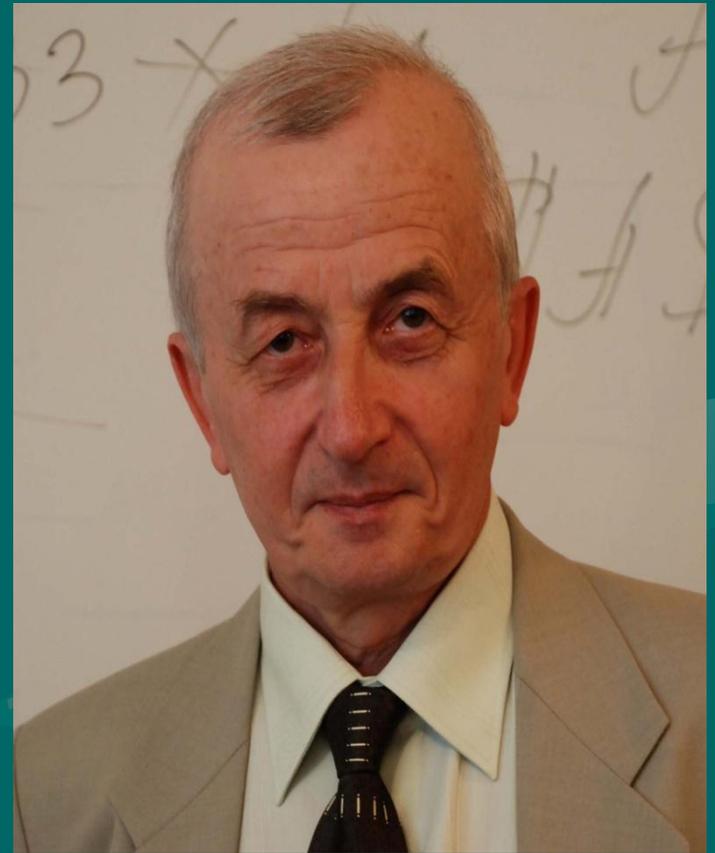


# Информатика, информация и её классификация

Косьмин Сергей  
Николаевич



© КсН, 2011-2016

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

Наука – это сфера человеческой деятельности,  
часть общественного сознания, функцией  
которой является выработка и теоретическая  
систематизация объективных знаний о реальной  
действительности.

**НАУКИ** (по предмету исследования)

Естественные

Общественные

Технические

Лингвистические

**НАУКИ** (по отношению к практике)  
подразделяются на: прикладные и фундаментальные.

**Прикладные науки** обеспечивают знаниями, необходимыми для эффективной деятельности в ограниченной прикладной области.

Науки, которые на основе накопленных знаний строят общую картину мира, называются **фундаментальными науками**.

**Информатика** – это естественная, фундаментальная наука, изучающая структуру и общие свойства информации, общие закономерности информационных процессов для построения машин обработки данных и информационных технологий.

*Предметом информатики является информация и общие закономерности информационных процессов.*

*Методом исследования мира, характерным для информатики, является диалектика, в своих двух частях: исторический (общество) и диалектический (природа) материализм.*

Метод науки прямо указывает на позицию исследователя и на то, каким образом добываются знания о мире.

Развитие определяется  
законами природы.

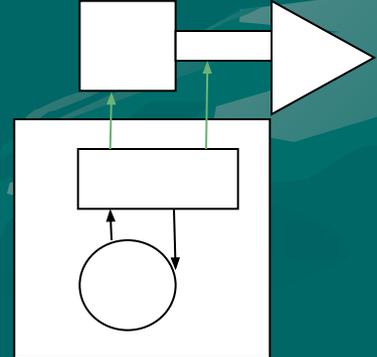
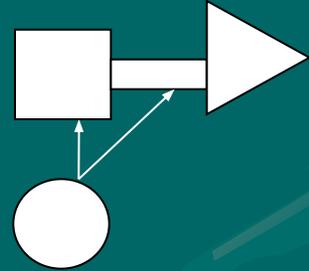
Развитие определяется  
законами природы и законами общества.

0

1789-94  
ГОДЫ

90-ые  
ГОДЫ XX века

time



Система: “человек-машина”

Эра развития природы.

