

# ЛЕКЦИЯ 2

---

## Язык как способ представления информации

1. **Информация**
2. **Естественный язык**
3. **Искусственный язык**
4. **Отличия естественных и искусственных языков**

# **Язык как способ представления информации**

**Информация**

# Определение

---

**Информация** – одно из базовых понятий в науке (как *материя, энергия*), поэтому сложно дать четкое определение:

- невозможно выразить через более простые понятия;
- объясняется только на примерах или в сравнении с другими понятиями.

*«Информация есть информация,  
а не материя и не энергия».*

*Н. Винер, «Кибернетика, или Управление  
и связь в животном и машине»*

# Виды информации

---

## По форме представления:

- текстовая – передаваемая в виде символов, предназначенных для обозначения лексем языка;
- числовая – передаваемая в виде цифр и знаков, обозначающих математические действия;
- графическая – передаваемая в виде изображений, графиков и т.п.;
- звуковая – устная или в виде записи для передачи лексем языка аудиальным путём;
- видеоинформация – передаваемая в виде видеозаписи.

# Свойства информации

---

Информация должна быть:

- **объективной** (не зависящей от чьего-либо мнения)  
~~«На улице тепло»~~, «На улице +28°C».
- **понятной** (выраженной на языке, понятном тому, для кого она предназначена)
- **полезной** (получатель решает свои задачи)
- **достоверной** (полученной без искажений)  
~~дезинформация, помехи, слухи, байки~~
- **актуальной** (ценной в данный момент)
- **полной** (достаточной для принятия правильного решения или понимания)  
~~«Концерт будет вечером»~~, «Концерт состоится 31 августа 2015 года в 17.00 в актовом зале Дворца студентов НТУ «ХПИ»»

# Информационные процессы

---

## 1. Создание

## 2. Хранение

- мозг, бумага, камень, береста...
- память ПК, дискеты, винчестеры, CD, DVD...

## 3. Обработка

- **кодирование** – запись информации в виде кода с помощью некоторой знаковой системы (языка)
- **поиск**
- **сортировка**

## 3. Передача

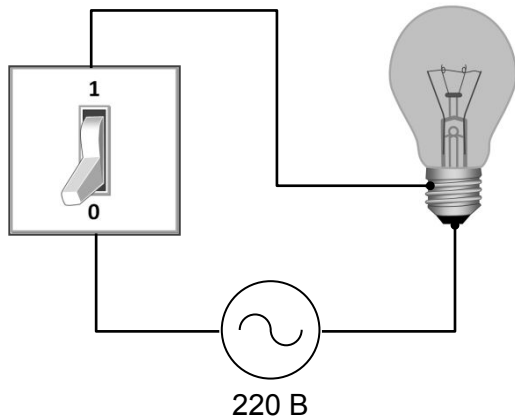


**Информация** – это содержание сигналов, передаваемых по каналам связи.

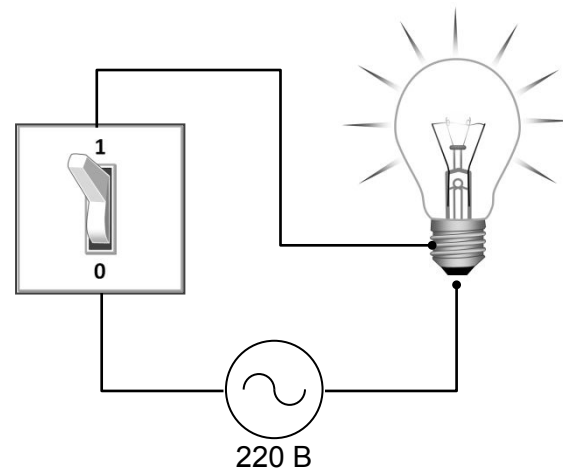
# Какой код использовать?



**Идея:** использовать тот код, который применяется в электронных устройствах



«0»



«1»

**Информация** – это сигнал (1, 0) или элементарное событие «да» – «нет», измеряемое в двоичных единицах – битах (мера Хартли).

