

удк 519.68

В. Л. Малых, С. Г. Юрченко

Документальный уровень представления знаний в интегрированной медицинской информационной системе

Аннотация. В работе документ рассматривается как носитель знаний. При этом на структуру и содержание документа не накладывается никаких существенных ограничений, за исключением четкого выделения в структуре документа понятий и концептов предметной области и удовлетворения требований к контекстам структурных элементов. Считается, что любые экземпляры документов, как и их прототипы (модели документов, архетипы документов), содержат заранее не формализованное знание о предметной области, описываемой в этих документах. Знания, заключенные в носителях (документах), рассматриваются и формализуются с помощью контекстного анализа документов. Вводится понятие контекста в различных смыслах, как для самого документа, так и для его структурных элементов. Указаны практические задачи, которые могут быть поставлены и решены с помощью контекста. Изложение ведется применительно к области медицинской информатики, но сам подход свободен от конкретики рассматриваемой предметной области.

Ключевые слова и фразы: Информационная Система, Документ, Контекст.

1. Введение

Основные проблемы, стоящие перед разработчиками медицинских информационных систем (МИС), описаны в [1]. Для связности изложения темы статьи приведем несколько выдержек из [1]:

«Примерно половина концептов (500 000) из всех выработанных человечеством непосредственно относится к медицине. Многоязычный тезаурус UMLS (Унифицированная система медицинского языка) содержит 730 000 концептов и 1,5 млн. понятий». Количество документов, ежегодно создаваемых только в одном крупном госпитальном учреждении, может измеряться сотнями тысяч и иметь объемы миллионов страниц. Понятийная и концептуальная база предметной области (медицины) недостаточно формализована и постоянно расширяется.

Перед разработчиками МИС постоянно возникает проблема необходимости расширения понятийной и концептуальной базы системы.

Необходимо по требованию пользователей вводить в систему новые понятия, концепты и документы. Этот процесс идет столь динамично, что, по мнению авторов [1], «процесс концептуализации предметной области . . . должен быть непосредственно встроен в архитектуру системы». Предоставляя достаточную свободу и легкость в расширении базы системы и конструировании документов, нельзя забывать о проблемах анализа данных. Наибольшую свободу при создании документов нам предоставляют редакторы типа Word. Однако анализ неструктурированных текстовых файлов весьма затруднен. Накопленные и апробированные методы разбора и анализа свободных текстов (искусственный интеллект) не могут удовлетворить нужды медицинской информатики в получении достоверной статистики или гарантированно точных ответов на поставленные пользователем формальные вопросы (демографические данные пациента, анамнез, жалобы, противопоказания и т. п.). Поэтому необходимо диалектически сочетать свободу при создании новых типов документов с формализацией содержимого документов. Понятия и концепты, входящие в документ, должны быть формализованы (вне самого документа), и ограничения на порядок расположения элементов в его структуре должны быть минимальны. При этом, в зависимости от месторасположения элемента в структуре документа, может меняться его семантическое значение. Полноценную семантическую интерпретацию формальных элементов данных структурной модели можно дать только с учетом их контекста, их положения в структуре документа. Это приводит к необходимости рассмотрения методов контекстного анализа данных, их возможностей для решения практических задач медицинской информатики.

Прежде всего, введем определения необходимых нам терминов. Сделаем это на основе работы [2].

2. Основные понятия структурной модели документа

2.1. Понятие. Все слова языка, которым мы владеем и на котором формулируем и передаем друг другу различную информацию, являются для нас семантически значимыми понятиями, абстрактными или конкретными.

Например: абстрактное понятие «человек» и конкретный человек — «Иван Иванович Иванов».

«Отвлеченное П. (понятие — прим. авторов) противопоставляется конкретному в том смысле, что отвлеченное обозначает собою признак, конкретное — предмет.» Брокгауз Ф. А., Ефрон И. А. Энциклопедический словарь.

«... абстрактное понятие в обычном смысле слова ... совпадает с понятием понятия как множества ситуаций ... Множеству, состоящему из одной ситуации, ... соответствует в традиционной логике конкретное понятие...» Турчин В. Ф. Феномен науки.

2.2. Атомарный факт. В качестве элементарной семантической единицы информации мы предлагаем рассматривать пару из абстрактного и конкретного понятия (выражения, фиксирующего некоторое эмпирическое знание), назовем ее атомарным фактом.