Наша эра (эра людей).

Первобытно-  
общинный  
строй

Рабовладель-  
ческий  
строй

Феодализм

Капитализм,  
Социализм

Комму-  
низм

Эволюционные этапы становления человека: из мира животных в мир людей.

Закономерности развития четвёртого звена рабочей машины и знания, необходимые для эффективного управления этим звеном четырёхзвенной рабочей машины накапливает наука, зародившаяся в 60–е годы XX века.

В мире существуют две научные школы, выполняющие такую работу:

- франко-итальянская (европейская) школа накапливает знания в науке **информатика**,
- американская школа — копит знания в **компьютерных науках**. В американской научной школе нет науки информатика[1].

[1] Брукшир, Дж., Гленн. Введение в компьютерные науки. Общий обзор, 6-е издание. : Пер. с англ. — М.: Издательский дом “Вильямс”, 2001. — 688 с.

Категориями [1] называются неформальные положения, не связывающие мышления стандартами формами и правилами, и тем самым открывающие путь свободному творчеству.

[1] Копнин П.В. Диалектика как логика и теория познания. М.: Наука, 1973, стр. 27, 92, 95.

Через взаимодействие категорий и других понятий методом определения в категориальный аппарат науки вводятся **ПОНЯТИЯ**.

Совокупность категорий и понятий, накопленная исследователями к данному моменту времени, составляет **категориальный аппарат науки**.

# Информация

*Материя      Структура      Система      Отражение*

*Материя и информация* - отдельные сущности мира.

**Материя**[1] – это объективная реальность, данная нам в ощущениях.

[1] Ленин В.И. Полн. собр. Соч., 5 изд., т.18, с. 131

Информация - не материя, но она о материализована через свой носитель.

# Информация

*Материя*

*Структура*

*Система*

*Отражение*

*Структура и информация*

- отдельные сущности мира.

Пример

Кристаллическая решетка – это структура кристалла, а то, как устроена кристаллическая решетка – это информация.

# Информация

*Материя*

*Структура*

*Система*

*Отражение*

*Система и информация*

- отдельные сущности мира.

**Система** – это совокупность элементов и связей между ними, обладающих системным качеством (свойством эмерджентности) таким, какое есть у системы в целом и отсутствует у любого из её элементов. (Людвиг фон Берталанфи)

**Информация** – это не система, но вне системы она существовать не может.

# Информация

*Материя      Структура      Система      Отражение*

*Отражение и информация* - отдельные сущности мира.

**Отражение** – это процесс изменения одного объекта или явления под воздействием другого.

В.И. Ленин. Ленинская теория отражения  
и современность, София, 1969.

**Информация** – компонента (составная часть) процесса отражения, выражающая его содержание.

*Материя      Структура      Система      Отражение      Информация*

Информация – это опосредованный формами связи результат отражения изменяющегося объекта изменяемым объектом с целью сохранения их системой целостности.

