

Тема 2. КОМПЬЮТЕРНЫЕ ДАННЫЕ. КОДИРОВАНИЕ И КЛАССИФИКАЦИЯ МЕДИЦИНСКОЙ ИНФОРМАЦИИ

ДАННЫЕ [data, information] - сведения, факты, показатели, выраженные как в числовой, так и любой другой форме.

Согласно другому определению: "Данные - некоторый факт, то, на чем основан вывод или любая **интеллектуальная система**". Компонентами данных являются цифры и символы естественного языка или их кодированное представление в виде строки двоичных **битов**.

БИТ [bit от *англ. binary digit*] - простое двоичное число (цифра или символ), принимающее значения 1 или 0 и служащее для записи и хранения данных в ЭВМ.

Бит является минимальной двоичной единицей измерения **энтропии и количества информации** в ЭВМ, соответствующей одному **двоичному разряду**. Энтропия сообщения, выраженная в битах, определяется средним числом символов, необходимых для записи этого сообщения. Определенное количество бит составляет размер других единиц – **двоичных слов**, в том числе – байта [byte], килобайта [kilobyte], мегабайта [megabyte], и т.д.

Байт [byte] – двоичное слово, способное записывать и хранить в памяти ЭВМ один буквенно-цифровой или другой символ данных.

Килобайт, кбайт [kilobyte] – 1024 байта или 10^3 байт.

Мегабайт, Мбайт [megabyte] - 1024 килобайт или 10^3 килобайт, или 10^6 байт.

Гигабайт, Гбайт [gigabyte] - 1024 мегабайт или 10^3 мегабайт, или 10^6 килобайт, или 10^9 байт.

Тип данных определяет:

- внутреннее представление данных в памяти компьютера;
- множество значений, которые могут принимать величины этого типа;
- операции и функции, которые можно применять к величинам этого типа.

Существующие типы компьютерных данных представлены на рис. 2.1 (классификация согласно [2]).



Рис.2.1. Типы компьютерных данных.

Кодирование данных – это переход от исходного представления данных, удобного для восприятия человеком, к представлению, удобному для хранения, передачи и обработки данных с использованием вычислительной техники.

Медицинские данные пациента – сведения о физиологических особенностях организма, перенесенных заболеваниях, состоянии здоровья и (или) оказанной пациенту медицинской помощи.

Для того чтобы медицинские данные можно было хранить в памяти компьютера, обрабатывать и передавать, они должны быть представлены определенным типом данных

Примеры представления медицинских данных:

- *целые числа*: некоторое дискретное число (напр., число лейкоцитов в образце крови);
- *действительные числа*: переменная, полученная при измерении того или иного показателя (напр., температура или давление крови);
- *код*: условное обозначение некоторой переменной (напр., боль);
- *текст*: напр., текст истории болезни или документация событий во время мониторинга;
- *изображения*: электрограммы – ЭКГ, ЭЭГ и др., реограммы, компьютерные томограммы, УЗИ, МРТ снимки и т.д.

Одной из основных проблем, связанных с представлением данных в памяти компьютера является их точность, корректность и полнота.

Точность данных – точность соответствия между значением свойства и истинным значением [3].

Корректность данных – это мера частоты появления ошибок в данных. Ошибки могут возникнуть во время сбора данных или их измерений.

Точность зависит от степени детализации. Например, количество десятичных знаков при измерении той или иной величины. Вес тела, выраженный как 89,12 кг, имеет большую точность, чем вес, выраженный, как 89,1 кг.

Полнота данных – необходимая достаточность и отсутствие избыточности данных. Например, представление температуры тела в виде 37,8° – достаточно, 37,831° – избыточно.

Обработка данных – процесс приведения данных к виду, удобному для использования.

Обработка данных включает три основные группы операций: подбор исходных, входных данных (сбор данных), собственно их обработку, получение и анализ результатов, т.е. выходных данных.

Собственно обработка данных компьютером – это арифметические операции сложения, вычитания, умножения и деления, а также логические операции принятия решений на основе сопоставления данных.

