

РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

А В Т О М Е Т Р И Я

№ 2

1993

МЕДИЦИНСКИЕ ИНФОРМАЦИОННЫЕ И ИЗМЕРИТЕЛЬНЫЕ КОМПЛЕКСЫ

УДК 519.683 : 681.3.06

Б. Л. Лисс, Ю. В. Рубо, С. П. Треккова

(*Новосибирск*)

СИСТЕМА ПРЕДСТАВЛЕНИЯ И ОБРАБОТКИ ДОКУМЕНТОВ (СПОД)

Описывается подход к представлению и обработке документов на ЭВМ. Документы представляются в виде совокупности разнородных данных (чисел, текстов, графических данных). Информация об их структуре и внешнем виде хранится отдельно от самих документов. Система поддерживает создание документов, их визуализацию в виде привычных «бумажных» документов и средства доступа к данным документов со стороны программ.

Введение. Система предназначена для автоматизации работы с документами на ПЭВМ и обеспечивает унифицированный способ представления документов и средства обработки их на ЭВМ.

В сфере человеческого общения под документом обычно понимается «деловая бумага, подтверждающая какой-либо факт или право на что-то». Функционирование современного общества и производства вызывает создание и оборот значительного числа документов. Применение вычислительной техники вплотную подвело к тому, чтобы документы хранились и обрабатывались на вычислительных машинах. Однако информация, получаемая в вычислительных системах и предназначенная для восприятия ее человеком в виде документов, как правило, представляется в текстовом (неформализованном!) виде (последовательностью символов). Если же эта информация должна обрабатываться на ЭВМ, то ее необходимо представить формализованным способом в виде набора данных так, чтобы она могла быть воспринята программами обработки.

Авторы описываемой системы совместили привычный для человека внешний вид документа с возможностью доступа к документу со стороны программ. Для этого были выделены две его составляющие: данные, семантически задающие документ, и его внешний вид (форма) — представление этих данных в удобном для восприятия виде.

Наш подход к представлению документа состоит в хранении собственно данных, входящих в документ, отдельно от информации о структуре и типах этих данных и об их внешнем представлении (отображении) на устройствах визуализации. Кроме того, мы стремились создать систему таким образом, чтобы введение новых видов документов не требовало изменений в программах, работающих с документами, а сводилось бы только к описанию видов документа.

В разд. 1 вводятся основные понятия и кратко излагается подход к представлению документов, в разд. 2 описывается структура системы и ее основные компоненты, в разд. 3 приводится пример создания нового вида документов.

1. **Основные понятия.** 1.1. *Понятие документа.* Под документом мы понимаем зафиксированную в определенном порядке совокупность данных (которая может быть представлена на различных материальных носителях инфор-

мации), предназначенную для общения между людьми,ющую иметь юридическую силу и носить служебный характер. Элемент документа — составная часть документа, которая имеет самостоятельное целостное (информационное) значение и с которой можно производить работу как с отдельной единицей. Элементы документа могут иметь значения следующих типов: целые и вещественные числа, тексты, перечислимые (строка из заданного множества строк), шкалы признаков (подмножество заданного множества строк), классификаторы, таблицы, графики, диаграммы, рисунки.

Сам документ является совокупностью элементов (в общем случае разнотипных).

1.2. Формуляры и бланки. Каждый документ имеет две формы представления. Первая — системная (машинная) — предназначена для хранения документа и его обработки на ЭВМ, вторая — внешняя человекоориентированная форма, определяющая вид документа на устройствах визуализации.

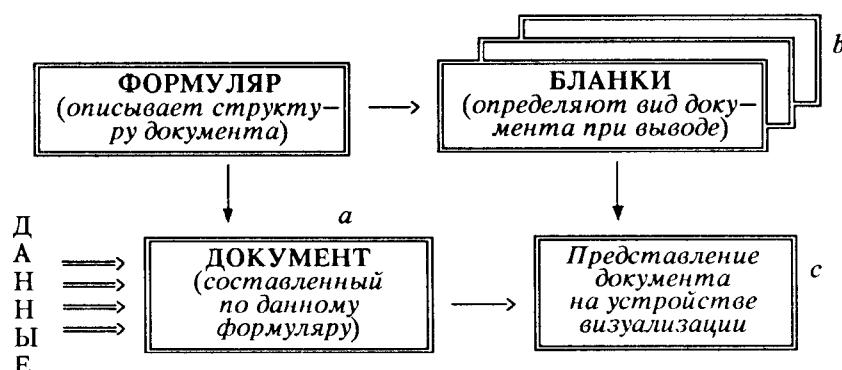
Первая форма задается формулляром — специальной информацией о структуре документа. Формулляр определяет состав и типы элементов документа. Он описывает класс однотипных документов (например, «платежное поручение» или «общий анализ крови»).