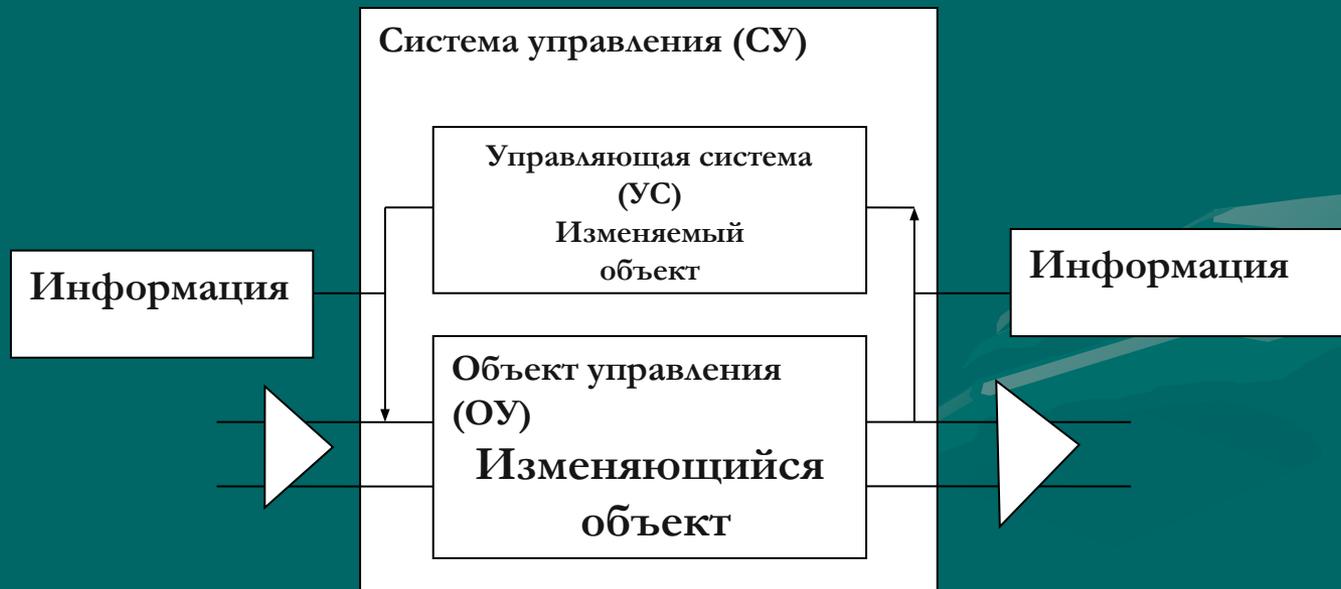


Иллюстрация описания (портрет) информации

Природа информации двойственна. Информация материальна по форме, и идеальна по содержанию.

Информация, носителями которой, с позиции внешнего наблюдателя системы, являются её изменяющийся и изменяемый объекты, называется связанной (в системе).

Связанная информация существует в системах молекулярного уровня.

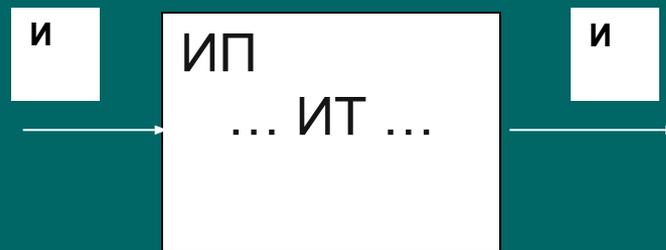
Информация, носитель которой способен существовать отдельно от изменяющегося и изменяемого объектов системы, называется выделенной.

Выделенная информация получила название — данные.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

Информационным процессом (ИП) называется процесс (явление, протекающее во времени) предметом и продуктом труда которого является информация (И).



- **Информационная технология (ИТ)** – это поэтапное, пооперационное описание информационного процесса, преобразующего входную информацию в выходную, с гарантированным качеством.
- **Новая информационная технология** – это технология, построенная на основе машин для обработки информации.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

Классификация – результат упорядоченного распределения множества объектов классификации (предметов и атрибутов) в соответствии со значениями их признаков (признаков классификации) и выбранным основанием деления (критерием классификации).

Основные принципы классификации:

