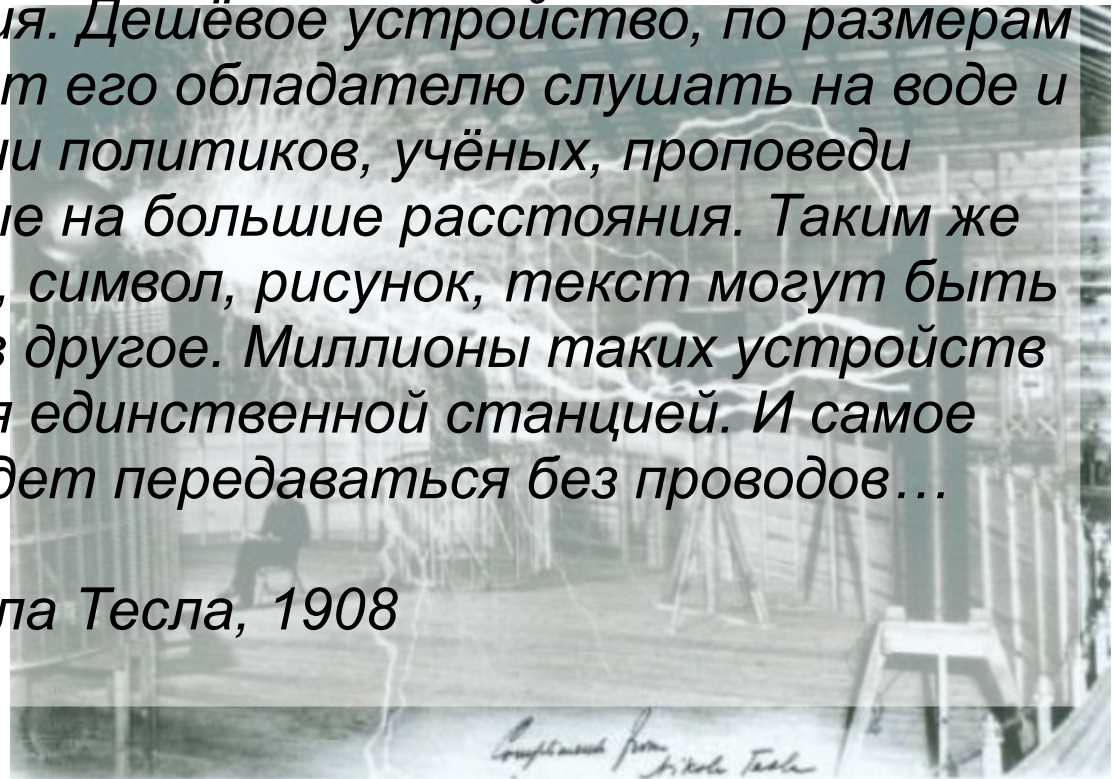


# Лекция №5

Поиск в Интернет

*Когда проект будет завершён, бизнесмен в Нью-Йорке сможет диктовать указания, и они будут немедленно появляться в его офисе в Лондоне или любом другом месте. Он сможет со своего рабочего места позвонить любому абоненту на планете, не меняя существующего оборудования. Дешёвое устройство, по размерам не больше чем часы, позволит его обладателю слушать на воде и суше музыку, песни, речи политиков, учёных, проповеди священников, доставляемые на большие расстояния. Таким же образом любое изображение, символ, рисунок, текст могут быть переданы из одного места в другое. Миллионы таких устройств могут контролироваться единственной станцией. И самое главное, что все это будет передаваться без проводов...*

