты предметной области, так и свободно конструируемые по заданным правилам информационные структуры. Можем констатировать, что наши предположения оправдались.

В США разработаны две архитектуры клинических документов — CDA (Clinical Document Architecture) и CCR (Continuity Care Record). CDA, предложенная в рамках развития стандарта HL7, стоит на позиции свободного конструирования документов из кодированных элементов. Эта архитектура определяет модель обмена клиническими документами. В ее основе лежит использование XML, информационной модели RIM и словарей кодов (coded vocabularies). CCR имеет более определенную структуру разделов.

Существует прямая аналогия между CDA и разработанной нами архитектурой HL-X:

- в обеих архитектурах предложен некоторый стандарт представления медицинских (клинических) документов;

- обе архитектуры не пытаются жестко определять структуры клинических документов, отдавая предпочтение свободному декларативному описанию и конструированию документов;

- особо важное место в обеих архитектурах занимает понятие контекста.

Если в МИС любые медицинские данные и их структуры будут стандартизованы и описаны на языке концептов, понятий (терминов) и контекстов их использования в базе знаний системы, перевод системы на общегосударственный стандарт будет означать переход от одного метаописания к другому, в противовес внедрению стандарта в систему, не имеющую явного стандартизованного метаописания. Архитектура современной МИС должна быть готова к работе со стандартами.

Модельные представления документа

В основе архитектуры HL-X лежит понятие документа, его свойств и методов работы с ним. Как всякую сложную систему, документ HL-X можно рассмотреть с различных точек зрения и для каждой из них можно предложить некоторую модель, отражающую различные аспекты документа. Выделим и рассмотрим следующие модели документа.

Понятийная модель представляет документ в виде структуры абстрактных и соответствующих им конкретных понятий. Она дает возможность:

- определить, какие понятия содержатся в документе и какими понятиями «владеет» база данных (БД) или база знаний (БЗ) документов HL-X;

- выполнить перевод понятийной модели на другие языки при наличии словарей;

- добавить семантические связи на понятиях и использовать технологии искусственного интеллекта для анализа свободных документов;

- формализовать структуру и семантику документа;

- использовать модель непосредственно как вариант простой визуализации документа;

- выполнять проверку правописания на основе словарей понятий.

Поддержка понятийной модели документа оказалась для разработчиков самым сложным делом. Проблема в том, что наиболее сильное отставание России от Запада в области архитектуры электронных медицинских документов наблюдается именно по стандартизации отечественной терминологии в медицине. Например, из используемых на Западе кодирующих словарей CPT (Current Procedural Terminology), ICD (International Classification of Diseases), LOINC (Logical Observation Identifier Names and Codes), NDC (National Drug Codes), RxNorm (Standard Names for Clinical Drugs), SNOMED (Systematized Nomenclature of Medicine), пожалуй, только ICD нашел в России широкое применение благодаря переводу на русский язык и национальной локализации этого классификатора. Совершенно понятно, что локализация указанных словарей или разработка их национальных аналогов в России требует немалых затрат и усилий. Быстрых изменений в этой области не ожидается. В рамках архитектуры HL-X мы попытались создать двуязычный тезаурус (на русском и английском языках) медицинской терминологии, используемой в клинических документах. Полученный опыт скорее отрицателен, так как терминологические проблемы пытались решать разработчики программного обеспечения, а не профессионалы в этой области знаний. Задача терминологического обеспечения национальной архитектуры электронных медицинских документов в России все еще ждет своего решения.

Информационная модель представляет документ HL-X в виде структуры из различных информационных объектов (концептов предметной области). Здесь под информационными объектами в широком смысле понимаются некоторые абстрактные связанные элементы модели, аналогами которых в различных языках программирования являются объекты, записи, структуры, массивы и т. п. Наличие у документа структуры позволяет декомпозировать его на элементы и выполнять манипуляции с этими элементами при необходимости. Информационная модель служит основой для:

- конструирования документа из элементарных информационных объектов;

— конструирования документа на основе уже имеющихся документов, с учетом уже накопленного знания о контекстах использования того или иного объекта (понятия), путем заимствования, полного или частичного, уже существующих информационных моделей;

— разбора и обработки данных документа (для структурированного хранения в БД или для экспорта и импорта данных, для усвоения знания и пополнения БЗ непосредственно из экземпляров документов и их информационных моделей).

