

**Поиск информации.
Программные поисковые
системы**

Ключевые слова

- Всемирная паутина
- веб-страница, веб-сайт
- язык HTML
- гиперссылка
- браузер
- поисковая система
- запрос



Всемирная паутина

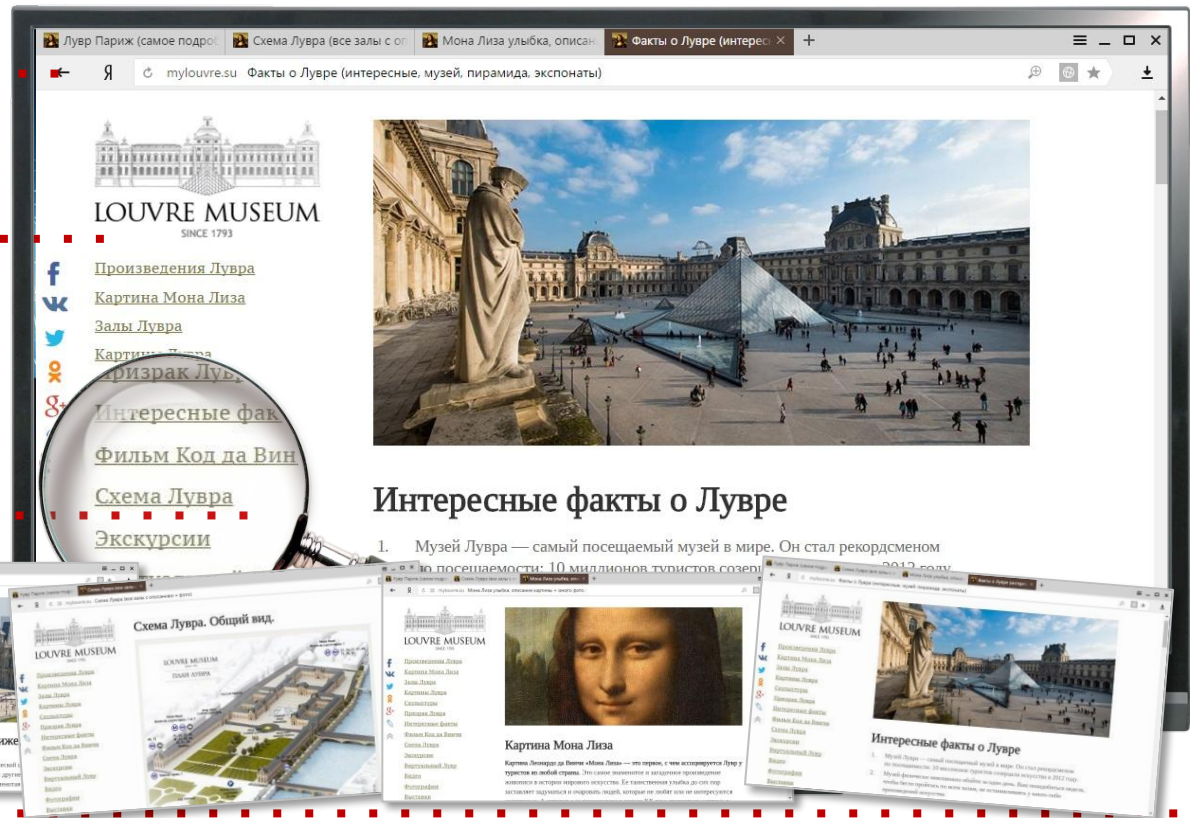
Браузер

Web-страница

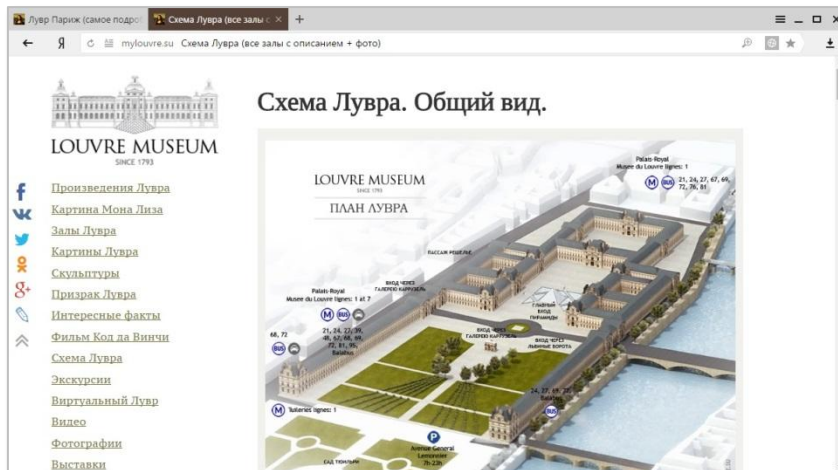
Гиперссылки

Web-сайт

Браузер (от англ. *browse* – просматривать) – специальная программа для просмотра веб-страниц.



Язык разметки гипертекста



```
1 <div id="fb-root"></div>
2 <script>(function(d, s, id) {
3   var js, fjs = d.getElementsByTagName(s)[0];
4   if (d.getElementById(id)) return;
5   js = d.createElement(s); js.id = id;
6   js.src = "//connect.facebook.net/ru_RU/sdk.js#xfbml=1&version=v2.3";