Для обработки данных на компьютере необходимо оборудование (аппаратное обеспечение) и программы (программное обеспечение), которые составляют *вычислительную систему*.

Одной из основных задач обработки медицинских данных является задача *классификации*.

Классификация – это осмысленный порядок вещей, явлений, разделение их на разновидности согласно каким-либо важным признакам [4].

Классификация предназначена для постоянного использования в какой-либо науке или области практической деятельности (например, классификация животных и растений, классификация заболеваний).

Критерий дифференциации (основание классификации) – критерий, по которому происходит разбиение на классы

Обычно в качестве основания деления в классификации выбирают признаки, существенные для данных предметов. В этом случае классификация называется *естественной*. Она выявляет существенные сходства и различия между предметами и имеет познавательное значение. В других случаях, когда цель классификации состоит лишь в систематизации предметов, в качестве основания выбираются признаки, удобные для этой цели, но несущественные для самих предметов (например, алфавитные каталоги). Такие классификации называют *искусственными*.

Наиболее ценными являются классификации, основанные на познании законов связи между видами, перехода от одного вида к другому в процессе развития (такова, например, классификация химических элементов, созданная Менделеевым) [5].

Задача классификации – формализованная задача, в которой имеется множество объектов, разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Классовая принадлежность остальных объектов неизвестна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Классифицировать объект – значит, указать класс, к которому он относится.

Задача диагностики в медицине – задача классификации. Врачу нужно найти правило, позволяющее по набору симптомов поставить диагноз, т.е. отнести объект к определенному классу заболеваний в системе классификации.

ВОЗ предписывает *создание международных классификаций здоровья* в целях обеспечения согласованной, значимой и применимой на практике основы для классифицирования, которую смогут использовать правительства стран мира и потребители медицинских услуг в качестве "единого языка для взаимодействия". Их использование позволит проводить сравнение данных в рамках одной или нескольких человеческих популяций, подготовку непротиворечивой национальной и международной статистики. Применение классификаций обеспечивает моментальное представление о состоянии здоровья популяции по таким параметрам как смертность, заболеваемость, социальное функционирование, инвалидность, здоровье, вмешательства, что гарантирует осведомленность руководящих органов и повышает эффективность принятых решений в области охраны общественного здоровья [6].

Семейство международных классификаций ВОЗ (WHO-FIC) состоит из *трех* следующих типов классификаций:

1. **Справочные классификации** (Reference Classifications): – основные классификации, описывающие наиболее существенные и основные параметры здоровья:

- *Международная классификация болезней (МКБ);*
- *Международная классификация функционирования, ограничений жизнедеятельности и здоровья (МКФ);*
- *Международная классификация медицинских услуг (ICHI);*

2. **Производные классификации** (Derived classifications) – основаны на справочных классификациях (т.е. МКБ и МКФ). Могут создаваться путем расширения структуры или дополнительной детализации информации, содержащейся в справочных классификациях, а также путем перестроения или агрегирования записей одной или нескольких справочных классификаций. К ним относятся:

- *Международная классификация болезней онкологии, 3-е издание (МКБ-О-3)*
- *МКБ-10 для психических и поведенческих расстройств - Клинические описания и руководство по диагностике;*
- *МКБ-10 для психических и поведенческих расстройств - диагностические показатели при проведении исследований;*
- *Применение Международной классификации болезней в неврологии (ICD-10-NA);*
- *Применение Международной классификации болезней в стоматологии и ортопедии, 3-е издание (ICD-DA);*

3. Родственные классификации (Related Classifications) – связаны или частично совпадают со справочными классификациями на определенных уровнях структуры. К ним относятся:

- *Международная классификация первичной медицинской помощи, 2-е издание (ICPC-2);*
- *Международная классификация внешних причин травм (ICECI);*
- *Классификация вспомогательных средств для лиц с ограниченными возможностями (ISO9999);*
- *Анатомическая, лечебная, фармацевтическая классификация с определенными суточными дозами (ATC/DDD);*
- *Международная классификация сестринской практики (ICNP);*

Код (франц. *code*, от лат. *codex* – свод законов) – уникальное алфавитно-цифровое обозначение объекта в классификации в соответствии с системой условных знаков (символов) для передачи, обработки и хранения различной информации.