Модели обработки данных документа позволяют специфицировать обработку документа как единого целого в интересах конкретной информационной системы (ИС). Они определяют, какие именно данные нужно обрабатывать и как это делать. Модель обработки данных документа дополняет информационную модель документа конкретными, зависящими от ИС инструкциями. На основе модели выполняются все основные манипуляции с данными документа.

Модели визуализации позволяют создавать различные визуальные представления данных документа. В основе визуальной модели лежит понятие визуальной компоненты как некоторой абстрактной элементарной формы представления данных. В различных конструкторах интерфейса аналогами визуальным компонентам служат компоненты визуальных палитр (таблицы, поля, панели, всевозможные элементы управления). Основное назначение модели — создать абстрактное, не привязанное к конкретному программному языку или конструктору интерфейса описание представления данных документа в интерфейсе пользователя. Модель визуализации предназначена для:

— визуализации документа в конкретной реализации пользовательского интерфейса со своими документами;

— предоставления, совместно с функциональной моделью, интерфейса для манипуляции данными документа;

— одновременной поддержки нескольких моделей визуализации одного документа (например, для формирования выписок из документов).

Функциональная модель документа определяет возможные манипуляции с данными документа и соответственно поддерживает определенные ограничения на данных (целостность данных). К области компетентности этой модели отнесем:

— транзакционность и инкапсуляцию — обработку документа и всех входящих в него объектов как единого целого в рамках одной транзакции;

— дисциплину коллективной работы над документом (пользовательские блокировки);

— определение возможных манипуляций над документом на основе данных самого документа (учет статуса и т. п.);

— целостность данных (обязательные элементы данных, домены возможных значений, контроль типов и форматов данных и т. п.);

— непосредственное включение пользователей в процесс концептуализации знаний, предоставление самим пользователям возможности формировать домены значений для полей документов;

— проверку правописания в документе;

— поддержку режима автосохранения документа (автотранзакции).

Модель безопасности документа определяет права доступа к элементам данных и права на манипуляцию данными. В качестве структурного элемента документа, на который можно будет установить права доступа, предлагается принять элементарное понятие из понятийной модели. В информационной модели этому элементу доступа соответствует атрибут информационного объекта или сам объект, а в визуальной модели — отдельное поле визуализации или ввода. Более элементарных целостных единиц данных, к которым можно было бы определить доступ, в документе просто нет. На уровне структуры документа предложена самая сильная модель безопасности. Мы сформулировали еще несколько новых требований к модели безопасности:

— управление доступом к отдельным полям документов на основе ролей, что обеспечивает коллективную работу и коллективное визирование документа;

— описание правил «обезличивания» документов для возможного открытого использования документов в консультативных и учебных целях без нарушения закона о защите персональных данных;

— цифровая подпись документа как механизм контроля его целостности;

— шифрование документов.

Необходимо смотреть на документ HL-X как на единство всех этих моделей — их синтез. Документ HL-X — это и сами данные, и организация этих данных в структуру, и знания, заключенные в структуре данных, и правила (ограничения) манипулирования данными, включая права доступа, и визуальное представление данных, и инструкции по обработке этих данных в ИС. Главное преимущество документа HL-X — в высоком уровне абстракции этих моделей, делающих его независимым от конкретной информационной системы, базы данных, технологических средств разработки, каналов связи и средств доставки документа:

— документ HL-X может быть легко доставлен из одной ИС в другую с помощью любых каналов связи и носителей;

— документ HL-X может быть понят и интерпретирован пользователем без использования мо-

дели визуализации на основе своей понятийной модели, благодаря которой он фактически является самодокументируемым;

— модель обработки данных документа в каждой ИС может быть своя, исходя из целей использования этих данных системой (степени интеграции этих данных системой);

— моделей визуализации может быть много, в соответствии с желаемыми представлениями данных.