Вторая форма представления документа задается бланком — специальной информацией о том, как выводить документ на устройство отображения (общее оформление, расположение и способ отображения элементов документа), а также о способе ввода с клавиатуры значений его элементов. Бланк составляется для всего класса документов, заданных формулляром, и ориентирован на восприятие документа человеком. Для одного и того же формулляра может быть создано несколько бланков, предназначенных для отображения документов данного вида на разных устройствах визуализации и учитывающих особенности этих устройств.

На рис. 1 приведена схема, представляющая взаимоотношение между формуллярами, бланками и документами, где: *a* — документ задается совокупностью данных (значениями элементов), типы и структура которых описываются формулляром; *b* — для устройств визуализации на основе формулляра составляются бланки (или один бланк для нескольких устройств); *c* — представление документа на устройстве визуализации формируется на основе совокупности данных, представленных в документе, и одного из бланков.

В описываемой системе каждый бланк, формуляр или документ размещаются в отдельных файлах. В документе отмечается имя соответствующего формуляра, а указание нужного бланка при отображении документа осуществляется на уровне прикладных программ.

1.3. Элементы документа. 1.3.1. Общие атрибуты элементов документа. При описании структуры документа (создании формулляра) задаются его эле-



PUC 1

менты и их характеристики (атрибуты). Определенные атрибуты присущи элементам любого типа. К ним относятся:

- имя элемента (предназначено для указания элемента при работе с документом);
- свойство изменяемости значения элемента (определяет, можно ли менять значение элемента после создания документа);
- свойство обязательности заполнения элемента (определяет, может ли элемент быть пустым, т. е. не иметь значения);
- допустимые значения (определяют множество значений, которые может принимать элемент);
- значение элемента по умолчанию (определяет значение, которое будет установлено в том случае, когда при формировании документа значение явно не задано).

1.3.2. Типы элементов документа. Элементы документов бывают простыми и составными. Простой элемент может принимать только одно значение. Составной элемент является совокупностью нескольких простых значений.

К простым относятся элементы целого, рационального, перечислимого и текстового типа, а также шкалы и классификаторы.

К составным относятся таблицы, диаграммы, графики и рисунки. Для составных элементов возможен доступ как ко всему элементу в целом, так и к его подэлементу (или группе подэлементов).

Элементы целого типа. Элементы целого типа могут принимать целочисленные значения из заданного диапазона.

Элементы рационального типа. Элементы данного типа могут принимать рациональные значения из заданного диапазона и с заданной точностью.

Элементы текстового типа. Значением элемента данного типа является текст. Кодировка символов зависит от конкретной реализации. В формуляре при описании элемента данного типа указываются размеры текста — максимальная длина строки и количество строк. Ограничение на количество строк может быть не задано, в этом случае текст может иметь любое число строк.

Элементы перечислимого типа. Значение данного типа — символьная строка из заданного перечня строк. Эти строки задаются в формуляре при описании элемента. В документе данное значение хранится в виде целого числа, являющегося порядковым номером соответствующей строки в перечне, а сама строка используется при отображении.

Шкалы признаков. Значением данного типа является подмножество из заданного множества символьных строк. Эти строки задаются в формуляре при описании элемента. В документе шкала признаков представляется в виде ряда битов, где каждый бит последовательно соответствует строке, заданной в перечне. Значение данного типа отображается как последовательность входящих в него строк.

Классификаторы. Значением данного типа является последовательность символьных строк, формируемая в результате прохождения по древовидной структуре, каждой вершине которой соответствует некоторая строка. Описание дерева классификатора задается в отдельном файле, в формуляре же делается ссылка на этот файл. Такая организация позволяет вносить изменения в классификатор, не меняя формуляр. В документе значение классификатора хранится как последовательность чисел (номеров вершин), а при визуализации значение отображается в виде единой строки (составленной из строк вершин).

