

Кодировка HTML

Подготовил:
Леонтьев С.В

• Существует (и применяется!) множество различных кодировок. Приведем краткий (!) список действующих:

- Кодировки Microsoft Windows:
 - Windows-1250 для языков Центральной Европы, которые используют латинское написание букв (польский, чешский, словацкий, венгерский, словенский, хорватский, румынский и албанский)
 - Windows-1251 для кириллических алфавитов
 - Windows-1252 для западных языков
 - Windows-1253 для греческого языка
 - Windows-1254 для турецкого языка
 - Windows-1255 для иврита
 - Windows-1256 для арабского языка
 - Windows-1257 для балтийских языков
 - Windows-1258 для вьетнамского языка
- MacRoman, MacCyrillic
- KOI8 (KOI8-R, KOI8-U...), KOI-7
- Болгарская кодировка
- ISCII
- VISCII
- Big5 (наиболее знаменитый вариант Microsoft CP950)
- HKSCS
- Guobiao
- GB2312
- GBK (Microsoft CP936)
- GB18030
- Shift JIS для японского языка (Microsoft CP932)
- EUC-KR для корейского языка (Microsoft CP949)
- ISO-2022 и EUC для китайской письменности
- Кодировки UTF-8, UTF-16 и UTF-32 набора символов Юникод

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 теле принтеры.
- 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
	 	пробел
!	!	восклицательный знак
"	"	компьютерная кавычка
#	#	номер
\$	$	знак доллара
%	%	процент
&	&	амперсанд
'	'	апостроф
((левая круглая скобка
))	правая круглая скобка
*	*	звездочка (астериск)
+	+	плюс
,	,	запятая
-	-	дефис
.	.	точка
/	/	слэш
0	0	цифра 0
1	1	цифра 1
2	2	цифра 2
3	3	цифра 3
4	4	цифра 4
5	5	цифра 5
6	6	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.