Процесс преобразования сообщения в комбинацию символов в соответствии с кодом называется *кодированием*, процесс восстановления сообщения из комбинации символов называется *декодированием*.

Создание единого информационного пространства здравоохранения в Российской Федерации подразумевает интеграцию разрозненных информационных систем медицинских учреждений, фондов ОМС, органов управления здравоохранением в системы, организованные по отраслевому или территориальному принципу, и дальнейшую интеграцию региональных систем в единую систему федерального уровня. Особенностью единой информационной системы должен стать свободный обмен данными медицинской тематики, статистических и экономических показателей, данных отчетности субъектов системы, при этом система должна предусматривать защиту личной информации пациентов и соблюдение юридических норм информационного обмена участников системы. Для реализации этой задачи используют различные стандарты кодирования и классификации медицинской информации.

Международная статистическая классификация болезней и проблем, связанных со здоровьем (англ. International Statistical Classification of Diseases and Related Health Problems) – документ, используемый как ведущая статистическая и классификационная основа в здравоохранении. Периодически (раз в десять лет) пересматривается под руководством Всемирной организации здравоохранения (ВОЗ).

В настоящее время действует Международная классификация болезней Десятого пересмотра (МКБ-10, ICD-10).

В России органы и учреждения здравоохранения осуществили переход статистического учёта на МКБ-10 в 1999 году.

Целью МКБ является создание условий для систематизированной регистрации, анализа, интерпретации и сравнения данных о смертности и заболеваемости, полученных в разных странах или регионах и в разное время. МКБ используется для преобразования словесной формулировки диагнозов болезней и других проблем, связанных со здоровьем, в буквенно-цифровые коды, которые обеспечивают удобство хранения, извлечения и анализа данных.

Основой классификации МКБ-10 является трёхзначный код, который служит обязательным уровнем кодирования данных о смертности, которые отдельные страны предоставляют ВОЗ, а также при проведении основных международных сравнений.

Структура МКБ-10 разработана на основе классификации, предложенной Уильямом Фарром. Его схема заключалась в том, что для всех практических и эпидемиологических целей статистические данные о болезнях должны быть сгруппированы следующим образом:

- ✓ *эпидемические болезни;*
- ✓ *конституциональные или общие болезни;*
- ✓ *местные болезни, сгруппированные по анатомической локализации;*
- ✓ *болезни, связанные с развитием;*
- ✓ *травмы.*

МКБ-10 состоит из трёх томов:

- том 1 содержит основную классификацию;
- том 2 содержит инструкции по применению для пользователей МКБ;
- том 3 представляет собой Алфавитный указатель к классификации.

Классификация разделена на **22 класса** (табл. 2.1.). Первым знаком кода в МКБ-10 является буква. Четыре класса (I, II, XIX и XX) используют более одной буквы в первом знаке своих кодов.

Классы I—XVII относятся к заболеваниям и другим патологическим состояниям, класс XIX — к травмам, отравлениям и некоторым другим последствиям воздействия внешних факторов. Остальные классы охватывают ряд современных понятий, касающихся диагностических данных.

Таблица 2.1.