Определение основных требований к документу как к основному носителю всей медицинской информации

Требования терминологические. Документ HL-X рассматривается нами как формализованная, в первую очередь, с понятийной точки зрения структура данных. Документ HL-X должен основываться на словаре понятий, достаточном для описания (формализации) документов основной предметной области (медицины), а также предметных областей, связанных с основной (управление организацией и персоналом, статистика, материально-техническое снабжение и пр.). Объем словаря понятий потенциально неограничен. Отдельно следует указать на сложность и неоднозначность процесса формирования словаря понятий. Поясним на примере. Пусть у нас имеются в словаре понятия: «группа», «инвалидность», «кровь». И пусть мы стоим перед выбором: следует ли включать в словарь понятия «группа инвалидности» и «группа крови»? Есть две возможности — либо включить эти понятия непосредственно в словарь, либо передать эти понятия через понятийную и информационную модель в виде соответствующих структур:

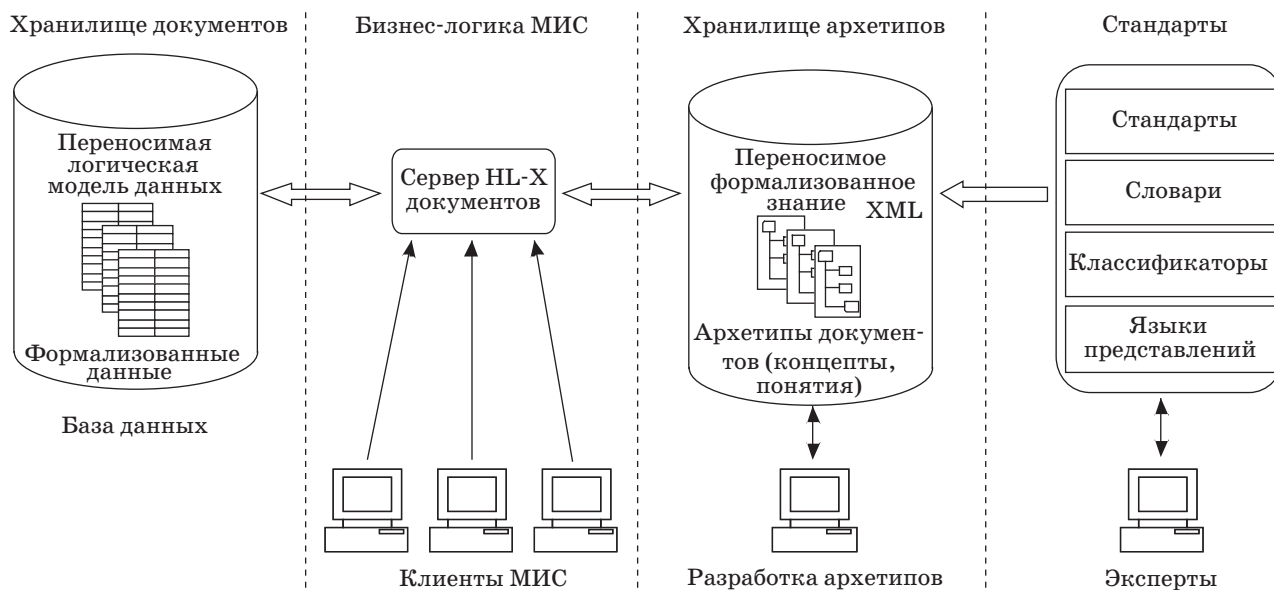
```
<Инвалидность>
  <Группа>Первая</Группа>
</Инвалидность>
<Кровь>
  <Группа>Первая</Группа>
  <Резус>Rh+</Резус>
</Кровь>
```

Теперь для правильного понимания значения высказывания «группа — первая» нам необходим контекст использования термина «группа». Самого термина уже недостаточно. Избавиться совсем от контекстной зависимости и необходимости проводить контекстный анализ нам не удастся даже в полностью формализованной системе. Единственная разница в том, что в формализованной системе все контексты заранее известны, а в открытой системе контексты возникают динамически и заранее не предопределены. В архитектуре HL-X тезаурус понятий должен быть дополнен всеми моделями и контекстами использования этих понятий, зафиксированных в архетипах документов.