<абстрактное понятие> — <конкретное понятие>

Например: «Температура» — «36,6 гр. Цельсия»

В языке мы часто указываем только на конкретное понятие, опуская соответствующее абстрактное. Но при этом подразумеваем существование такой связи.

2.3. Информационная модель атомарного факта. Обычно требуется дополнительно описывать конкретное знание, входящее в атомарный факт. В частности, может понадобиться определение типа данных, а также формата представления этого знания.

(Имя абстрактного понятия, (Конкретное понятие, Тип данных, Формат данных))

Например: («Температура», («36,6», «число», «гр. Цельсия»))

2.4. Документ. По своему смыслу понятие документа близко примыкает к понятию текста.

«Текст — это последовательность осмысленных высказываний, передающих информацию, объединенных общей темой ... обладающая свойствами связности и цельности». В. Руднев. Словарь культуры XX в.

По аналогии с понятием текста определим документ как *последовательность осмысленных атомарных фактов, передающих информацию, объединенных общей темой, обладающих свойствами связности и цельности.*

Документ — это совокупность логически и семантически связанной информации. При этом он может включать как текст, так и изображения и звуки (закодированные определенным образом).

Можно постулировать, что документ имеет древовидную структуру, в узлах которой находятся атомарные факты.

2.5. Концепт. Концепт обладает заданной структурой и является «более крупным» структурным элементом, чем атомарный факт. В качестве примера концепта приведем сокращенное описание персоны из стандарта обмена медицинской информацией HL7. В рамках этого широко распространенного американского стандарта персона описывается с помощью набора атомарных фактов: пол, дата и место рождения, имя и фамилия, адрес, псевдоним, родной язык, гражданство, этническая группа, исповедуемая религия, девичья фамилия матери и т. д. Ясно, что в предметной области могут быть выделены достаточно «устойчивые», подробно разработанные концепты, использование которых облегчает задачу конструирования документов и задачу анализа данных.

Следует отметить, что концепт фактически является фреймом, как он понимается в теории представления знаний. Концепты, описанные в тезаурусе, будут составлять множество аксиом описанной в нем формальной системы.

Документ, по нашему утверждению, является структурой из атомарных фактов, или, с учетом введенных выше определений, *структурой из понятий и концептов.*

Структурная модель документа фактически представляет собой вышеназванные элементы, размеченные определенным образом. Для описания структурной модели документа предлагается использовать язык разметки XML, ставший де-факто стандартом при решении подобных задач.

Приведем пример концепта — заголовок истории болезни:

```
<ЗАГОЛОВОК_ИСТОРИИ_БОЛЕЗНИ>  
<ИДЕНТИФИКАТОР/>  
<ЗАГОЛОВОК/>  
<ФИО_ПАЦИЕНТА/>  
<КОД_ИСТОРИИ_БОЛЕЗНИ/>
```

```

<ДАТА_РОЖДЕНИЯ/>
<ЛЕЧАЩИЙ_ВРАЧ/>
<ДАТА_ГОСПИТАЛИЗАЦИИ/>
<ДЛИТЕЛЬНОСТЬ_ГОСПИТАЛИЗАЦИИ/>
<ВОЗРАСТ/>
<ПОЛ/>
<ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА/>
</ЗАГОЛОВОК_ИСТОРИИ_БОЛЕЗНИ>

```

В реальном документе он будет заполнен конкретными данными, например, такими:

```

<ЗАГОЛОВОК_ИСТОРИИ_БОЛЕЗНИ>
<ИДЕНТИФИКАТОР>1234</ИДЕНТИФИКАТОР>
<ЗАГОЛОВОК>ИБ 1 Петров Иван Иванович д.р. 30.10.1960</ЗАГОЛОВОК>
<ФИО_ПАЦИЕНТА>Петров Иван Иванович</ФИО_ПАЦИЕНТА>
<КОД_ИСТОРИИ_БОЛЕЗНИ>1</КОД_ИСТОРИИ_БОЛЕЗНИ>
<ДАТА_РОЖДЕНИЯ>30.10.1960</ДАТА_РОЖДЕНИЯ>
<ЛЕЧАЩИЙ_ВРАЧ>Иванов Петр Петрович</ЛЕЧАЩИЙ_ВРАЧ>
<ДАТА_ГОСПИТАЛИЗАЦИИ>10.03.2000</ДАТА_ГОСПИТАЛИЗАЦИИ>
<ДЛИТЕЛЬНОСТЬ_ГОСПИТАЛИЗАЦИИ>12</ДЛИТЕЛЬНОСТЬ_ГОСПИТАЛИЗАЦИИ>
<ВОЗРАСТ>39</ВОЗРАСТ>
<ПОЛ>М</ПОЛ>
<ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>4321
</ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>
</ЗАГОЛОВОК_ИСТОРИИ_БОЛЕЗНИ>