*Никола Тесла, 1908*



# Интернет



Компьютер

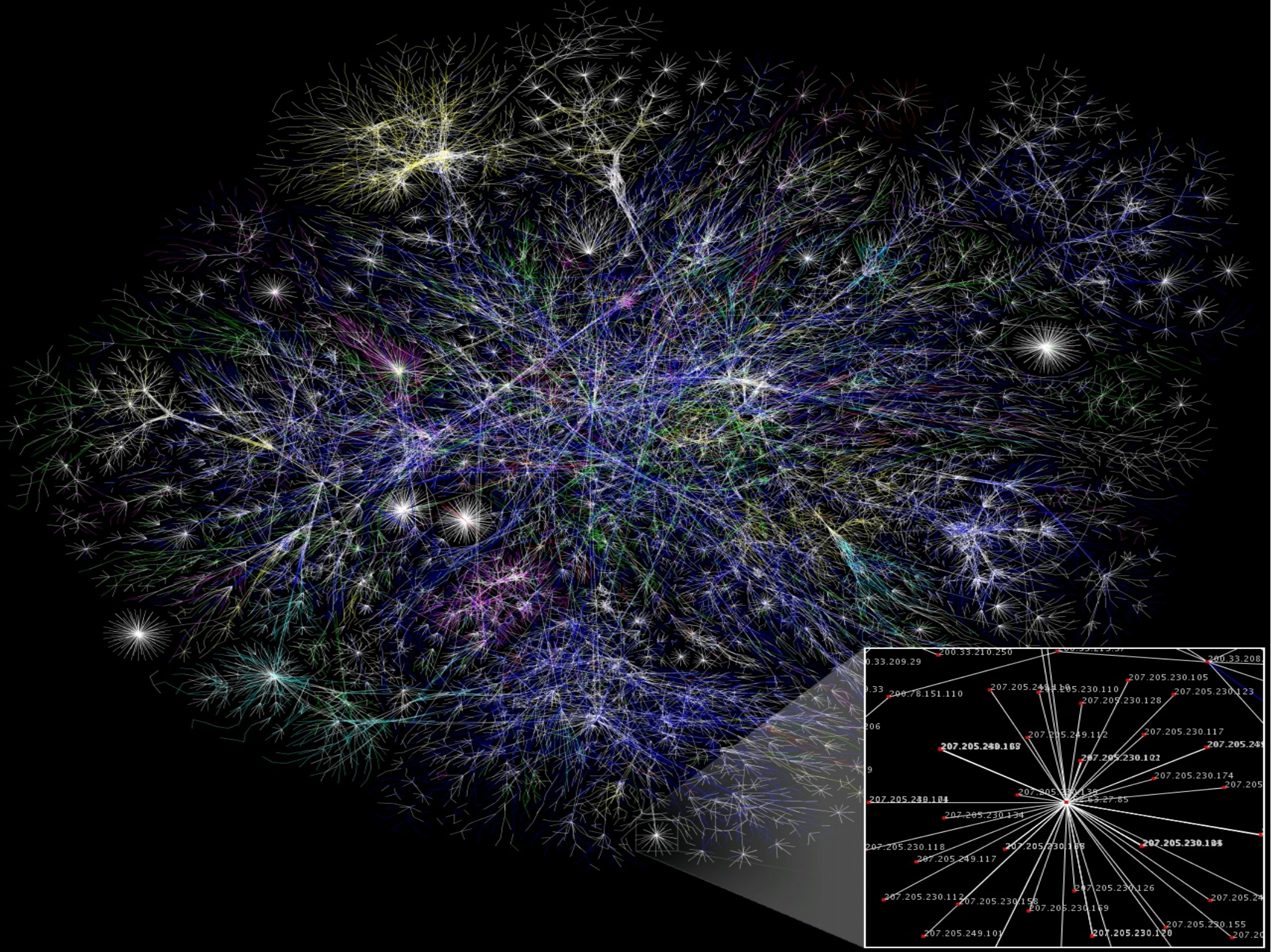


# Интернет



# Провайдер





# Интернет

**Интерне́т** (произносится [интэрнэ́т]; англ. **Internet**) — всемирная система объединённых компьютерных сетей, построенная на использовании протокола **IP** и маршрутизации пакетов данных. Интернет образует глобальное информационное пространство, служит физической основой для Всемирной паутины и множества других систем (протоколов) передачи данных. Часто упоминается как «Всемирная сеть» и «Глобальная сеть», в обиходе иногда употребляют сокращённое наименование «Инэ́т».

В настоящее время, когда слово «Интернет» употребляется в обиходе, чаще всего имеется в виду Всемирная паутина и доступная в ней информация, а не сама физическая сеть.

**Всеми́рная паути́на** (англ. **World Wide Web**) — распределенная система, предоставляющая доступ к связанным между собой документам, расположенным на различных компьютерах, подключенных к Интернету. Всемирную паутину образуют миллионы web-серверов. Большинство ресурсов всемирной паутины представляет собой гипертекст.

# Интернет

IP-адрес (IPv4 192.168.000.001)

DNS-имя (myhost.ru)

URL/URI

# URI

**URI (англ. Uniform Resource Identifier)** — унифицированный (единообразный) идентификатор ресурса. На английский манер произносится как [ю-ар-ай], по-русски чаще говорят [ури]. URI — это последовательность символов, идентифицирующая абстрактный или физический ресурс. Ранее назывался Universal Resource Identifier — универсальный идентификатор ресурса.