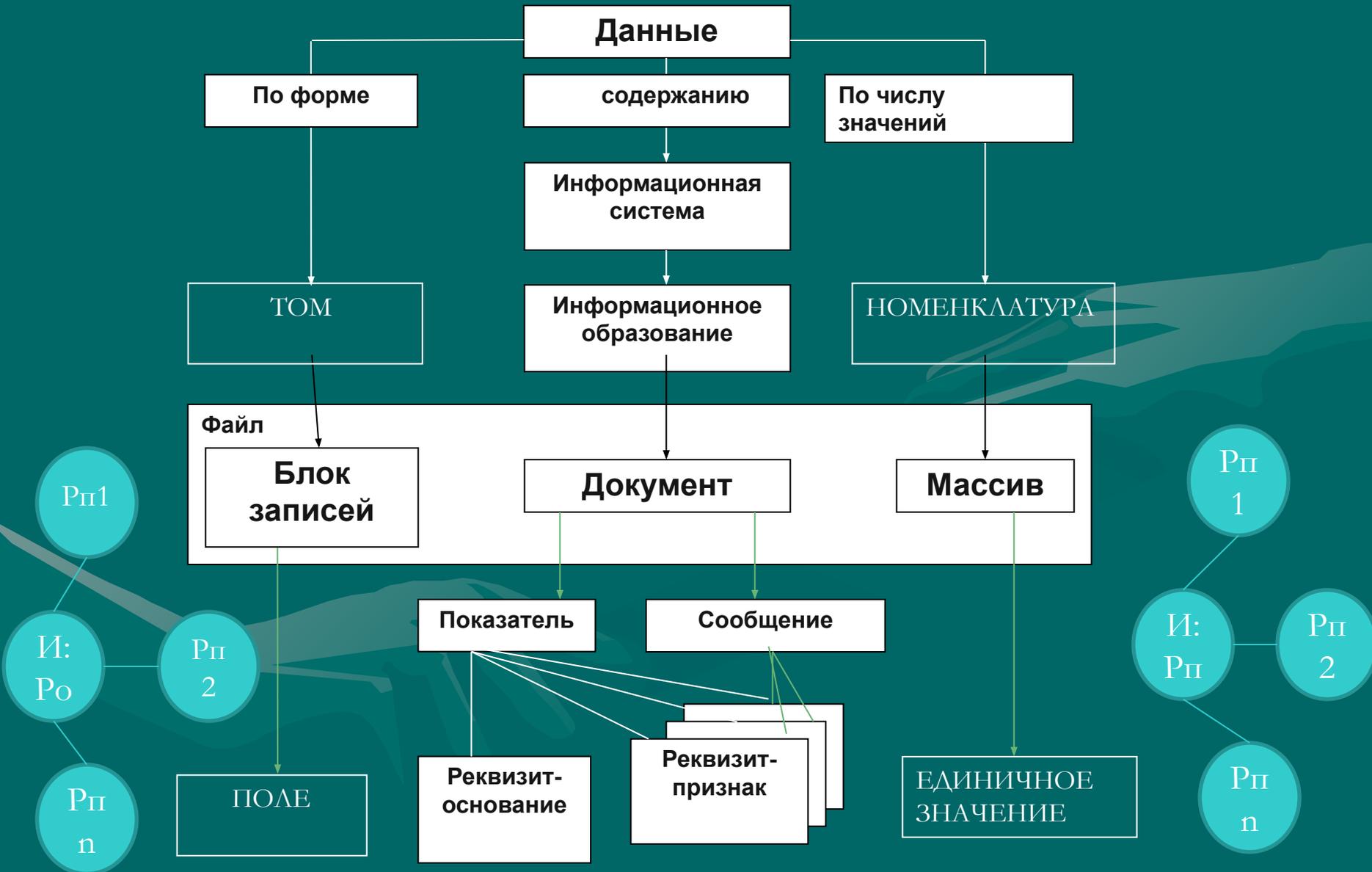
- *Поэтапность классификации* – заключается в пошаговом разбиении исследуемой сущности.
- Принцип *единственности основания деления на этапе* утверждает, что разбиение исходного множества на подмножества должно производиться на основе единого критерия.
- Принцип *исключающего деления классифицируемого множества* утверждает обратный переход от части к целому происходит без потери объектов классификации.

Информация многогранна и описывается множеством классификаций.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

# Основные классификации данных



Классификация данных по содержанию предложена Михаилом Антоновичем Королевым и используется при описании *логической структуры данных*.

Мера классификации: **Информационная совокупность** – это группа данных, обладающая общей формой и характеризующая одну сущность реального мира.