Таблицы. Таблица определяется как последовательность строк, каждая из которых состоит из последовательности в общем случае разнотипных простых элементов. Пользуясь терминами языков программирования высокого уровня, можно сказать, что таблица — это одномерный массив, каждый элемент которого является структурой. Стока таблицы соответствуют структуре, а элементы строки (столбцы) — полям этой структуры. Как следует из определения, в каждом столбце таблицы находятся только однотипные зна-

чения. Таблица может содержать строго фиксированное число строк (таблица фиксированной длины) или переменное число строк (таблица переменной длины). Работа с таблицей может осуществляться как с единым целым, так и на уровне ее подэлементов — строк или отдельных элементов.

Г р а ф и к. Значением данного типа является одна или несколько функций, графики которых подлежат выводу на устройства отображения. Функции в документе представляются табличным способом — последовательностью пар рациональных значений (значение аргумента и соответствующее ему значение функции). Доступ к графику возможен либо целиком, либо к отдельной функции.

Д и а г р а м м ы. Значением данного типа является диаграмма, которая при визуализации может быть отображена как столбиковая, столбиковая совмещенная или секторная. Значение представляется в виде одного или нескольких последовательностей (рядов) рациональных чисел (величин). Число рядов диаграммы и число величин в ряде задаются при описании формулляра. Если рядов несколько, то количество величин в каждом из них должно быть одинаково. Доступ к диаграмме возможен либо целиком, либо к отдельному ряду.

Р и с у н к и. Значением данного типа является рисунок, который подлежит отображению. Значение представляется массивом цветных точек (пикселов). Каждый пиксель представлен в виде набора битов, количество которых зависит от числа используемых цветов. Характеристики рисунка (количество точек по вертикали и по горизонтали, число битов, занимаемых одним пикселям, таблица кодов цветов) могут как задаваться в формулляре, так и храниться в документе вместе с самим рисунком. Доступ к рисунку возможен либо целиком, либо к отдельной пиксельной строке рисунка.

1.4. Документ как единица хранения и обработки информации. В отличие от реляционных баз данных, представляющих информацию в виде совокупности таблиц, в нашем подходе данные (в общем случае разнородные) сгруппированы по признаку принадлежности к документу. Такой способ организации применим в том случае, когда наиболее интенсивная работа с документом производится как с единым целым. Так, амбулаторная карта пациента является ярко выраженным примером представления информации о пациенте в виде совокупности документов, полученных при его обследовании. Эту информацию нужно представлять как в привычном текстовом виде, так и в формализованном виде, чтобы ее можно было обработать на ЭВМ.

2. Структура СПОД. Система представления и обработки документов состоит из трех основных компонентов: подсистемы создания формулляров и бланков, базовой подсистемы работы с документами, диалоговой подсистемы создания и просмотра документов.

2.1. Подсистема создания формулляров и бланков. Подсистема создания формулляров и бланков является отдельной программной системой. Она состоит из трансляторов формулляров, классификаторов и бланков. Назначение данной подсистемы — проверка правильности описания формулляра (классификатора, бланка) на языке описания формулляров (классификаторов, бланков) и трансляция его во внутреннее представление, с которым работают остальные компоненты системы.

2.2. Базовая подсистема работы с документами. Вся работа с документами (не как с файлами) осуществляется через базовую подсистему (БП). Она обеспечивает создание документов, доступ к элементам и подэлементам документов, получение информации о структуре документов и атрибутах элементов, движение по классификаторам. БП позволяет одновременно работать с несколькими документами. Обращение к подсистеме осуществляется посредством программных запросов.

Базовая подсистема представлена в виде двух библиотек подпрограмм, реализованных на языке Си.

2.3. Диалоговая подсистема создания и просмотра документов. Данная подсистема предназначена для поддержки процесса создания (заполнения) и корректировки документов непосредственно с клавиатуры ПЭВМ, просмотра их на экране дисплея и подготовки для печати. Вся работа с документами

осуществляется подсистемой через базовую подсистему, а для отображения документов на дисплей и при подготовке их к печати она опирается на информацию, представляемую в бланках.