```

Итак, мы можем конструировать из понятий концепты, а из понятий и концептов — архетипы документов. В процессе концептуализации и конструирования мы обладаем определенной свободой выбора структуры элементов и документов.

Документ рассматривается нами как формализованная, в первую очередь с понятийной точки зрения, структура данных. Элементы, в него входящие, описываются вне документа, в так называемом тезаурусе, содержащем понятия и концепты как основной (медицины), так и вспомогательных предметных областей (управление организацией и персоналом, статистика, материально-техническое снабжение, диетпитание и т. п.). Объем тезауруса потенциально не ограничен. В тезаурусе используется фреймовая модель представления знаний. Для построения формальной логической модели в тезаурусе могут быть описаны множество синтаксических правил, гарантирующих то, что документы и концепты будут описывать синтаксически правильные совокупности, а также множество правил вывода, обеспечивающих возможность формирования на основе имеющихся концептов новых

синтаксически правильных совокупностей. Это обеспечит возможность автоматизации процесса конструирования новых документов.

Отдельно следует указать на сложность процесса формирования тезауруса понятий и концептов. Этот процесс вовсе не однозначный. Поясним это на примере из [1].

Пусть у нас имеются в словаре понятия: «группа», «инвалидность», «кровь». Перед нами стоит выбор — 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 как концепты в виде соответствующих структур:

```
<ИНВАЛИДНОСТЬ>
  <ГРУППА>ПЕРВАЯ</ГРУППА>
</ИНВАЛИДНОСТЬ>
```

```
<КРОВЬ>
  <ГРУППА>ПЕРВАЯ</ГРУППА>
  <РЕЗУС>РН+</РЕЗУС>
</КРОВЬ>
```

Для правильного понимания значения высказывания «группа — первая», нам необходим контекст использования термина «группа». самого термина (соответствующего атомарного факта) уже недостаточно. В этом заключается основное отличие всех контекстно-зависимых языков — естественных языков — от формальных алгоритмических и математических языков. Другой вариант вышеуказанных структур будет выглядеть так:

```
<ИНВАЛИДНОСТЬ>
  <ГРУППА_ИНВАЛИДНОСТИ>ПЕРВАЯ</ГРУППА_ИНВАЛИДНОСТИ>
</ИНВАЛИДНОСТЬ>
```

```
<КРОВЬ>
  <ГРУППА_КРОВИ>ПЕРВАЯ</ГРУППА_КРОВИ>
  <РЕЗУС>РН+<РЕЗУС>
</КРОВЬ>
```

Кажется, на первый взгляд, что мы избавились от контекстной зависимости, но это не так. Кто инвалид? Чья группа крови? Все равно для понимания этих данных необходим дополнительный контекст. Избавиться совсем от контекстной зависимости и необходимости проведения контекстного анализа 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.

Представленный пример показывает необходимость введения в рассмотрение понятия контекста. С формальной точки зрения каждый элемент в структуре документа является атомарным фактом, возможно, снабженным дополнительными техническими атрибутами. Введем для каждого элемента понятие контекстов различного уровня и определим механизм их формирования.

3. Определения контекстов

Контекст структурного элемента — множество ассоциированных (приписанных) структурному элементу атомарных фактов — контекстных переменных.