Список классов МКБ 10-го пересмотра

Литера	Класс	Заголовок
A	Класс I	Некоторые инфекционные и паразитарные болезни
B		
C	Класс II	Новообразования
D	Класс III	Болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм
E	Класс IV	Болезни эндокринной системы, расстройства питания и нарушения обмена веществ
F	Класс V	Психические расстройства и расстройства поведения
G	Класс VI	Болезни нервной системы
H	Класс VII	Болезни глаза и его придаточного аппарата
	Класс VIII	Болезни уха и сосцевидного отростка
I	Класс IX	Болезни системы кровообращения
J	Класс X	Болезни органов дыхания
K	Класс XI	Болезни органов пищеварения
L	Класс XII	Болезни кожи и подкожной клетчатки
M	Класс XIII	Болезни костно-мышечной системы и соединительной ткани
N	Класс XIV	Болезни мочеполовой системы
O	Класс XV	Беременность, роды и послеродовой период
P	Класс XVI	Отдельные состояния, возникающие в перинатальном периоде
Q	Класс XVII	Врожденные аномалии (пороки крови), деформации и хромосомные нарушения

R	Класс XVIII	Симптомы, признаки и отклонения от нормы, выявленные при клинических и лабораторных исследованиях, не классифицированные в других рубрика
S	Класс XIX	Травмы, отравления и некоторые другие последствия воздействия внешних причин
T		
V	Класс XX	Внешние причины заболеваемости и смертности
W		
X		
Y		
Z	Класс XXI	Факторы, влияющие на состояние здоровья населения и обращения в учреждения здравоохранения.
U	Класс XXII	Зарезервировано

Классы подразделяются на однородные «блоки» трёхзначных рубрик. В рамках каждого блока некоторые из трёхзначных рубрик предназначены только для одной болезни, отобранной вследствие её частоты, тяжести, восприимчивости к действиям служб здравоохранения, в то время как другие трёхзначные рубрики предназначены для групп болезней с некоторыми общими характеристиками. В блоке обычно имеются рубрики для «других» состояний, дающие возможность классифицировать большое число различных, но редко встречающихся состояний, а также «неуточнённые» состояния.

Большинство трёхзначных рубрик подразделены посредством четвёртого цифрового знака после десятичной точки, с тем чтобы можно было использовать ещё до 10 подрубрик. Если трёхзначная рубрика не подразделена, рекомендуется использовать букву «X» для заполнения места четвёртого знака, чтобы коды имели стандартный размер для статистической обработки данных.

Четырёхзначные подрубрики используют любым подходящим способом, определяя, например, различные локализации или разновидности одной болезни.

Коды U00—U49 следует использовать для временного обозначения новых болезней неясной этиологии. Они составляют XXII класс болезней [7].

SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) – это систематизированная машинно-обрабатываемая медицинская номенклатура. SNOMED CT является базовой терминологией, используемой для ведения электронных медицинских записей. Номенклатура SNOMED состоит из 11 связанных взаимными ссылками классификаций, называемых модулями (табл. 2.2.). Структура каждого модуля похожа на структуру классификации МКБ.

Таблица 2.2.

Литера	Модуль
T	топография — анатомические термины
M	морфология — клетки, ткани, органы
L	тип живого организма, например, бактерия, вирус
C	химия, лекарства
F	функции — признаки и симптомы
J	род занятий
D	диагностические термины
P	процедуры — административные, диагностические, терапевтические
A	физические агенты, деятельность, связанная с болезнью
S	социальные условия
G	синтаксические связи и определители

К примеру, диагностический термин «DE-14810 Лёгочный туберкулёз» содержит ссылки на термины «T-28000 Лёгкое (топография)», «L-21801 Бацилла туберкулёза человека (живые организмы)», а также ссылку на код МКБ 011.9.

SNOMED в первую очередь предназначена для формализации описания результатов клинических наблюдений. В качестве её логической модели можно предложить следующую формулу: «Для пациента из социальной среды S с профессией J (процедура P, применённая к месту T, показала/выявила морфологию M, живой организм L, химический,

лекарственный и биологический продукт С, функцию F, физическое воздействие А, заболевание или диагноз D) (с модификатором G)».

Учитывая описанные выше направленность, гибкость и интернациональный характер номенклатуры SNOMED, комитет по стандартизации передачи медицинских изображений DICOM выбрал её в качестве рекомендуемой системы кодирования передаваемых вместе с изображениями текстовых данных [8].