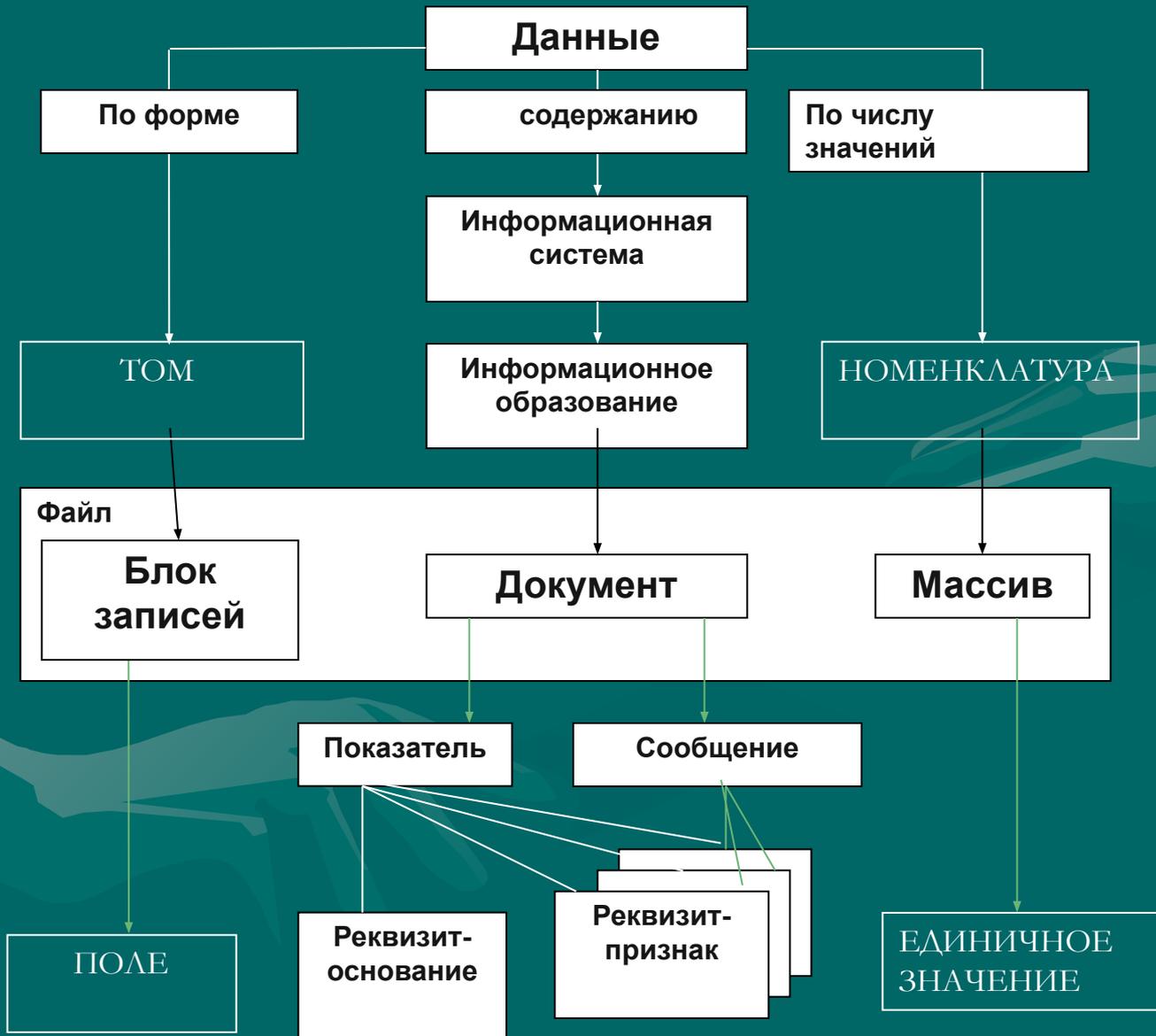
- Минимальная по объему, не расчленяемая без потери смысла информационная совокупность называется **реквизитом**.
- **Реквизиты-основания** характеризуют идентифицируемую сущность мира с количественной стороны, а **реквизиты-признаки** – с качественной.

- **Показатель** несет количественную информацию, определяемую рядом реквизитов—признаков в информационном пространстве.
- **Сообщение** несет качественную информацию, определяемую рядом реквизитов—признаков в информационном пространстве.
- **Документ** – это информационная совокупность предназначенная для подготовки или принятия решения.
- Информационная совокупность, охватывающая документы конкретной функции управления, называется **информационным образованием**.
- Информационная совокупность, охватывающая данные определённой системы управления, называется **информационной системой**.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