7   fjs.parentNode.insertBefore(js, fjs);
8 })(document, 'script', 'facebook-jssdk');</script>
9 <div id="fb-root"></div>
10 <!DOCTYPE html>
11 <html lang="en">
12 <head>
13   <meta charset="utf-8" /><link rel="shortcut icon" href="/images/favi
14
15 <!-- This site is optimized with the Yoast SEO plugin v4.2.1 - https://y
16 <title>Схема Лувра (все залы с описанием + фото)</title>
17 <meta name="description" content="Самые знаменитые произведения искусств
18 если вы хотите посмотреть «самое-самое»./>
19 <meta name="robots" content="nooodp"/>
```

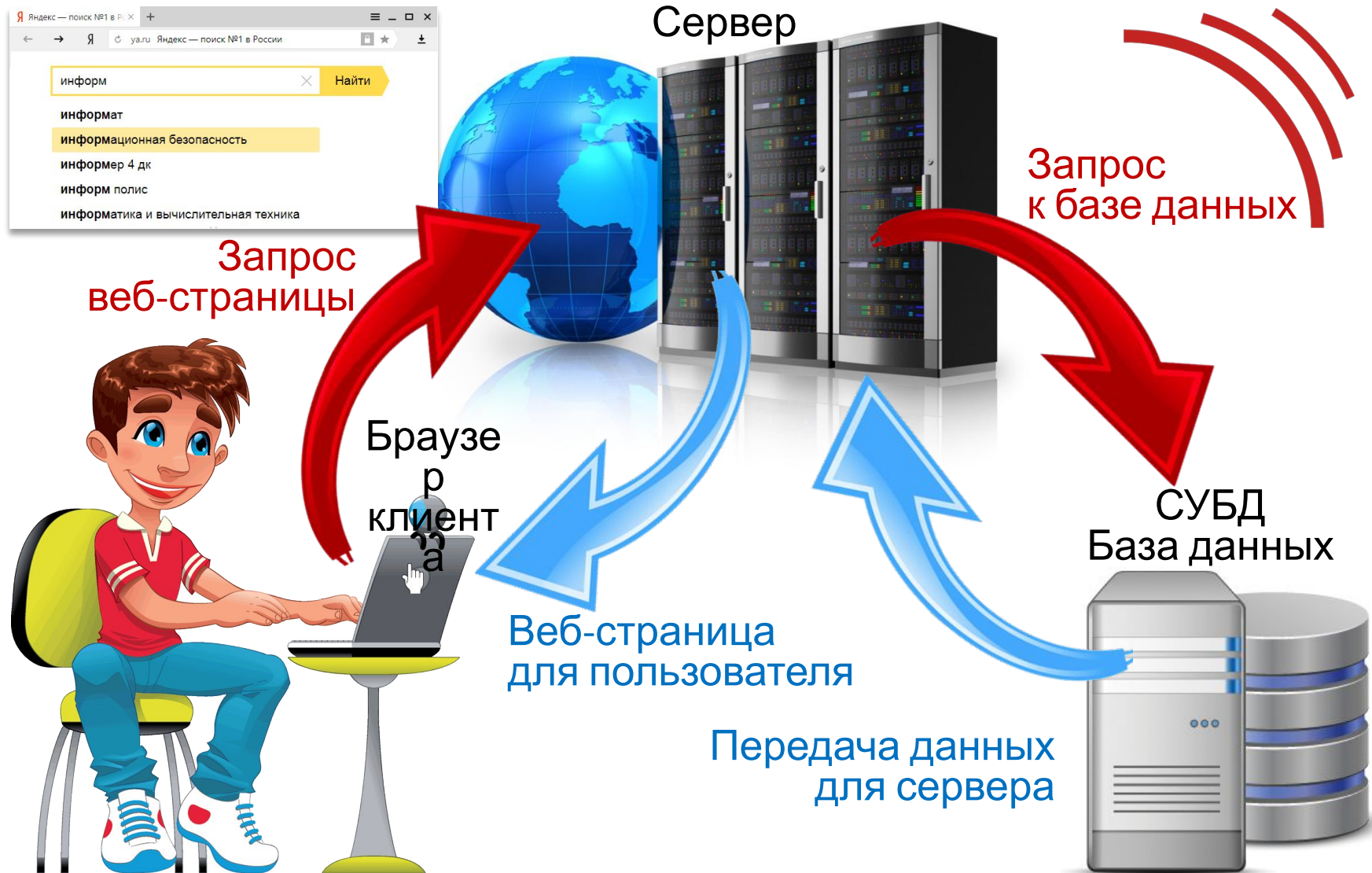
CSS (англ. *Cascading Style Sheets* – каскадные таблицы стилей) – формальный язык описания внешнего вида документа, составленного с использованием языка разметки.