Клинические коды Рида RCC (Read Clinical Codes). Первая версия этих кодов была разработана Джеймсом Ридом в 80-х гг. и предназначалась для более точной и унифицированной регистрации на компьютере сведений о состоянии здоровья пациентов, обращавшихся за первичной медицинской помощью. С помощью этих кодов автоматически формировались эпикризы, обеспечивалось ведение полностью безбумажной истории болезни (в этом случае коды сочетаются со свободным текстом), выдавались стандартные отчеты о заболеваемости, предусмотренные органами управления здравоохранением, обеспечивалась выписка и повторение рецептов. В настоящее время последние версии охватывают 347 569 терминов, многие из которых составлены автоматически, путем комбинирования базовых терминов с дополнительными данными.

DICOM (англ. Digital Imaging and Communications in Medicine) - отраслевой стандарт создания, хранения, передачи и визуализации медицинских изображений и документов обследованных пациентов, поддерживается основными производителями медицинского оборудования и медицинского программного обеспечения. Стандарт позволяет организовать цифровую связь между различным диагностическим и терапевтическим оборудованием, использующимся в системах различных производителей.

Файловый уровень стандарта DICOM 3.0 описывает:

1. Атрибуты и демографические данные пациента.
2. Модель и фирму производителя аппарата, на котором проводилось обследование.
3. Атрибуты медицинского учреждения, где было проведено обследование.
4. Атрибуты персонала, проводившего обследование пациента.
5. Вид обследования и дата/время его проведения.
6. Условия и параметры проведения исследования пациента.
7. Параметры изображения или серии изображений, записанных в DICOM-файле.
8. Уникальные ключи идентификации Unique Identifier (UID) групп данных, описанных в DICOM-файле.
9. Изображение, серию или набор серий, полученных при обследовании пациента.
10. Представление, в первую очередь, PDF-документов в DICOM-файле.
11. Представление DICOM-записи на оптические носители, включая DVD формат.
12. DICOM-протокол для передачи/приема по TCP/IP компьютерным сетям.

HL7, Health Level 7 («Седьмой уровень») - стандарт обмена, управления и интеграции электронной медицинской информации. Поддерживает выполнение таких задач как:

- Структурирование передаваемых данных;
- Возможности проектирования систем;
- Достижение согласованности передач;
- Безопасность;
- Идентификация участников;
- Доступность.

Важным направлением деятельности в стандартизации является стандартизация передачи электрокардиограмм. В настоящее время переведен международный стандарт EN 1064:2005+A1 Health informatics - Standard communication protocol - Computer-assisted electrocardiography. Он необходим нашим отечественным производителям электрокардиографических систем [9].

Список литературы

1. Бородина А.И. Технологии баз данных и знаний. – Мн.: БГЭУ, 2008. – 505 с.
2. Вирт Н. Алгоритмы и структуры данных. – М.: Мир, 1989. – 135 с.
3. ГОСТ Р ИСО 8000-102-2011: Качество данных.
4. Классификация // Философский словарь / Под ред. И. Т. Фролова. – 4-е изд. – М.: Политиздат, 1981. – 445 с.
5. <https://ru.wikipedia.org/wiki/Классификация>
6. <http://who-fic.ru/>
7. <https://ru.wikipedia.org/МКБ-10>
8. http://medic.oplib.ru/zabolevaniya/view/32132_nomenklatura_snomed
9. Бокерия Л.А., Самородская И.В., Ступаков И.Н. Проблемы создания классификаторов и методик учета операций на сердце и сосудах // Вестник новых медицинских технологий, 2010. – Т. XVII, № 4. – С.88-92.

Контрольные вопросы

1. Что такое данные?
2. Как представляются данные в памяти компьютера?
3. Какие типы данных используются для хранения, обработки и передачи медицинской информации?
4. Что включает в себя обработка медицинских данных?
5. Что такое классификация?
6. Что такое кодирование?
7. Что такое SNOMED?
8. Для чего используются клинические коды Рида RCC?
9. Стандарты DICOM, HL7.