В зависимости от методов формирования множества атомарных фактов можно ввести в рассмотрение контексты различного уровня. Механизм формирования контекста основывается на однозначно заданном порядке обхода структуры документа, начиная с корневого элемента — обходе дерева в глубину, с однозначным порядком выбора из каждого подмножества узлов, имеющих общего родителя. В ходе обхода каждый из непосредственных предков какого-либо узла может явно установить или переопределить для него, точнее для любого из своих потомков, произвольное множество контекстных переменных. Множество контекстных переменных, ассоциированное в силу описанного механизма с данным узлом, называется *локальным контекстом*. В модели документа для указания того, какие атомарные факты необходимо поместить в динамически формируемый локальный контекст, используются определенные в грамматике документа теги разметки.

```
<ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>54321  
  <ToContext>ПЕРСОНАЛЬНЫЙ ИДЕНТИФИКАТОР ПАЦИЕНТА</ToContext>  
  <КОД_АМБУЛАТОРНОЙ_КАРТЫ>105</КОД_АМБУЛАТОРНОЙ_КАРТЫ>  
</ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>
```

В данном примере элемент `<КОД_АМБУЛАТОРНОЙ_КАРТЫ>` в своем локальном контексте имеет атомарный факт «ПЕРСОНАЛЬНЫЙ ИДЕНТИФИКАТОР ПАЦИЕНТА» — «54321», что семантически означает принадлежность амбулаторной карты с указанным кодом 105 персоне с указанным в локальном контексте идентификатором 54321.

Локальный контекст, в силу указанного механизма своего формирования, может быть установлен для элемента структурной модели только элементами, являющимися в отношении него непосредственными предками. *Глобальный контекст структурного элемента* определяется как множество атомарных фактов (контекстных переменных), установленных или переопределенных глобально всеми структурными элементами, пройденными до заданного, в силу однозначного порядка обхода документа.

```
<ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>54321
```

```
<ToContext Area="Global">ПЕРСОНАЛЬНЫЙ ИДЕНТИФИКАТОР ПАЦИЕНТА</ToContext>
```

```
<КОД_АМБУЛАТОРНОЙ_КАРТЫ>105</КОД_АМБУЛАТОРНОЙ_КАРТЫ>
```

```
</ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА>
```

Данная разметка означает, что при обходе указанного структурного элемента атомарный факт «ПЕРСОНАЛЬНЫЙ ИДЕНТИФИКАТОР ПАЦИЕНТА» — «54321» будет помещен в глобальный контекст. Глобальный, динамически создаваемый контекст структурных элементов используется для их семантической интерпретации при разборке документа. Удобство глобального контекста состоит в том, что структурные элементы могут установить между собой «контекстные» отношения даже в том случае, когда они не находятся в отношении «предок — потомок» относительно структуры документа. Формальный недостаток данного определения глобального контекста состоит в том, что элементы, которые в силу обхода документа будут пройдены позже заданных элементов, не могут предоставить своего контекста элементам, пройденным ранее. Корневой узел структуры документа, с которого начинается обход, может предоставить свой контекст локально или глобально любому другому элементу структуры. Локальный контекст корневого элемента документа называется также *локальным контекстом документа*.

В принятой модели разметки документа корневой элемент (концепт) называется просто «контекстом». Приведем пример структуры этого концепта:

```
<КОНТЕКСТ>
```

```
<ТИП_ДОКУМЕНТА/>
```

```
<ГЛАВНЫЙ_ОБЪЕКТ_ДОКУМЕНТА/>
```

```
<ИДЕНТИФИКАТОР_АРХЕТИПА_ДОКУМЕНТА/>
```

```
<ИДЕНТИФИКАТОР_МОДЕЛИ_ВИЗУАЛИЗАЦИИ/>
```

```
<ОПЕРАЦИЯ_НАД_ДОКУМЕНТОМ/>
```

```
<ИДЕНТИФИКАТОР_ ДОКУМЕНТА/>
```

```
<БЛОКИРОВКА/>
```

```
<КЛЮЧ_БЛОКИРОВКИ/>
```



```
<ДАТА_ПОСЛЕДНЕЙ_МОДИФИКАЦИИ/>  
<ИДЕНТИФИКАТОР_ОПЕРАТОРА/>  
<ИДЕНТИФИКАТОР_АВТОРА/>  
<ИДЕНТИФИКАТОР_ИСТОРИИ_БОЛЕЗНИ/>  
<ПЕРСОНАЛЬНЫЙ_ИДЕНТИФИКАТОР_ПАЦИЕНТА/>  
<ДАТА_СОЗДАНИЯ/>  
<СТАТУС/>  
</КОНТЕКСТ>
```

Дадим определение *глобального контекста документа*. Под глобальным контекстом документа будем понимать множество всех атомарных фактов, включенных в документ. Глобальный контекст документа будет содержать все данные документа, но при этом утратит информацию о структуре данных.