# Двоичное кодирование

**Двоичное кодирование** – это представление всех видов информации с помощью последовательностей двух знаков – 0 и 1.



- в такой форме можно закодировать **все виды** информации;
- нужны только устройства с **двумя состояниями**;
- почти **нет ошибок** при передаче данных;
- **компьютеру легче** обрабатывать данные.



**человеку сложно** воспринимать двоичные коды



# Двоичный код

---

Код, в котором используются только два знака (0 и 1), называется **двоичным**.

Знаки 0 и 1 = *двоичная цифра*, англ. **binary digit** = *bit*

**1 бит** – это количество информации, которое можно передать с помощью 1 знака в двоичном коде («0» или «1»).

**1 байт** – это количество информации, которое можно передать с помощью 8 знаков двоичного кода.

**Слово** – это группа из 16 взаимосвязанных бит (2 взаимосвязанных байтов).

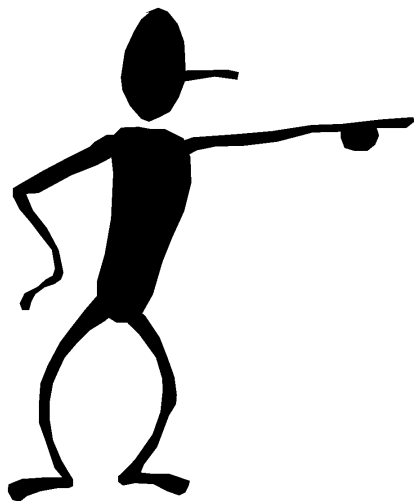
**Удвоенное слово** – это группа из 32 взаимосвязанных бит (4 взаимосвязанных байтов).

**Учетверенное слово** – это группа из 64 взаимосвязанных бит (8 взаимосвязанных байтов).

# Сколько информации?

---

Определите количество информации:



**01**

2 бита

**10101**

5 бит

**10101011**

8 бит, 1 байт

**1010101001100111**

16 бит, слово



# Единицы измерения

---

**1 байт (*byte*)** – это объем компьютерной памяти, который имеет индивидуальный адрес.

**1 байт (*byte*) = 8 бит**

**1 Кбайт (килобайт) = 1024 байта**

**1 Мбайт (мегабайт) = 1024 Кбайт**

**1 Гбайт (гигабайт) = 1024 Мбайт**

**1 Тбайт (терабайт) = 1024 Гбайт**

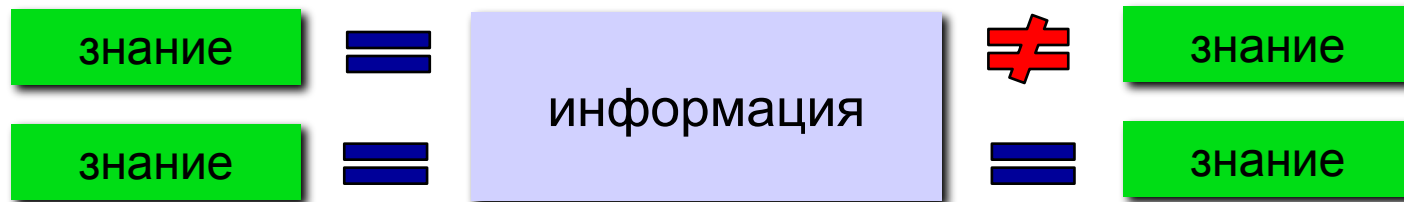
**1 Пбайт (петабайт) = 1024 Тбайт**

# Информация, знания, данные

---

**Информация** – одно из базовых понятий в науке, поэтому нет четкого единственно правильного определения.

**Знания** – воспринятая, тщательно отобранная, проанализированная информация, хранящаяся и используемая индивидуумом в его практической деятельности.



**Данные** – это любая, отобранная по определенному критерию и определенным образом упорядоченная информация, позволяющая решать определенный круг задач.

# Информационное общество

---

## Информационное (постиндустриальное)

**общество** – главными продуктами производства являются информация и знания.

### Этапы развития:

- **письменность** (Египет, Междуречье, Китай, 3000 лет до н.э.);
- **книгопечатание** (И. Гутенберг, XV век);
- **средства связи** (телеграф, телефон, радио, TV, конец XIX века);
- **компьютерная техника** (XX век).

