

# Кодировка HTML

Подготовил:  
Леонтьев С.В

# Кодировка ASCII

- ASCII расшифровывается как Американский Стандартный Код для Обмена информацией (American Standard Code for Information Interchange). Он был разработан в начале 60-х годов 20-го века как стандартная кодировка для компьютеров и аппаратных устройств, таких как телепринтеры.
- ASCII — это семибитная кодировка, содержащая 128 символов.
- Она состоит из прописных и строчных латинских символов, цифр от 1 до 9 и некоторых специальных символов.
- Кодировки, используемые в современных **компьютерах**, HTML и в интернете, основаны на ASCII.
- В таблице ниже представлены все 128 символов ASCII и их аналоги в HTML.

Символ ASCII	HTML кодировка	Описание
	&#32;	пробел
!	&#33;	восклицательный знак
"	&#34;	компьютерная кавычка
#	&#35;	номер
\$	&#36;	знак <a href="#">доллара</a>
%	&#37;	процент
&	&#38;	амперсанд
'	&#39;	апостроф
(	&#40;	левая круглая скобка
)	&#41;	правая круглая скобка
*	&#42;	звездочка (астериск)
+	&#43;	плюс
,	&#44;	запятая
-	&#45;	дефис
.	&#46;	точка
/	&#47;	слэш
0	&#48;	цифра 0
1	&#49;	цифра 1
2	&#50;	цифра 2
3	&#51;	цифра 3
4	&#52;	цифра 4
5	&#53;	цифра 5
6	&#54;	цифра 6

# Кодировка Unicode

- Поскольку кодировки, перечисленные выше, ограничены по размеру и не совместимы с многоязычной средой, Консорциум Unicode разрабатывал свой стандарт.
- Стандарт Unicode включает в себя все символы, используемые в мире.
- Стандарт Unicode позволяет обрабатывать, хранить и обмениваться текстовой информацией независимо от платформы, программного обеспечения и языка.

- Консорциум Unicode разработал стандарт Unicode. Цель консорциума — заменить существующие кодировки своим стандартным Форматом Преобразования Unicode (Unicode Transformation Format — UTF).
- Стандарт Unicode был успешно введен в XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0, WML и т.д. Кроме того, стандарт Unicode поддерживается во многих операционных системах и современных браузерах.
- Консорциум Unicode сотрудничает с ведущими организациями по разработке стандартов, такими как ISO, W3C и ECMA.
- Чаще всего используют кодировки Unicode UTF-8 и UTF-16.

# UTF-8 и UTF-16

- Символы в UTF-8 могут быть длиной от 1 до 4 байт. С помощью этой кодировки можно отобразить любой символ. UTF-8 обратно совместим со стандартом ASCII. UTF-8 — рекомендованная кодировка для электронной почты и веб-страниц.
- 16-битный формат преобразования Unicode позволяет кодировать любые символы переменной длины. Эта кодировка используется в основных операционных системах и средах разработки, таких как Microsoft Windows 2000/XP/2003/Vista/CE, Java и .NET.