Требования информационные. Современное направление развития информатики как науки связано, в первую очередь, с интеллектуализацией информационных систем и переходом от парадигмы накопления формализованных данных (базы данных) к накоплению формализованных знаний (базы знаний), «открытию» новых неизвестных знаний. В основе этого процесса лежит контекстно-зависимая интерпретация данных. Следует учесть стратегию развития информатики и в этом направлении. В архитектуре создается новый документальный уровень представления знаний. Каждый документ HL-X, структурированный в виде понятий и концептов (объектов) предметной области, рассматривается нами как носитель знания. Это знание в формализованном виде будет представляться в виде тезауруса понятий и в виде множества контекстов использования любого понятия. Знание должно автоматически «усваиваться» системой непосредственно из документов, сразу же после появления в системе нового архетипа и соответствующих ему экземпляров документов. На основе этого знания создается конструктор документов. При выборе понятия для добавления его в конструируемый документ система сама будет предлагать разработчику документа все известные ей контексты использования этого понятия. Подробнее о контекстном анализе знаний изложено в работах [5, 6].

Требования функциональные. Основными системообразующими функциональными свойствами документа HL-X, отражающими его целостность, должны стать транзакционность и инкапсуляция данных документа. Эти свойства означают, что документ и все его содержимое (структурированные формализованные данные и знания) обрабатываются как единое целое, как одна транзакция. Изменение данных документа, например непосредственно в таблицах БД, вне рамок документа, недопустимо. Информационной системе с поддержкой документа HL-X присуща следующая функциональность:

- транзакционность и инкапсуляция данных (целостность);
- авторизация любой информации, содержащейся в документе HL-X;
- историчность представления документа;
- шаблонизируемость содержания документа;
- поддержка мультимедийных типов данных;
- возможность разметки графических данных;
- поддержка неограниченного объема данных;
- возможность получения твердой копии любого документа;
- внедрение (заимствование) информации из других документов;
- обмен информацией с другими системами;
- конструирование документов на уровне прикладного разработчика;



■ Эволюционный процесс концептуализации предметной области в архитектуре HL-X

- поддержка коллективной работы с документами;
- динамическая генерация пользовательского интерфейса к документам;
- возможность работы с документами HL-X в тонком клиенте;
- возможность автономной работы с документами HL-X в off-line;
- возможность выгрузки документа в формате XML.

Требования архитектурные. Как уже отмечалось, документ HL-X должен быть реализован в открытой архитектуре и обладать известной инвариантностью и независимостью от быстро развивающихся информационных технологий. Это требование приводит к необходимости создавать высокоабстрактное модельное представление документов и средств работы с моделями. Документы HL-X должны «уметь» работать как в сосредоточенном информационном пространстве в рамках одного информационного узла, так и в распределенном информационном пространстве в рамках различных информационных узлов. Особо следует рассмотреть возможность автономной работы документов HL-X в узлах информационного пространства, не обеспечивающих связь on-line с центральным узлом. Речь идет об удаленных от центрального ядра системы участках, характеризующихся невозможностью или экономической нецелесообразностью организации связи on-line с центральным ядром системы. Архитектура документов HL-X должна уметь разворачиваться в информационных пространствах самой разной конфигурации и на самых разных технических средствах, начиная с мощного многосерверного

центрального узла и вплоть до единственного персонального компьютера, совмещающего функции клиента и сервера. В связи с этим предлагается отделить хранилище и модели документов от базы данных, поместив их в базу знаний. База знаний должна обладать свойствами переносимости на узлы произвольной конфигурации и требуемой масштабируемостью. Основная работа при использовании архитектуры HL-X в МИС заключается в построении множества архетипов документов над понятиями и концептами предметной области (проблемно-ориентированные модели документов). Одна из важнейших идей архитектуры HL-X — *введение процесса концептуализации предметной области непосредственно в саму информационную систему*. Это ответ на отмеченные выше трудности, связанные с недостаточной формализацией и с динамичностью предметной области. Иллюстрацией к сказанному служит рисунок.

Заключение

Приведем основные тезисы, отражающие суть предлагаемой архитектуры HL-X, и укажем на ее преимущества.

- Архитектура признает неизбежность стандартизации медицинской информации и предполагает, что в основу будущих открытых для расширения стандартов лягут словари понятий и концептов. Стандарты будут определять жесткие модели данных (понятия и концепты предметной области) и правила (ограничения) конструирования информационных структур на основе понятийного и концептуального тезауруса предметной области.