**Информатизация** – процесс перехода к информационному обществу.

# Информационное общество

---

## Критерии:

- число компьютеров;
- уровень развития компьютерных сетей;
- доля населения, занятого в сфере ИТ.

## Характерные черты:

- ✓ изменение структуры экономики;
- ✓ развитие телекоммуникаций;
- ✓ свобода доступа к информации;
- ✓ рост информационной культуры (потребность и умение человека использовать ИТ);
- ✓ доступность образования;
- ✓ изменение уклада жизни.



**Ближе всего** США, Япония, Германия

# Информационное общество

---



- усиление влияния средств массовой информации;
- разрушение частной жизни личности и организаций;
- проблема отбора качественной и достоверной информации;
- многим людям будет сложно приспособиться к новым условиям.

# Информационная культура

---

## Информационная культура – это:

- совокупность знаний, умений та навыков поиска, отбора, анализа информации в соответствии с поставленной задачей;
- деятельность, направленная на удовлетворение информационных потребностей.



# Информационная культура

---

Составляющие:

- поисковая культура – поиск информации;
- компьютерная культура – использование программного обеспечения (ПО) различных уровней, знание теоретических основ функционирования компьютеров;
- деятельностьная культура – создание нового ПО, а также выбор и использование существующего ПО для решения поставленных задач.

# Информационные ресурсы

---

**Ресурс** – средство, «обработав» которое можно получить желаемый результат.

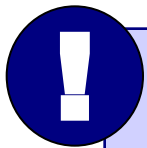
## Ресурсы общества:

- материальные;
- природные (сырьевые);
- энергетические;
- трудовые;
- финансовые;
- **информационные** – документы, хранящиеся в различных информационных системах, в том числе в библиотеках, архивах, фондах, банках данных и др.

# Информационные ресурсы

---

- библиотечные;
- архивные;
- научно-техническая информация;
- правовая информация;
- финансовая и экономическая информация;
- информация о природных ресурсах;
- информация государственных органов;
- отраслевая информация;
- информация предприятий и учреждений.



**Информационные ресурсы становятся **товаром**, стоимость которого сопоставима со стоимостью других ресурсов!**

# Информационные услуги

---

- СВЯЗЬ;
- Интернет;
- рекламные агентства;
- подбор персонала (кадровые агентства);
- обучение;
- консалтинг (услуги по оптимизации бизнеса);
- поиск и подбор информации;
- создание программного обеспечения;
- создание web-сайтов и др.

# Информационные технологии

---

**Технология** – способ сделать «продукт» из исходных материалов (с гарантированным результатом!).

**Новые (компьютерные) информационные технологии** – это технологии, связанные с использованием компьютеров и программного обеспечения для осуществления информационных процессов, то есть создания, хранения, обработки и передачи информации.

# Информационные технологии

---

- ❑ подготовка документов;
- ❑ поиск информации;
- ❑ сфера телекоммуникаций (локальные сети, Интернет, электронная почта и т.п.);
- ❑ автоматизированные системы управления (АСУ);
- ❑ системы автоматизированного проектирования (САПР);
- ❑ геоинформационные системы;
- ❑ обучение:
  - электронные учебники;
  - компьютерные тренажеры;
  - дистанционное обучение и т.д.

# **Язык как способ представления информации**

**Естественный язык**

# Определения

---

Для представления информации используется специальная знаковая система – **язык**.

**Алфавит** – конечный набор знаков (символов) любой природы, из которых конструируются сообщения данного языка.

**Слово** – последовательность символов алфавита, кодирующая информацию.

**Грамматика** – правила, по которым из символов алфавита строятся слова.

**Синтаксис** – правила, по которым из слов строятся предложения и более крупные конструкции языка.

**Семантика** – определение смыслового значения отдельных конструкций языка.

**Прагматика** – использование отдельных конструкций языка конкретными людьми в конкретной ситуации.



# Языки

---

## ■ естественные

Любой **естественный язык** – это сложная **многоуровневая** система знаков, возникшая для обмена информацией между людьми, выработанная в процессе практической деятельности человека, и постоянно изменяющаяся в связи с этой деятельностью;

## ■ искусственные

Любой **искусственный язык** – это знаковая система, создаваемая людьми для использования в тех областях науки и техники, где применение естественного языка ограничено, менее эффективно или невозможно.