Диалоговая подсистема построена таким образом, что предполагает наличие некой «оболочки» — прикладной программы, которая сама определяет стратегию работы с документами и использует подсистему в тех случаях, когда необходимо предоставить пользователю, работающему за ПЭВМ, возможность вручную вводить данные в создаваемый или корректируемый документ(ы), визуализировать документ или несколько документов на экране дисплея или подготовить для печати документа его образ (печать которого будет осуществлена оболочкой). Другими словами, эта подсистема предоставляет средства поддержки наиболее часто встречающихся способов работы с документами.

Диалоговая подсистема создания и просмотра документов реализована в виде библиотеки подпрограмм, написанных на языке Си и использующих библиотеки подпрограмм базовой подсистемы.

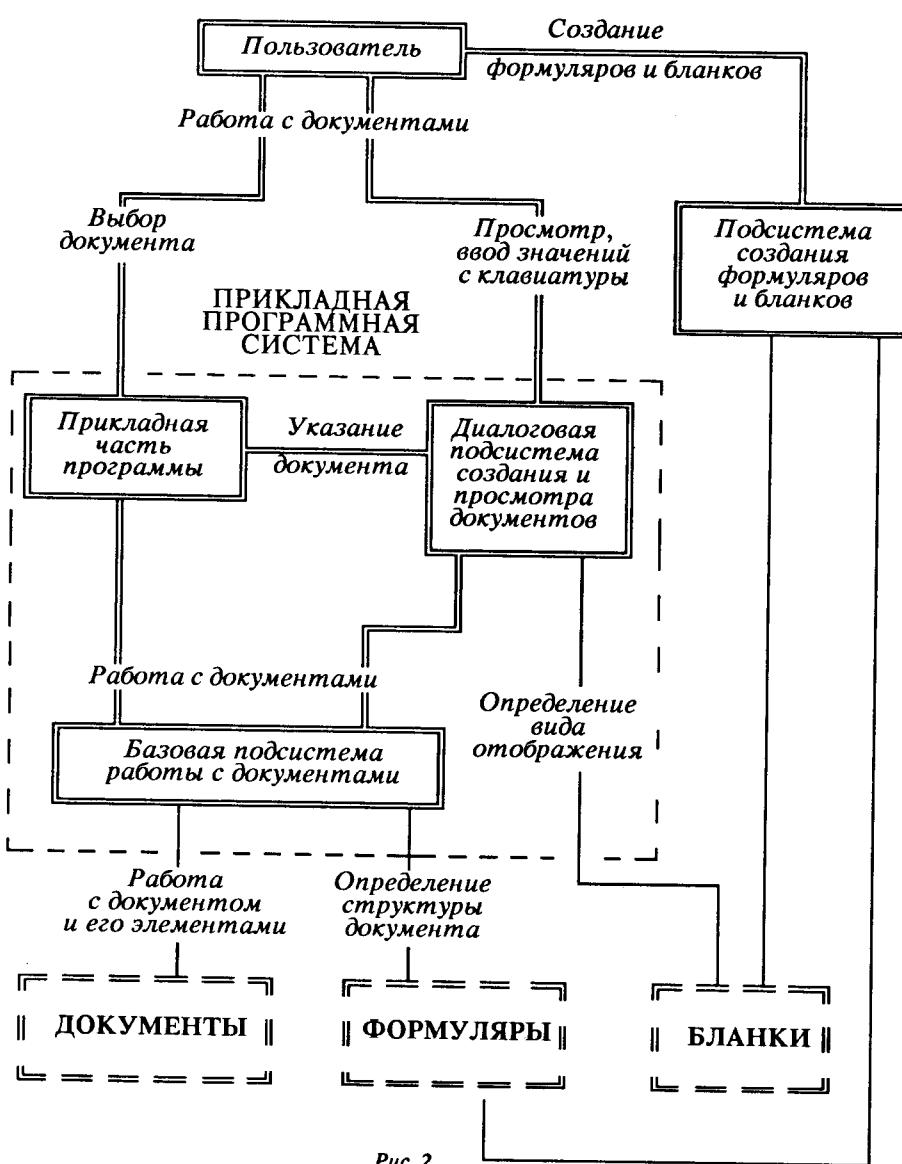


Рис. 2

2.4. Схема использования СПОД. Система представления и обработки документов ориентирована на использование ее в прикладных программах, работающих с документами. Базовая подсистема предоставляет средства для доступа к документу и его обработки. Диалоговая подсистема — средства визуализации и ввода данных в документ с клавиатуры. Подсистема создания формулейров и бланков обеспечивает разработку новых видов документов (рис. 2).