URI = URL + URN



# URL

**Единый указатель ресурсов (англ. URL — Uniform Resource Locator)** — единообразный локатор (определитель местонахождения) ресурса. Ранее назывался Universal Resource Locator — универсальный локатор ресурса. URL — это стандартизированный способ записи адреса ресурса в сети Интернет

<схема>://[<логин>[:<пароль>]@]<хост>[:<порт>]/

**URN (Uniform Resource Name):**

[<локальный путь к ресурсу>/]<имя ресурса>[? [<параметры>]  
[#<якорь>]]

# Схемы

Общепринятые схемы (протоколы) URL включают:

ftp — Протокол передачи файлов FTP

http — Протокол передачи гипертекста HTTP

https — Специальная реализация протокола HTTP, использующая шифрование (как правило, SSL или TLS)

gopher — Протокол Gopher

mailto — Адрес электронной почты

news — Новости Usenet

nntp — Новости Usenet через протокол NNTP

irc — Протокол IRC

prospero — Служба каталогов Prospero Directory Service

telnet — Ссылка на интерактивную сессию Telnet

wais — База данных системы WAIS

xmpp — Протокол XMPP (часть Jabber)

file — Имя локального файла

data — Непосредственные данные (Data: URL)

# Кодирование URL

Появление адресов URL стало существенным нововведением в Интернете. Однако стандарт URL обладает серьёзным недостатком — в нём можно использовать только ограниченный набор символов, даже меньший, нежели в ASCII: латинские буквы, цифры и лишь некоторые знаки препинания. Если мы захотим использовать в URL символы кириллицы, или иероглифы, или, скажем, специфические символы французского языка, то нужные символы должны быть перекодированы особым образом. Например, строка вида:

`http://ru.wikipedia.org/wiki/Микрокредит`

кодируется в URL как:

`http://ru.wikipedia.org/wiki/%D0%9C%D0%B8%D0%BA%D1%80%D0%BE  
%D0%BA%D1%80%D0%B5%D0%B4%D0%B8%D1%82`

Такое преобразование происходит в два этапа: сначала каждый символ кириллицы кодируется в Юникоде (UTF-8) в последовательность из двух байтов, а затем каждый байт этой последовательности записывается в шестнадцатеричном представлении:

М → D0 и 9C → %D0%9C

# Виды информации

Гипертекст (HTML)

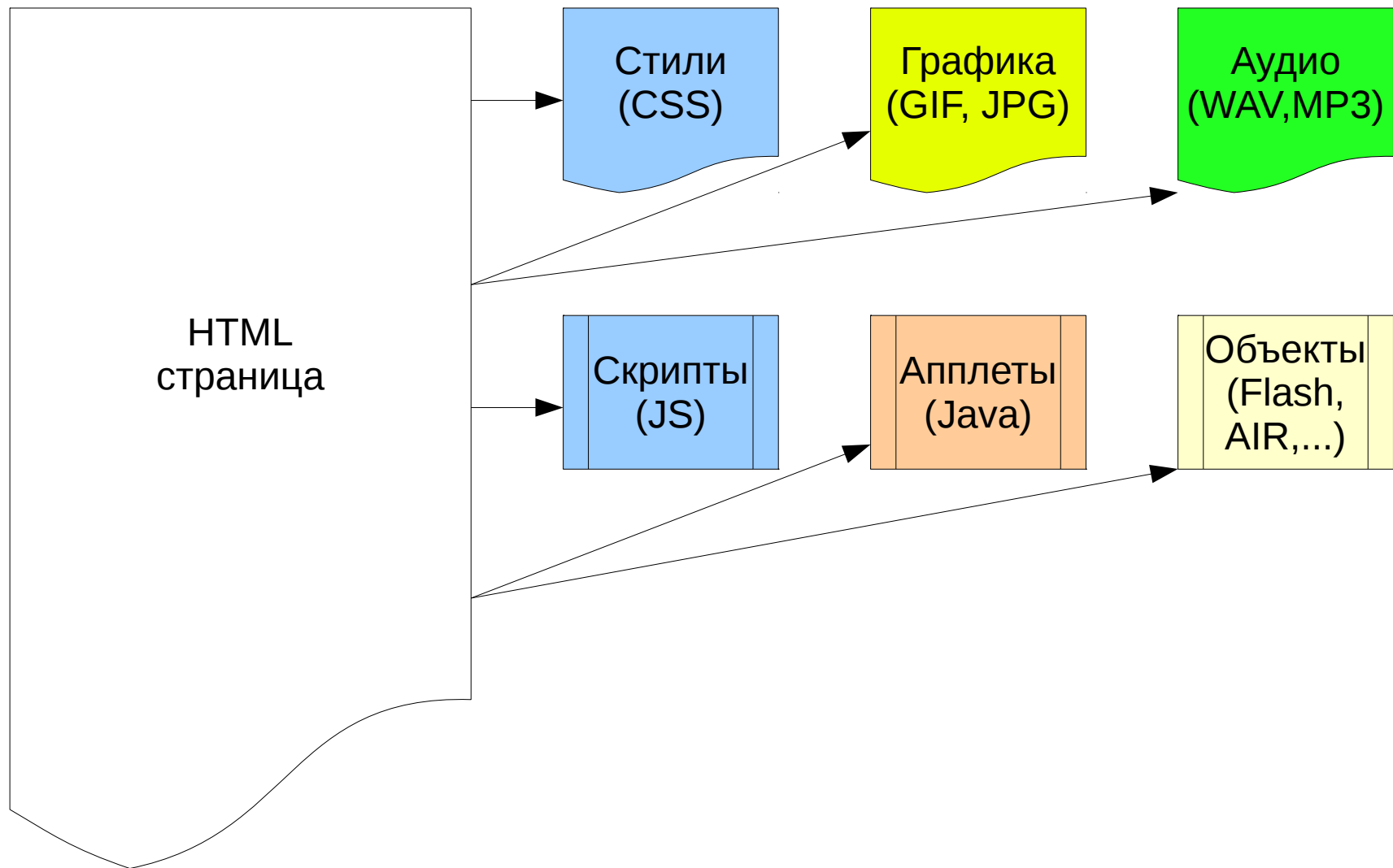
Файлы (FTP://, аудио, видео, графические)