# Основные классификации данных



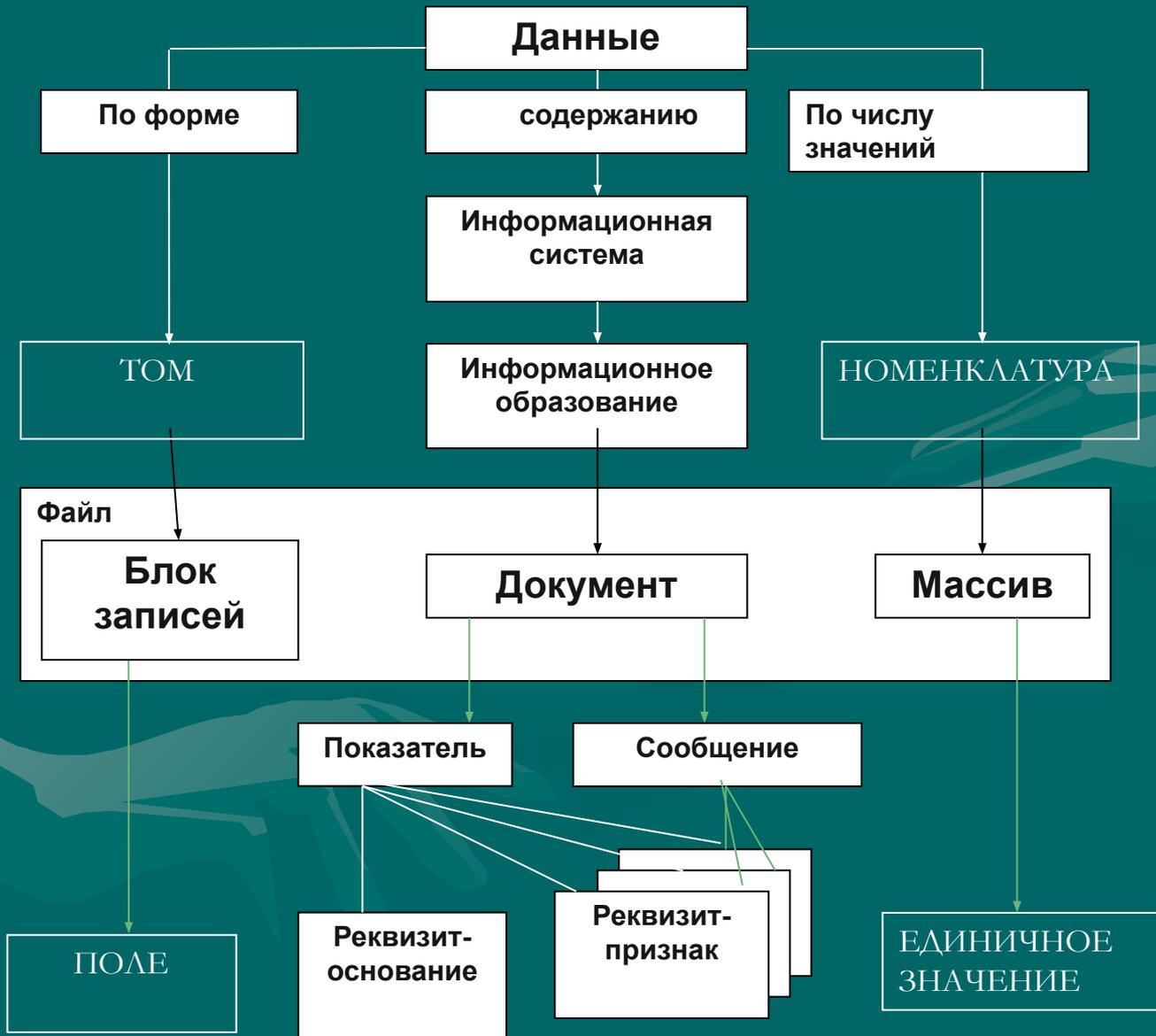
Классификация данных по числу значений в информационной совокупности предложена Михаилом Антоновичем Королёвым.

- **Единичное значение** – это информационная совокупность из одного элемента, то есть единичное множество.
- **Массив** – это информационная совокупность, состоящая из множества однотипных элементов.
- **Номенклатура** – это информационная совокупность, состоящая из единичных значений и массивов.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

# Основные классификации данных



Классификацию данных по форме (Рис. 3.) предложил Джеймс Мартин[1] – учёный из института системных исследований фирмы IBM. Данная классификация используется при описании *физической структуры данных*.