Эта технология позволяет принципиально разделить содержание и представление документа:

- описание содержания и логической структуры веб-страницы производится с помощью HTML или других языков разметки;
- описание внешнего вида веб-страницы производится с помощью CSS.

Стиль хранится в отдельном CSS-файле, что позволяет создавать документы в едином стиле.

Схема обработки запросов пользователя



Взаимодействие клиент–сервер происходит по протоколу HTTP

Поиск информации в сети Интернет



Поисковая система – это программно-аппаратный комплекс, предназначенный для поиска информации во Всемирной паутине.

Поисковая машина – программная часть поисковой системы; комплекс программ, предназначенный для поиска информации

ПОИСКОВЫЕ СИСТЕМЫ

Поисковые каталоги, управляемые человеком

содержат базу данных ссылок на веб-сайты, распределённых по отдельным тематическим рубрикам; поиск осуществляется вручную спуском по дереву каталогов

Системы, использующие поисковых роботов

для каждого документа составляется его поисковый образ – набор ключевых слов, отражающих содержание этого документа; поиск осуществляется по запросу

Гибридные поисковые системы

сочетают в себе функции систем, управляемых человеком, и систем, использующих поисковых роботов

Типы поисковых запросов

Поиск по любому из слов

результатом является огромный список всех страниц, содержащих хотя бы одно из ключевых слов

Поиск по всем словам

в этом режиме формируется список всех страниц, содержащих все ключевые слова в любом порядке

Поиск точно по фразе

в результате составляется список всех страниц, содержащих фразу, точно совпадающую с ключевой (знаки препинания игнорируются)

Чтобы поиск стал более продуктивным, используйте языки формирования запросов, которые есть во всех поисковых системах.

Язык поисковых запросов

The screenshot shows a browser window with the Yandex search engine. The page title is 'Морфология и поисковый контекст'. The left sidebar contains a navigation menu with 'Морфология и поисковый контекст' highlighted. The main content area has a heading 'Морфология и поисковый контекст' and two paragraphs of text. The first paragraph explains that search results are based on word morphology. The second paragraph explains that by default, Yandex searches for all forms of a word. A table follows, detailing search operators. The table has four columns: 'Оператор', 'Описание', 'Синтаксис', and 'Пример запроса'. The first row describes the exclamation mark operator (!) used for exact word matching.