Приложения/Объекты (Java, JavaScript, Flash)

# Гипертекст

```
<td height= 10 style= width: 1px; background-color: #4D6D82; ></td>
<td colspan="3" valign="bottom">
  <table width="100%" border="0" align="center" cellpadding="0" cellspacing="0">
    <tr><td><span class="style1">МИНТРАНС РФ - Федеральное агентство железнодорожного транспорта</span>
  </td>
    <td width='20'><a href='index.php?lang=1'><img src='image/rus.gif' width='15' height='10'
  alt='русский' border='0'></a></td><td width='20'><a href='index.php?lang=2'><img src='image/eng.gif'
  width='15' height='10' alt='английский' border='0'></a></td>
    </tr></table> </td>
  <td width="1" bgcolor="#4D6D82"></td>
    <td width="166" valign="bottom">
      <form name="form1" method="post" action="search.php?lang=1" style="margin:0; padding:0;">
        <input type="text" name="textfield" class="form_line" value="поиск...">
        <input type="submit" name="Submit" value="поиск" class="form_button"></form></td>
      <td width="1" bgcolor="#4D6D82"></td>
      <td width="10"></td>
    </tr>
  <tr bgcolor="#CCCCCC">
    <td height="1"></td>
    <td height="1"></td>
    <td></td>
    <td width="1"></td>
    <td></td>
    <td></td>
    <td></td>
    <td bgcolor="#CCCCCC"></td>
    <td></td>
  </tr>
  <tr>
    <td bgcolor="#003366">&nbsp;</td>
    <td class="w_menu"></td>
    <td width= 170 align="center" valign="top" bgcolor="#003366" onmouseover="P7_autoLayers(0)">
      <div align="center" style="padding-top: 10px; padding-bottom: 10px;"><a href="http://www.iriit.irk.ru
  /"></a>
      </div>
      <div class="style5">
        <div align="center">ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ</div>
      </div> </td>
    <td class="w_menu"></td>
    <td valign="top" bgcolor="#003366" onmouseover="P7_autoLayers(0)"><table width="100%" border="0"
  cellpadding="0" cellspacing="0">
      <tr>
        <td valign="middle"><div align="center">
```

# Гипертекст

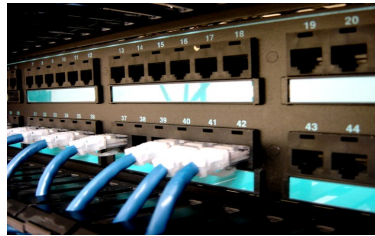


# Сайты в Интернет



ПК

канал связи



Провайдер



Сервер

81.19.85.116 — новости  
94.100.191.201 - почта

Lenta.ru - 81.19.85.116  
Mail.ru - 94.100.191.201

Altavista.com (1995)  
Yahoo.com (1995)  
Rambler.ru (1996)

Yandex.ru (1997)  
Google.com (1998)  
(Bing, Baidu, ...)