# Уровни естественного языка

---

## ❖ **Фонологический** или **уровень фонем**.

**Фонемы** – отдельные звуки, с помощью которых формируются и различаются слова.

Фонологический уровень выделяется для устной речи, для письменных текстов в языках с алфавитным способом записи он соответствует **уровню символов (графематическому уровню)**, т.к. фонемы примерно соответствуют буквам алфавита.

# Уровни естественного языка

---

- ❖ **Морфологический уровень** или **уровень слов (словоформ)**.

**Словоформы** – слова в определенной грамматической форме.

Морфологический уровень включает **подуровень морфем**.


**Морфема** – это минимальная значащая часть слова (приставка, корень, суффикс, окончание).

# Уровни естественного языка

---

- ❖ Как отдельный может быть выделен **лексический уровень** или **уровень лексем**.

**Лексема** – это слово как совокупность всех его конкретных грамматических форм (например, лексему **стол** образуют формы стол, стола, столу, столом).

 В тексте встречаются словоформы, а в словаре – лексемы, точнее, в словаре записывается **лемма** – каноническая словоформа лексемы.

# Уровни естественного языка

---

- ❖ **Синтаксический уровень** или **уровень предложений** (высказываний).

Синтаксический уровень включает:

- **подуровень словосочетаний**;
- **надуровень сложного синтаксического целого**, которому примерно соответствует абзац текста.

# Уровни естественного языка

---

**Словосочетание** – это синтаксически связанная группа слов.

**Сложное синтаксическое целое, или сверхфразовое единство** – это последовательность предложений (высказываний), объединенных смыслом и лексико-грамматическими средствами.

К лексико-грамматическим средствам относятся:

- лексические повторы;
- **анафорические ссылки** – это ссылки на предшествующие слова текста, реализуемые при помощи местоимений и местоименных слов.

# Уровни естественного языка

---

- ❖ Также можно выделить **уровень дискурса**.

**Дискурс** – это

- последовательность взаимосвязанных друг с другом предложений текста, обладающая определенной смысловой целостностью, за счет чего она выполняет определенную прагматическую задачу;
- связный текст в его коммуникативной направленности.

# Уровни естественного языка

---

- ❖ Отдельным является вопрос об **уровне семантики**.

Подтверждением самостоятельности уровня семантики считается то, что человек обычно запоминает смысл высказывания, а не его конкретную языковую форму.

**Семы** – элементарные единицы уровня семантики, при помощи которых можно выразить смысл любого высказывания.

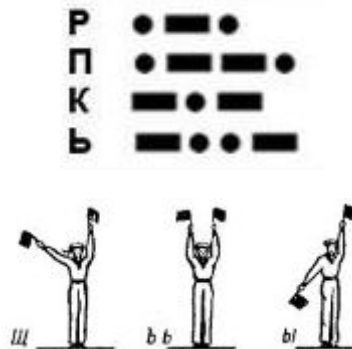
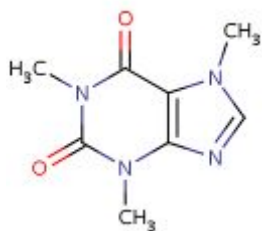
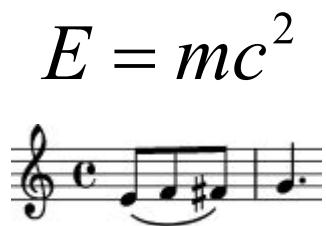


# **Язык как способ представления информации**

**Искусственный язык**

# Классификация искусственных языков

1. **Неспециализированные языки общего назначения** (например, эсперанто, волапюк и др.).
2. **Специализированные языки различного назначения:**
  - символические языки наук;
  - **языки программирования;**
  - языки операционных систем;
  - языки информационных систем и др.



$$16 = 10_{16} = 20_8 = 10000_2$$

```
program first;  
begin  
  writeln("Hello, world!");  
end.
```

# Языки программирования

---

**Языки программирования** – это:

- класс искусственных языков, предназначенных для обработки информации с помощью компьютера;
- строгая (формальная) знаковая система, при помощи которой записываются компьютерные программы.