Все вышеопределенные контексты будем также называть *контекстами в широком смысле*. Эти контексты не несут структурной информации. Все практические задачи, которые легко могут быть решены с помощью этих контекстов, должны сводиться к теоретико-множественным операциям. Наиболее важной из таких задач является задача поиска документов по заданному контексту. Определяется образец — некоторое множество атомарных фактов, и отбираются те документы, локальные контексты которых включают в себя образец. Например, таким образом можно отобрать все документы, относящиеся к заданному пациенту, или к заданной медицинской карте, или имеющие заданного автора или дату создания. Аналогичные задачи поиска можно определить и для отдельных структурных элементов, задавая условия отбора в виде теоретико-множественных операций над образцами и контекстами элементов. Для усиления возможностей анализа и отбора необходимо определение различных операций над атомарными фактами, которые, как мы уже указывали выше, обладают типом и форматом. Например, для атомарных фактов типа числа или даты могут быть определены операции отношений «меньше», «меньше или равно», «равно», «больше», «больше или равно». Это позволит задавать с помощью операций более сложные условия отбора в виде логических выражений. Будут отбираться те документы, у которых не только заданные структурные элементы своим контекстом удовлетворяют контекстным образцам поиска, но и заданные атомарные факты удовлетворяют определенным логическим выражениям. Например, можно будет поставить задачу отбора документов, в которых определен тест на билирубин в биохимическом анализе крови, лежащий в заданном диапазоне значений, для пациентов

определенной возрастной категории с заданным диагнозом (заданным кодом диагноза по классификатору МКБ 10). Необходимо будет реализовать возможность использования логических операндов (логические «И», «ИЛИ», «НЕ» и т. п.).

Еще одной из важнейших задач является задача построения конструктора архетипов документов. В ее решении контекст также играет большую роль. Для структурного элемента документа — понятия или концепта — вводится понятие его *рабочего контекста* как множества атомарных фактов, необходимо присутствующих в объединении локального и глобального контекстов структурного элемента. Необходимость «удовлетворения» требований рабочего контекста накладывает дополнительные ограничения на архетипы документов, обеспечивая в некотором смысле семантическую корректность архетипа. Например, не следует включать в документ данные о результатах лабораторных исследований вне контекста конкретного пациента и его медкарты, даты забора лабораторных материалов и т. п. Однако, эти же исследования на калибровочных образцах, применяемых для контроля качества работы лабораторных анализаторов, уже не нуждаются в контексте пациента. Это означает, что рабочих контекстов для структурного элемента может быть определено несколько. Процесс конструирования архетипа выглядит следующим образом. Пусть уже имеется некоторая начальная структура архетипа (в тривиальном случае, один корневой элемент — носитель контекста документа), удовлетворяющая все рабочие контексты структурных элементов. Любое потенциальное изменение структуры (добавление или удаление структурных элементов, выбор другого рабочего контекста элемента) допускается только в том случае, если при этом удовлетворяются все рабочие контексты элементов новой структуры. Таким образом обеспечивается конструктивная целостность архетипа. Проверка целостности сводится к проверке вхождения рабочего контекста элемента в объединение локального и глобального контекстов этого элемента. В ходе конструирования из уже созданных архетипов документов могут заимствоваться любые подструктуры, удовлетворяющие при своем включении в новый архетип рабочие контексты элементов этих подструктур.

Введенное в работе понятие контекста структурного элемента оказалось достаточно конструктивным. Имеются и другие практически значимые задачи, для решения которых данное определение

контекста оказывается полезным. К таким задачам, широко распространенным в медицинской практике, относятся операции заимствования данных из одного документа в другой. Например, заимствование жалоб пациента из первичного осмотра в дневник. Формально в этом случае мы должны определить отображение атомарных фактов из одного документа в атомарные факты другого документа. В частном случае это может быть изоморфизм между двумя подструктурами. Подобные отображения желательно уметь строить алгоритмически вместо их явного определения в системе. В тривиальном случае отображение может быть основано на совпадении с точностью до синонимии имен абстрактных понятий отображаемых друг в друга атомарных фактов. В этом случае в вышеприведенном примере атомарный факт «группа», находящийся в контексте атомарного факта «инвалидность», может быть отображен в атомарный факт «группа», находящийся в контексте атомарного факта «кровь», что очевидно неверно. Для построения более корректного отображения можно использовать контексты элементов. Например, требовать не только совпадения имен абстрактных понятий, но и включения контекста отображаемого атомарного факта в контекст факта-отображения.