3. Пример создания нового вида документов. Для создания нового класса документов необходимо определить, какую информацию должны содержать документы этого класса, данными каких типов она будет представлена и каков будет внешний вид этих документов.

Сначала определяется назначение документов требуемого класса и на основе этого состав и характеристики его элементов (их тип и прочие атрибуты).

Структура и внешний вид класса документов описываются соответственно на языке описания формулейров (и классификаторов) и бланков. Затем эти описания переводятся соответствующими трансляторами во внутренние представления, с которыми в дальнейшем работают остальные компоненты системы. Оттранслированные формулейры, классификаторы и бланки хранятся в отдельных файлах.

Предположим, что необходимо составить новый вид документов, который назовем «основные данные о пациенте», для медицинского учреждения. Документы данного вида должны содержать следующую информацию:

а) паспортные данные (фамилия, имя, отчество, дата рождения; домашний адрес; место работы и должность и т. п.);

б) данные медицинского характера (группа и резус крови; факторы риска хронических заболеваний; диагнозы хронических болезней и годы их возникновения; применяемые постоянно медикаменты и др.).

3.1. Составление формулейра. При описании формулейра необходимо указать состав и типы элементов документа, а также их атрибуты (изменяемость, обязательность заполнения и др.).

Например, дата рождения будет состоять из трех обязательно заполняемых, неизменяемых элементов: двух элементов целого типа (день и год) соответственно с диапазонами 1—31 и 1880—2100 и одного элемента перечислимого типа со значениями «января», «февраля», «марта», ..., «декабря». Факторы риска будут реализованы шкалой, задаваемой множеством строк, из которых могут быть указаны одновременно несколько. Диагнозы хронических болезней задаются изменяемой, не обязательно заполняемой таблицей переменной длины, состоящей из трех столбцов:

— первый столбец задает наименование хронического заболевания и имеет тип «классификатор» и атрибуты неизменяемости и обязательности заполнения;

— второй столбец задает год постановки диагноза и является целым числом в диапазоне от 1900 до 2100, неизменяемым и не обязательно заполняемым;

— третий столбец задает примечание и является не обязательно заполняемым изменяемым текстом длиной до 120 символов.

Ниже приведены фрагменты описания формулейра для данного вида документов (на языке описания формулейров).

```
DPC:    ФОРМУЛЯР
{
ФАМ:    (ИЗМ)        ТЕКСТ      ДЛИНА 30;      \фамилия\
ИМЯ:    (ИЗМ)        ТЕКСТ      ДЛИНА 25;      \имя\
ОТЧ:    (ИЗМ НЕЗАП)  ТЕКСТ      ДЛИНА 25;      \отчество\
ДРОЖ:   ЦЕЛ          ОТ 1 ДО 31;    \день рождения\
МРОЖ:   ПЕР          ('января',
                     'февраля',
```

'марта',
'апреля',
'мая',
'июня',
'июля',
'августа',
'сентября',
'октября',
'ноября',
'декабря');

ГРОЖ: ЦЕЛ ОТ 1880 ДО 2100; \год рождения\

\...далее описание данных, содержащих информацию о домашнем адресе, телефонах,
месте работы, должности и пр.\

ГРКР: ПЕР ('O (I)',
'A (II)',
'B (III)',
'AB (IV)');

РЕЗФ: ПЕР ('+', '-'); \резус\

ФРИС: (НЕЗАП) ШКАЛА (\факторы риска\
'Гиперлипидемия',
'Избыточная масса тела',
'Артериальная гипертензия',
'Сахарный диабет',
'Курение',
'Гиподинамия',
'Психоэмоциональные перегрузки',
'Нерациональное питание');

