

## Сортировка и структурирование данных

Большинство собранных медицинских данных являются весьма неоднородными по характеру, имеют разнообразную форму и вид. Наименьшей единицей измерения данных является *байт* (он равняется 8 битам). Одним байтом, как правило, кодируется один символ текстовой информации. Два логически взаимосвязанных байта составляют *слово*. Условно 1000 байт составляет *1 Кбайт*, 1000 Кбайт = 1 Мбайт, 1000 Мбайт -- 1 Гбайт, 1000 Гбайт = 1 Тбайт.

При организации упорядоченных структур медицинских данных пользователь компьютерных технологий должен иметь представление об ориентировочном объеме этих данных. Так, одна страница машинописного текста может иметь объем в зависимости от используемого формата от 10 до 30 Кбайт. Значительный объем имеют медицинские диагностические изображения. В зависимости от размеров матрицы и ее строения изображение органов человека, например, на радионуклидной сцинтиграмме будет иметь объем несколько сот Кбайт, А вот уже цифровые рентгенограммы грудной клетки, кровеносных сосудов, молочных желез по своему объему могут достигать 5-10 Мбайт.

В качестве единицы хранения данных в информатике принят файл. *Файл – это определенное количество данных, хранящихся в памяти компьютера и имеющих уникальное собственное имя.* В медицинской информатике файлом могут являться страницы текста (истории болезни, отчеты, отдельные записи), визуальные картины (гистологический срез, ультразвуковая сканограмма, рентгенограмма). Файлами могут быть учебные или иллюстративные видеофильмы, звуковой ряд (музыка, речь, зарегистрированные тоны сердца, сосудистые шумы). Каждый файл имеет атрибуты, которые включают в себя имя файла и тип содержимого, дату и время создания, имя владельца, размер, права и метод доступа к нему. Для идентификации типа файла использую *расширения*.

Уникальность имени файла очевидна, ибо это обеспечивает единственно правильный путь к нему доступа. Хранение файлов осуществляется в иерархической структуре, которая называется *файловой структурой*. На вершине файловой структуры указывается имя магнитного носителя. Далее файлы группируются в *каталоги*, или *папки*. Внутри каждой папки (каталога) могут быть созданы *вложенные папки (каталоги)*. Характерным примером такой структуры в медицинской документации может служить электронная история болезни.

Полное имя файла включает в себя наряду с собственным именем файла также и путь доступа к нему. Например:

*C:\заболевание\инфекционные\легкие\пневмония.doc.*

В этом примере: «С» – имя носителя (диска), «заболевание» – адрес коренного каталога, «инфекционные\легкие» – вложенные в каталог папки, «пневмония» – имя файла (страницы), «doc» – расширение файла.

Существуют также другие расширения, которые определяет его тип: *.exe, .com* – программы, *.txt, .doc* – текстовые файлы (документы), *.bmp, .gif, .jpg* – графические файлы, *.avi* – видеофайлы. Необходимо добавить, что только расширения *.exe* и *.com* служат признаками исполняемости указанных файлов. Для работы с файлами, имеющими другие расширения, на компьютере пользователя должны быть установлены специальные программные служебные модули.

*Структурирование медицинских данных – это упорядочение массива данных по заранее заданному алгоритму.* Существует три основных типа структурирования данных: *линейный, табличный и иерархический*. Для каждого из вышеуказанного типа характерны свой принцип разделения данных и система адресации элементов.

Линейные, или списочные, структуры данных состоят из простого перечня элементов данных. Подобную структуру, например, имеет список больных, находящихся в отделении. Чаще всего этот список представляет собою журнал, где каждый пациент зарегистрирован построчно. При этом разделителем элементов является конец каждой строки. Например:

1. Сидоров П. Е.
2. Куликов И. С.
3. Семенов В. П. и т.д.

В таких случаях нахождение нужного элемента решается очень просто: он определяется по номеру строки, например, №2 – Куликов И.С.

Разделителем элементов может быть и какой-либо специальный символ, например, символ . Тогда линейная структура данных приобретет следующий вид: Сидоров П.Е. Куликов И.С. Семенов В.П и т.д. В такой структуре нужный элемент с номером  $n$  разыскивается по формуле  $n-1$ , где  $n$  – число разделителей, начиная с первого, зарегистрированного в строке. Заканчивается искомым элементом тогда, когда будет зарегистрирован следующий за искомым элементом разделитель.

*Табличные, или матричные, структуры данных* представляют собою таблицы, разбитые на ячейки горизонтальными и вертикальными линиями. Совокупность ячеек по горизонтали называется строкой, по вертикали – столбцом. Каждая ячейка содержит отдельный элемент данных (ячейка может быть и пустой). Пример такой структуры данных приведен в табл. 1.3.

Таблица 1.3

**Табличная (матричная) структура данных**

Ф.И.О. пациента	Диагноз заболевания	Температура тела, °С	Рост пациента, см	Масса тела пациента, кг
Сидоров П.Е.	Холецистит	37,7	171	87,3
Куликов И.С.	Панкреатит	38,4	168	66,7
Семенов В.П.	Пневмония	39,1	180	96,6

Другим примером табличной структуризации данных является офисное приложение *MS Excel*, которое часто используется при обработке медицинской информации.

Адресация элементов в таблицах осуществляется двумя параметрами – номерами (или буквенными указателями) столбца и строки, например, А,7; В,5. В медицинской практике иногда приходится иметь дело с таблицами, все ячейки которых имеют одинаковую длину. Такие таблицы называются *матрицами*. С матрицами сталкиваются, например, при тестировании биосубстратов, определении диагностической эффективности различ-

ных методов исследования. Недостатком линейной и табличной структур данных является неудобство их развития. При добавлении какого-либо нового элемента внутри этих структур будет изменяться адресация всех других элементов.

Ниже приведен пример (рис.1.2) *иерархической структуры данных*. Она используется при сложной зависимости отдельных элементов, нерегулярности данных, а также при выполнении классификационных схем. Этот принцип лежит также в основе построения файловой структуры хранения данных в компьютерах.



Рис. 1.2. Иерархическая структура данных

Адрес элемента в иерархической структуре определяется путем доступа (маршрута), который начинается у вершины структуры и оканчивается в искомом элементе. Так, например, в приведенной схеме доступ к митральным порокам сердца осуществляется по маршруту: *заболевания сердца* → *приобретенные пороки* → *митральные пороки*.

Иногда медицинские данные могут представляться более упрощенно – в *дихотомическом виде* (рис.1.3). В дихотомической схеме построения данных путь доступа к искомому элементу указывать проще и короче. Его можно обозначить как серию поворотов маршрута налево (0) или направо (1) и отобразить в виде двоичной записи. Например, в указанной схеме митральные пороки у взрослых могут быть найдены по маршруту *01*.



Рис.1.3. Дихотомическая система группировки данных

Преимуществом иерархической и дихотомической структур данных, по сравнению с табличными и списочными методами, является более простой способ развития схемы: при этом необходимо лишь удлинить запись маршрута.