[1] Дж. Мартин. Организация баз данных в вычислительных системах. Изд. 2-е, доп. М.: Мир, 1980.- 662 с.

- **Поле** – это часть носителя для регистрации простой единицы информации.
- **Запись или блок записей** – это часть внешней памяти одновременно доступная одной макрокомандой ввода – вывода.
- **Том** – это сменный носитель информации.
- Информационный объект по форме представляющий собой запись, по содержанию - документ, а по числу значений массив называется **файлом**.
- **Файл** – это основная единица хранения данных во внешней памяти.

# Информатика, информация и её классификация

- Информатика, её предмет и метод
- Основные понятия информатики
- Классификация информации
- Классификация данных по содержанию
- Классификация данных по числу значений
- Классификация данных по форме
- Измерение количества информации

Самый простой канал связи может быть построен на наличии или отсутствии сигнала. Сигнал – это одна из форм отражения, способная переносить информацию. Если события “наличие сигнала” и “отсутствие сигнала” распределены равномерно, то в канале связи наблюдается только несущая. Информация в канале связи отсутствует. Канал готов к передаче данных.

Символы сообщения



Информационный процесс канала связи

- Если передача производится поочерёдно то в одном, то в другом направлении, связь источника с потребителем называется **полудуплексной**.
- Если передача производится одновременно в обоих направлениях, то связь источника с потребителем называется **дуплексной**.
- Какой бы ни была связь, в канале связи символы сообщения представляются двоичным кодом.
- Любое сообщение выражается словами определённого языка. Язык имеет конечное множество символов, называемых **алфавитом языка**.
- Алфавит можно перенумеровать. Если нумерация производится в двоичной системе счисления, то в результате каждый символ алфавита получает двоичный порядковый код.

В 1928 году английский инженер Роберт Хартли предложил длину кодового обозначения символа алфавита, то есть число разрядов двоичного числа, использовать в качестве меры количества информации в сообщении  $I$ , передаваемом по каналу связи. Эта мера количества информации стала называться логарифмической, так как логарифм выражает количество разрядов числа:

$$I = \log_2 V = k, \quad (1)$$

где  $V$  - двоичное число из  $k$  разрядов, представляющее собой код последнего символа алфавита (самый длинный).

Мера Хартли характеризует источник информации ФОРМАЛЬНО: через длину его алфавита.

Вероятностная мера количества информации в сообщении, передаваемом по каналу связи, введённая Клодом Шенноном в 1946 году, получила название энтропия источника информации.

- Клод Шеннон предложил рассматривать процесс появления определённого информационного символа  $a_i$  в канале связи как случайное событие  $A$  генерации символа источником информации.

Код формируется поразрядно. Разряды определяются независимо друг от друга. Следовательно, по теореме об умножении событий вероятность генерации кода символа источником:

- $$P(a_i) = \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \dots * \frac{1}{2} = 2^{-k} \quad \text{или}$$

умноженное  $k$  раз

- $$k = -\log_2 P(a_i) \quad (2)$$

Всё сообщение представляется статистическим рядом, члены которого являются символами сообщения, каждый из которых в канале связи появляется с определённой вероятностью  $P(ai)$ .

В качестве меры количества информации, приходящейся на символ сообщения Шеннон предложил среднее взвешенное по вероятности логарифмической меры Хартли, рассчитанной по несущей:

$$H = \sum_{i=1}^m P(ai) * k = - \sum_{i=1}^m P(ai) * \log_2 P(ai) \quad (3)$$

- Если известна энтропия источника информации  $H$  и длина сообщения  $L$  в символах (байтах), то количество информации этого сообщения определится по формуле:

$$Q = H * L \quad (4)$$

- Количество информации сообщения позволяет рассчитать размер оперативной памяти машины, необходимой для хранения данных сообщения, если машина работает в двоичной системе счисления, то есть способна чётко различать одно из двух состояний.

**Мера количества информации в сообщении  
Клода Шеннона учитывает ФОРМУ и  
СОДЕРЖАНИЕ ИНФОРМАЦИИ**

© KcH, 2011-2016