КОММФР: (ИЗМ НЕЗАП) ТЕКСТ ПЕРЕМ ДЛИНА 45; \комментарий\

МЕД: (ИЗМ) ТАБЛИЦА \применяемые постоянно медикаменты\
{

СТОЛБЦЫ

{

МЕДИКАМ: (НЕИЗМ ЗАПОЛ) ПЕР (\медикамент\
'Инсулин',
'Кортикоиды',
'Препараты наперстянки',
'Бета-адреноблокаторы',
'Нитроглицерин',
'Нитросорбид',
'Изоптин (финоптин)',
'Хинидин',
'Этацизин',
'Амиодарон (кордарон)',
'Диуретики',

```

'Препараты раувольфии',
'Клофелин',
'Транквилизаторы',
'Снотворные средства',
'Препараты, содержащие железо',
'Другие медикаменты'

);

КОММЕД: (ИЗМ НЕЗАП) ТЕКСТ ДЛИНА 100; \комментарий\

};

\диагнозы хронических заболеваний\

ДИАГХР3: (ИЗМ) ТАБЛИЦА

{

СТОЛБЦЫ

{

ЗАБОЛ: (НЕИЗМ ЗАПОЛ) КЛАССИФ 'CHRON_ILL'; \заболевание\
\задано имя классификатора\

ГОД_ЗАБОЛ: (НЕИЗМ НЕЗАП) ЦЕЛ ОТ 1900 ДО 2100; \год\

КОМ_ЗАБОЛ: (ИЗМ НЕЗАП) ТЕКСТ ДЛИНА 120; \примечание\

};

СТРОКИ

[

ПЕРЕМ;

];

};

}

```

3.2. Составление бланка. При составлении бланка необходимо определить:

1) внешний вид документа: а) фон, на котором будут отображаться значения элементов документа; б) места (окна), в которых должны отображаться значения; в) способ отображения каждого элемента;

2) тексты запросов, которые должны выводиться диалоговой системой при вводе значений элементов документа с клавиатуры.

Окна, в которых отображаются элементы документа, могут быть фиксированные или переменные. Размеры фиксированного окна постоянны, размеры переменного окна постоянны только по горизонтали, по вертикали же они не фиксированы (т. е. в зависимости от значения они могут занимать нужное количество строк). В переменном окне могут отображаться элементы типа шкалы, текста или классификатора. Таблицы отображаются в так называемых композиционных окнах. В бланках имеется возможность указать расположение значения в окне (справа, слева, сверху, снизу, в центре). Для ряда

элементов можно указать способ изображения значения (например, для рационального — в виде десятичной дроби или с указанием степени десяти, для целого можно указать изображение нуля и т. п.).

Ниже приведено описание бланка для документов, содержащих основные сведения о пациенте. Заметим, что описание состоит из двух частей: а) последовательность заданий для каждого элемента способа отображения и запроса для ввода его значения с клавиатуры, б) трафарет, описывающий общий вид документа и расположение окон.

DPC: БЛАНК
{
'*' — (ФАМ) ТЕКСТ {ПОДСКАЗКА "Фамилия:";};
'*' — (ИМЯ) ТЕКСТ {ПОДСКАЗКА "Имя:";};
'*' — (ОТЧ) ТЕКСТ {ПОДСКАЗКА "Отчество:";};
'*' — (ДРОЖ) ЦЕЛ {ПОДСКАЗКА "Дата рождения (день):";
OKHO '#0';};
'*' — (МРОЖ) ПЕР {ПОДСКАЗКА "Дата рождения (м-ц):";};
'*' — (ГРОЖ) ЦЕЛ {ПОДСКАЗКА "Дата рождения (год):";};
'*' — (ПИНД) ТЕКСТ {ПОДСКАЗКА "Почтовый индекс:";};
'*' — (УЛИЦ) ТЕКСТ {ПОДСКАЗКА "Домашний адрес (улица):";};
'*' — (ДОМ) ТЕКСТ {ПОДСКАЗКА "Домашний адрес (N дома):";};
'*' — (КВ) ТЕКСТ {ПОДСКАЗКА "Домашний адрес (квартира):";};
'*' — (ДТЕЛ) ТЕКСТ {ПОДСКАЗКА "Телефон домашний:";};
'*' — (РТЕЛ) ТЕКСТ {ПОДСКАЗКА "Телефон служебный:";};
'*' — (МРАБ) ТЕКСТ {ПОДСКАЗКА "Место работы:";};
'*' — (ДОЛЖ) ТЕКСТ {ПОДСКАЗКА "Должность:";};
'*' — (УСЗВ) ПЕР {ПОДСКАЗКА "Ученая степень/звание:";};
'*' — (ГРКР) ПЕР {ПОДСКАЗКА "Группа крови:";};
'*' — (РЕЗФ) ПЕР {ПОДСКАЗКА "Резус:";};
'*' — (ФРИС) ШКАЛА {ПОДСКАЗКА "Факторы риска:";};
'*' — (КОММФР) ТЕКСТ {ПОДСКАЗКА "Комментарий к факторам риска:";};
 (МЕД) ТАБЛИЦА
 {ПОДСКАЗКА "Применяемые постоянно медикаменты:";
 '*' — (МЕДИКАМ) ПЕР {ПОДСКАЗКА "Медикаменты:";};
 '*' — (КОММЕД) ТЕКСТ
 {ПОДСКАЗКА "Комментарий к применяемым медикаментам:";};
 (ДИАГХРЗ) ТАБЛИЦА
 {ПОДСКАЗКА "Диагнозы хронических заболеваний:";
 '*' — (ЗАБОЛ) КЛАССИФ {ПОДСКАЗКА "Заболевание:";};
 '*' — (ГОД_ЗАБОЛ) ЦЕЛ
 {ПОДСКАЗКА "Год возникновения заболевания:";};
 '*' — (КОМ_ЗАБОЛ) ТЕКСТ
 {ПОДСКАЗКА "Комментарий к хроническому заболеванию:";};
 };
ВЫРАВ ДУБЛИР;