- Архитектура характеризуется достаточно высоким уровнем абстракции, что делает ее независимой от быстро изменяющихся информационных технологий. Документ HL-X является синтезом нескольких высокоабстрактных моделей, описывающих его структурные и функциональные свойства.

- Архитектура предлагает разделить хранилище архетипов документов HL-X и БД. Предполагается возможность распределенного хранения архетипов между узлами единого информационного пространства системы с обменом знаниями об архетипах между узлами.

- Процесс концептуализации предметной области, заключающийся во вводе в систему новых понятий, концептов и архетипов документов, должен быть непосредственно встроен в архитектуру системы.

Преимущества документов HL-X:

- свобода от конкретных, подверженных постоянному развитию и изменению технологий реализации ИС;

- самоценность документа HL-X как носителя модели информации, описывающей определенную предметную область, как носителя знания об этой предметной области;

- продление времени жизни модели документа в силу ее инвариантности по отношению к реализации документа HL-X в ИС;

- гибкость при реализации ИС на базе документов HL-X: возможность произвольного выбора БД (реляционной, постреляционной или объектной) и различных технологий реализации интерфейса документа HL-X. В качестве языков реализации архитектуры использованы PL/SQL, JavaScript, Java, HTML, XML, DotNet;

- восприимчивость к инновациям в информационных технологиях — возможность быстрого переноса накопленных моделей документов HL-X в новую информационную среду без их существенных переделок и развертывания ИС на базе новых технологий.

Глубина и широта реализации концепции документа HL-X в ИС определяется конкретными требованиями системы. В простейшем случае обработка данных документов может не производиться. Система будет оперировать документами как единым целым. В этом случае могут не понадобиться модель обработки данных и функциональная модель документа, особенно если эти документы импортируются из другой ИС и их изменение в принимающей ИС не предполагается. Визуализация может быть сведена к понятийной модели документа и не потребует дополнительных моделей визуализации. Полная реализация документа HL-X в ИС должна, по нашему мне-

нию, обладать следующими важными характеристиками:

- поддерживать развитые (многоязычные, синонимические) словари понятий для определенных предметных областей;

- предоставлять банки данных моделей документов HL-X (архетипы — совершенные образцы), описывающих определенную предметную область. Ценность моделей должна быть подтверждена их реальным практическим использованием в различных МИС, соответствовать государственным или отраслевым стандартам, поддерживаться мнением экспертов; поддерживать базу знаний «усвоенных» из архетипов документов;

- поддерживать средства разработки моделей документов HL-X (конструктор документов) на основе БЗ;

- предоставлять средства импорта/экспорта словарей понятий и моделей документов;

- предлагать различные технологические реализации механизмов работы с HL-X документами (хранилища документов, пользовательский интерфейс к документам).

Литература

1. Гулиев Я. И., Малых В. Л. Архитектура HL-X // Программные системы: Теория и приложения. М.: Физматлит, 2004. Т. II. С. 147–168.
2. McDonald C. J., Overhage J. M., et al. What is done, what is needed and what is realistic to expect from medical informatics standard // International Journal of Medical Informatics. Febr. 1998. Vol. 48. N. 1. P. 5–12.
3. Емелин И. В., Перов Ю. Л., Серегин Ю. С., Эльчиан Р. А. Концепция построения открытых медицинских информационных систем // Кремлевская медицина. Клинический вестник. 1998. № 1. С. 147–156.
4. Емелин И. В. Интеграция стандартов медицинской информатики // Кремлевская медицина. Клинический вестник. 2000. № 4. С. 68–76.
5. Малых В. Л., Юрченко С. Г. Документальный уровень представления знаний в интегрированной медицинской информационной системе // Программные системы: Теория и приложения. М.: Физматлит, 2004. Т. II. С. 217–230.
6. Гулиев Я. И., Малых В. Л., Юрченко С. Г. Контекстный анализ событий и синтез структуры медицинских знаний // Современные информационные и телемедицинские технологии для здравоохранения: Материалы II Междунар. конф. АИТН'2008 / Объединенный институт проблем информатики Национальной академии наук Беларуси. Минск, 2008. С. 164–168.