Яндекс **Помощь** Поиск

Морфология и поисковый контекст

Документные операторы
Исправление запроса
Семейный поиск
Расширенный поиск
Настройка поиска
Поисковые подсказки
Персональный поиск
Вопросы и ответы
Будьте осторожны

Морфология и поисковый контекст

При поиске с учетом морфологии принимаются во внимание:

- форма заданного слова (падеж, род, число, склонение и т. д.);
- часть речи (существительное, прилагательное, глагол и т. д.).

По умолчанию Яндекс ищет все формы слова, указанного в запросе. Например, при запросе [рассказал] поиск будет производиться по глагольным формам: «рассказать», «расскажу», «рассказывать» и т. д., но не по однокоренным словам типа «рассказ», «рассказчик». Исключение составляют случаи, когда используются операторы ! и " .

Также вы можете конкретизировать поисковый запрос с помощью операторов, которые уточняют наличие запрашиваемых слов в документе.

Оператор	Описание	Синтаксис	Пример запроса
!	Поиск слова в заданной форме. Допустимо использовать несколько операторов ! в рамках одного запроса.	<code>!слово</code>	[!рассказал] Будут найдены документы со словом «рассказал» в заданной форме.

Оценка эффективности поиска

ЭФФЕКТИВНОСТЬ (КАЧЕСТВО) ПОИСКА

Точность поиска – это отношение числа выданных релевантных (соответствующих информационному запросу) документов к общему числу документов, выданных системой по данному запросу.

$$\text{Точность} = \frac{D_{\text{релевантных}} \cap D_{\text{найденных}}}{D_{\text{найденных}}}$$

Значение этого параметра колеблется от 0,1 до 1.



Полнота поиска – это отношение числа выданных релевантных документов к общему числу релевантных документов, имеющихся в базе данных поисковой системы.

$$\text{Полнота} = \frac{D_{\text{релевантных}} \cap D_{\text{найденных}}}{D_{\text{релевантных}}}$$

В идеале это число должно равняться 1; на практике может достигать значений 0,7–0,9.

Достоверность информации на веб-ресурсах



Выяснение репутации сайта

Проверенные данные публикуют официальные сайты государственных, коммерческих, научных и других структур, являющиеся первоисточниками информации. Ответственность за любую опубликованную ими информацию несут ресурсы, имеющие свидетельство о регистрации средства массовой информации

Получение информации об авторе материала

Убедиться, что на сайте приведены данные об авторе (ФИО, квалификация, контактная информация и т.п.). Можно найти и ознакомиться с другими работами этого автора, отзывами на его работы.

Проверка фактического материала

Фактические и статистические данные можно выборочно сверить с официальными источниками. Они должны согласоваться с тем, что вы изучали в школе или узнали из других проверенных источников.

Самое главное

Веб-страница – это файл, содержащий собственно текст, несущий определённую информацию для пользователя, и служебную информацию для браузера (тэги разметки) на языке HTML (англ. *HyperText Markup Language* – язык разметки гипертекста).

HTML – один из веб-стандартов, по которым разрабатываются сайты во всём мире. Ещё одним из таких стандартов является технология **CSS** (англ. *Cascading Style Sheets* – каскадные таблицы стилей) – формальный язык описания внешнего вида документа, составленного с использованием языка разметки.

Технология CSS позволяет принципиально разделить содержание и представление документа: описание содержания и логической структуры веб-страницы производится с помощью HTML или других языков разметки, а описание внешнего вида веб-страницы производится с помощью CSS.



Самое главное

Веб-страницы предназначены для воспроизведения на самых разных экранах самых разных компьютеров. Поэтому они не имеют «жёсткого» форматирования. Оформление веб-страницы выполняется непосредственно во время её воспроизведения на компьютере клиента в соответствии с настройками используемого браузера.

Поисковая система – это программно-аппаратный комплекс, предназначенный для поиска информации во Всемирной паутине.

Поисковая машина – программная часть поисковой системы; комплекс программ, предназначенный для поиска информации.