ТРАФАРЕТ

3.3. Пример заполненного документа. Далее приведен пример заполненного документа. Объем, занимаемый данным документом на внешней памяти, 1479 байтов (685 байтов в упакованном (ZIP) виде).

Заключение. В настоящее время на базе системы представления и обработки документов разработана прикладная система ведения электронных медицинских карт пациентов. Опыт создания данной прикладной системы показал следующее:

- СПОД может включаться в автоматизированные рабочие места различного профиля как средство работы с документами;
 - СПОД позволяет легко создавать новые классы документов, поскольку для этого достаточно только описать структуру и внешний вид документов нового класса на соответствующих специализированных языках;
 - СПОД обеспечивает компактное хранение документов, поскольку в файле, содержащем документ, хранятся только значения его элементов, а все, что относится к структуре информации и к представлению документа при

ОСНОВНЫЕ ДАННЫЕ О ПАЦИЕНТЕ

Фамилия Иванов Имя Петр Отчество Иванович
--

Дата рождения 3 февраля 1956 г.
Домашний адрес: почт. индекс 630090
 ул. Жемчужная
 дом 7 кв. 98
Телефоны: домашний 34-97-08 служебный 32-88-34
Место работы Институт математики
Должность Ст. научный сотрудник
Ученая степень/звание Кандидат наук

Группа крови В(III) резус +

ФАКТОРЫ РИСКА

<i>Избыточная масса тела</i> <i>Курение</i> <i>Психоэмоциональные перегрузки</i> <i>Нерациональное питание</i>

Комментарий _____

Курение до 30 сигарет в день

ПРИМЕНЯЕМЫЕ ПОСТОЯННО МЕДИКАМЕНТЫ

Медикаменты	Комментарий
-------------	-------------

<i>Нитроглицерин</i> <i>Снотворные средства</i> <i>Препараты, содержащие железо</i>	<i>Практически ежедневно</i>
---	------------------------------

ДИАГНОЗЫ ХРОНИЧЕСКИХ ЗАБОЛЕВАНИЙ

Заболевание	Год	Примечание
-------------	-----	------------

<i>Врожденный гипотиреоз</i> <i>Нефролитиаз</i>	<i>1980</i>	<i>В 1982 году проведена операция по удалению камней из правой почки</i>
<i>Хронический колит</i>	<i>1988</i>	

визуализации, хранится в единственном экземпляре для всех документов одного вида.

По мнению авторов, дальнейшие усилия по совершенствованию системы должны быть направлены в сторону: а) введения в описание документов средств указания элементов, значения которых зависят от значения других элементов (например, строка «Итого» в таблицах); б) создания универсальных средств задания выборок и статистических подсчетов; в) создания средств сопряжения системы с другими программными системами (в частности, с базами данных и т. п.).

Поступила в редакцию 7 декабря 1992 г.