Мы привели достаточно неформальное концептуальное описание идей введения и использования понятия контекста документа и его структурных элементов. Дальнейшее направление развития этой темы предполагает более строгую формализацию.

В заключение еще раз опишем процесс концептуализации знаний на документальном уровне.

- (1) *Введение в тезаурус новых понятий*, необходимых для построения архетипов документов. Обычно для этого анализируются документы, которые необходимо создать и добавить в систему. Для каждого поля или раздела документа проверяется, существует ли уже в тезаурусе такое же или близкое (синонимичное) понятие; если нет, то оно добавляется. Одновременно добавляется и перевод понятия с русского на различные языки (английский и др.) в случае поддержки многоязычного тезауруса понятий.

- (2) *Построение на основе имеющихся понятий концептов*, описывающих концептуализированные знания о предметной области, и необходимых для построения архетипов новых документов. Концепты создаются для общепринятых сущностей (объектов) предметной области, имеющих унифицированные представления и модели. Для создания нового концепта необходимо вначале создать понятия, входящие в него.
- (3) *Создание архетипов документов* — по сути, концептуализация контекстных отношений между структурными элементами архетипа. Имеется в виду построение (древовидной) структуры документа из существующих понятий и концептов. При этом осуществляется проверка контекстных зависимостей для того, чтобы документ содержал все данные, необходимые для понимания как всего документа, так и его структурных элементов.

4. Заключение

- В работе рассматривается проблема представления и анализа документов произвольной структуры.
- В качестве основного структурного элемента модели представления данных предлагается принять атомарный факт. На основе атомарных фактов конструируются концепты и архетипы документов.
- Жесткая формализация понятий и концептов предметной области сочетается со свободой в определении структуры документа, при условии удовлетворения структурой необходимых рабочих контекстов ее элементов.
- Свобода в выборе структуры архетипа документа приводит в общем случае к необходимости использования контекстного анализа для обработки и интерпретации данных и формализации знаний, заключенных в документе.
- Введены различные понятия контекста, как для документа, так и для отдельных его элементов.
- Указана возможность определения контекстов как множеств атомарных фактов, а также вытекающие из этого определения практические возможности.
- Указана неизбежность структурного контекстного анализа для решения ряда практических задач (отображения документов друг в друга).

- Намечены пути развития контекстного анализа документов, соответствующих архитектуре HL-X [1].

Список литературы

- [1] Гулиев Я. И., Малых В. Л. *Архитектура HL-X // Программные системы: теория и приложения: Труды конференции.* — Москва: Физматлит, 2004 (готовится к публикации). ↑1, 2.5, 4
- [2] *Концепция интеграции системы Интернет с другими информационными системами:* Отчет по НИР/ ИПС РАН. — Переславль-Залесский, 2001. ↑1

ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР МЕДИЦИНСКОЙ ИНФОРМАТИКИ ИПС РАН

V. L. Malykh, S. G. Yurchenko. *Documentary level of knowledge representation in integrated medical information system.* (in russian.)

АБСТРАКТ. In this article document is considered a knowledge medium. At that no significant restrictions are applied to structure and contents of the document, except for well-defined marking of terms and concepts of data domain in the structure of the document, and implementation of requirements about context of structural elements. Any instance of document, as well as document prototype (document model, document archetype) a reputed to contain knowledge about data domain that hasn't been formalized yet, knowledge that is described in that document. Knowledge contained in that medium — in documents is viewed and formalized with the help of contextual analysis of the document. Different meanings of term 'context' are introduced, both for the document and for it's structural elements. Real-world problems that can be posed and solved with the help of context are denoted in the article. Medical informatics domain is discussed in the article, but the approach is not limited to it.