По принципу действия поисковые системы делят на: поисковые каталоги, управляемые человеком; системы, использующие поисковых роботов; гибридные поисковые системы.

Важные данные, найденные в Интернете, необходимо проверять: выяснить репутацию сайта; получить информацию об авторе материала; проверить факты по данным из авторитетных источников.



Вопросы и задания



Задание 1. В языке запросов некой поисковой машины для обозначения логической операции ИЛИ используется символ «|», а для логической операции И — символ «&». Известны запросы и количество найденных по ним страниц некоторого сегмента сети Интернет:

Запрос	Найдено страниц (млн)
Информатика & Математика & Физика	12
Информатика & Математика	26
Информатика & Физика	34

Сколько страниц (в миллионах) будет найдено по запросу

Информатика & (Математика| Физика) ?

Решение

Вопросы и задания



Запрос	Найдено страниц (млн)
Информатика & Математика & Физика	12
Информатика & Математика	26
Информатика & Физика	34
Информатика & (Математика Физика)	?

Решение:

$$|И \cap М \cap Ф| = 12, |И \cap М| = 26, |И \cap Ф| = 34, И \cap (М \cup Ф) = ?$$

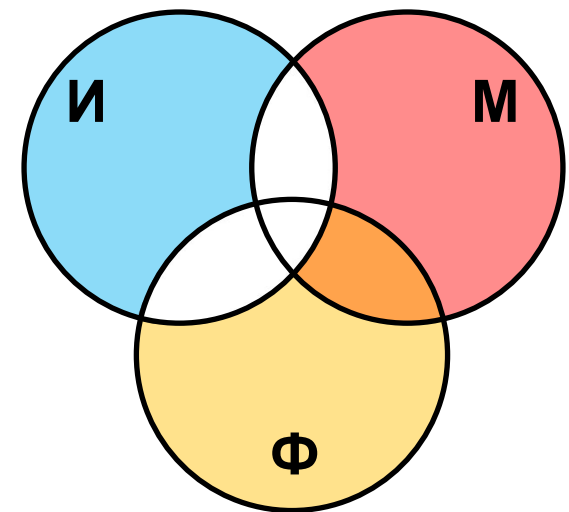
Преобразуем выражение:

$$И \cap (М \cup Ф) = (И \cap М) \cup (И \cap Ф).$$

Вспомним принцип включений-исключений для двух множеств: $|X \cup Y| = |X| + |Y| - |X \cap Y|$.

$$\begin{aligned} \text{Тогда } |(И \cap М) \cup (И \cap Ф)| &= \\ &= |И \cap М| + |И \cap Ф| - |(И \cap М) \cap (И \cap Ф)| = \\ &= |И \cap М| + |И \cap Ф| - |И \cap М \cap Ф| = \end{aligned}$$

Ответ: 48 (млн).



Информационные источники

- <http://mylouvre.su>
- <https://yandex.ru/support/search/query-language/qlanguage.html>
- http://pngimg.com/uploads/monitor/laptop_PNG5887.png
- http://gagauzpravda.md/wp-content/uploads/2016/10/loupe_PNG10020.png
- <https://blogukrnames.idua.org/wp-content/uploads/2016/07/world-wide-web.png>
- <http://mojthemes.cosmo-coder.com/dual-flow/Dark/images/browsers.png.pagespeed.ce.WC32NffDFX.png>
- <https://pollpath.com/user-img/1496491110contactbg-student.png>
- <https://privatfinance.com/wp-content/uploads/2016/04/Kak-pravilno-vyibrat-VPS-server-1024x819.jpg>
- <http://beandesoftware.com/img/DB-DEV-3.png>
- <http://images.easyfreeclipart.com/76/curved-arrow-blue-76226.png>
- <http://www.prof-lead.ru/wp-content/uploads/2012/08/31660297.jpg>
- <https://3.bp.blogspot.com/-VbthDDPwbKQ/WRHCvBdgDrI/AAAAAAAAAHw/eDACxFSJUWgNiBAZx2BXtOkAmxy52pvHQCLcB/s1600/Online.jpg>