Сервер

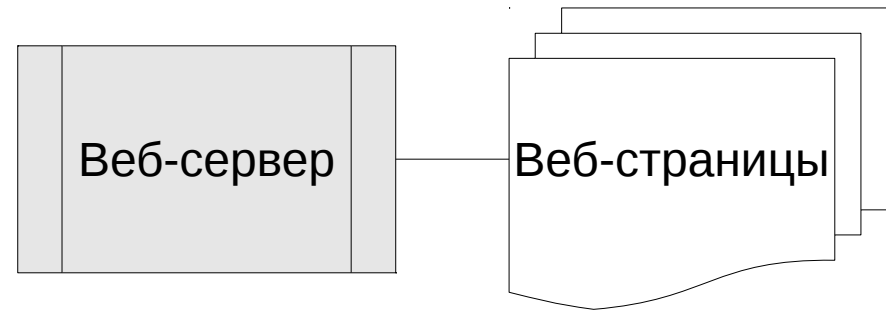
# Поиск

Поиск по сайтам

Поиск по сайту



# Поиск по сайту



# Поисковая система



# Поисковые системы

Всеязычные:

Google (37,2 %)

Bing (0,8 %)

Yahoo! (0,2 %, включая Inktomi, AltaVista, Alltheweb)

Русскоязычные — большинство «русскоязычных» поисковых систем индексируют и ищут тексты на многих языках — украинском, белорусском, английском, татарском и др. Отличаются же они от других систем, индексирующих все документы подряд, тем, что в основном индексируют ресурсы, расположенные в доменных зонах, где доминирует русский язык или другими способами ограничивают своих роботов русскоязычными сайтами.

Яндекс (48,1 %)

Mail.ru (5,9 %)

Рамблер (1,2 %)

Нигма (0,3 %)

Некоторые из поисковых систем используют внешние алгоритмы поиска. Nigma сочетает в себе как свой алгоритм, так и сборную выдачу от других поисковиков.

# Базы данных, знаний

*Энциклопедии*

bse.sci-lib.com — БСЭ

www.encyclopedia.ru

slovari.yandex.ru

wikipedia.org

*ГОСТы*

gost.ru

*Карты*

maps.yandex.ru, maps.google.com, 2gis.ru

Enter what you want to calculate or know about:

2+2



[Examples](#) [Random](#)

Input:

2 + 2

Result:

4

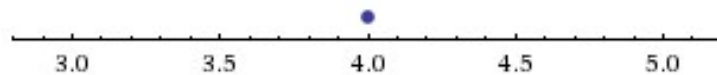
Number name:

four

Visual representation:



Number line:





2+2   [Найти!](#)

В найденном    в Иркутске    Поисковики    Язык    Сортировка    Настройки

**Справка**

- [2x2 \(2\\*2\)](#)

Название канала
2x2
Полное название
ЗАО «Телеканал 2x2»
Страна
Россия
<a href="#">- подробнее</a>
- [2+2](#)  
[- подробнее](#)

*Дано:*  $2 + 2$

*Ответ:*  
(Вычисление)  
 $4$

Понравилось, как решаются задачи? <sup>x</sup>  
Сделай Нигма.рф [поиском для FireFox!](#)

**Как правильно ввести задачу?**

- $\sqrt{a}$  – "корень a"
- $x^2$  – "x^(2)"
- $\frac{a}{b}$  – "(a)/(b)"

**Внимание:**  
Чтобы Нигма поняла Вас правильно, необходимо расставить скобки!  
[Служба поддержки](#)

[Что это такое?](#)   [Список решаемых задач](#)   [Пожаловаться](#)

**Фильтр**

Как это работает

- [скачать бесплатно](#)
- [телеканал 2+2](#)
- [онлайн](#)
- [программа телепередач](#)
- [интернет магазины](#)
- [Русскоязычные сайты](#)
- [Англоязычные сайты](#)

Применить ко всем

1. [Официальный сайт - 2x2](#)



«2x2» (читается как «дважды два») — первый в истории СССР и России коммерческий телеканал. С 19 марта 2007 года тематика телеканала — мультфильмы. В кабельных сетях Москвы и Санкт-Петербурга, он начал вещание 1 апреля 2007 года на частоте телеканала Rambler Телесеть. На...

[Читать дальше на Википедии →](#)

Сайт (рус.): [www.2x2tv.ru](http://www.2x2tv.ru)

2. [Официальный сайт - 2+2](#)

Телеканал «2+2» — кабельный развлекательный украинский телеканал, начавший вещание в 2006 г. на смену телеканалу Гравис-7 (183,25 МГц). Телеканал ориентируется на аудиторию возрастом 18—35 лет. Вещание производится преимущественно на украинском языке. Как правило, демонстрируются различные фильмы, сериалы и

Прокачай  
свой браузер  
ПОИСКОМ  
от Нигма.РФ



27%