**Программа** – это структурно строгий текст, записанный по формально заданным правилам языка программирования.

# Уровни языка программирования

---

- ❖ **Символьный уровень** – элементы алфавита, например, буквы, спецсимволы (в естественном языке – *графематический уровень*).
- ❖ **Уровень имен**, например, зарезервированных слов, выражений и т.п. (в естественном языке – *лексический уровень*).
- ❖ **Операторный (командный) уровень**, где синтаксические конструкции имеют повелительный характер (в естественном языке – *синтаксический уровень*).
- ❖ **Уровень программы**, всегда являющейся синтаксически и семантически законченной последовательностью предписаний-команд (в естественном языке – *семантический уровень*).

# **Язык как способ представления информации**

**Отличия естественных и  
искусственных языков**

# Отличия естественных и искусственных языков

---

- **Естественный язык** – продукт долгого исторического, и в определенной степени стихийного развития, а **искусственные языки** – это результат целенаправленной деятельности человека.
- **Открытость системы естественного языка**: язык постоянно изменяется. Следствие открытости – принципиальная невозможность один раз описать конкретный естественный язык и построить соответствующий лингвистический процессор. Необходимо пополнение знаний о языке на всех его уровнях.

**Лингвистический процессор** – это программа (или часть прикладной системы) для обработки текстов на естественном языке, использующая ту или иную формальную модель языка.

# Отличия естественных и искусственных языков

---

- ❑ **Характерной особенностью искусственных языков является однозначная определенность их словаря, правил образования выражений и правил их интерпретации и использования.**
- ❑ **Нестандартная сочетаемость единиц на каждом уровне естественного языка.** В частности, если в искусственных языках синтаксическая сочетаемость знаков определяется их семантикой, то в естественном языке соединение слов на уровне предложений лишь частично может быть описана законами грамматики. В любом языке достаточно большое количество грамматически правильных сочетаний реально не употребляется.

# Отличия естественных и искусственных языков

---

- **Большая системность естественного языка**, т.е. в нем больше уровней, четче границы между ними, а также более выражена асимметрия.

**Асимметрия** – нарушения регулярности связей между единицами языка и выражаемыми ими смыслами, проявляющиеся на всех уровнях языковой системы.

Асимметрия выражается в таких явлениях как:

- **полисемия (многозначность)** – наличие у одной единицы языка нескольких связанных между собой значений;
- **синонимия** – полное или частичное совпадение значений разных единиц;
- **омонимия** – совпадение по форме двух разных по смыслу единиц.



# Виды омонимии

---

- **Лексическая омонимия** означает одинаково звучащие и пишущиеся слова, не имеющие общих элементов смысла.

*Например, слово **ключ** может означать «родник» и «предмет, открывающий замок»; слово **каток** в предложении «Я гонялся за другом на катке» может означать «специальную дорожную машину для уплотнения грунта, асфальта и т.д.» и «гладкую ледяную поверхность для катания на коньках».*

- **Морфологическая омонимия** – совпадение форм одного и того же слова (лексемы).

*Например, словоформа **круг** соответствует именительному и винительному падежам единственного числа.*

# Виды омонимии

---

- **Лексико-морфологическая омонимия** (наиболее частый вид) возникает при совпадении словоформ двух разных лексем.

*Например, слово **стих** может быть глаголом в единственном числе мужского рода и существительным в единственном числе, именительном падеже; слово **души** в предложении «Души прекрасные порывы» может быть существительным в единственном числе родительном падеже и глаголом повелительного наклонения.*

- **Синтаксическая омонимия** означает неоднозначность синтаксической структуры, что приводит к нескольким интерпретациям.

*Например, Студенты из Львова поехали в Киев, Я догонял девушку с веслом, Flying planes can be dangerous.*