加密相关算法概要

华工

2016年11月30日

目录

1.字符编码

- 字符编码（英语：Character encoding）也称字集码，是把字符集中的字符编码为指定集合中某一对象（例如：比特模式、自然数序列、8位组或者电脉冲），以便文本在计算机中存储和通过通信网络的传递。

1.1 ASCII编码

- ASCII（American Standard Code for Information Interchange，美国信息交换标准代码）是基于拉丁字母的一套电脑编码系统，主要用于显示现代英语和其他西欧语言。它是现今最通用的单字节编码系统，并等同于国际标准ISO/IEC 646

- ASCII 码使用指定的7 位或8 位二进制数组合来表示128 或256 种可能的字符

1.1 ASCII编码

00101111	57	47	2F	/	斜杠
00110000	60	48	30	0	数字0
00110001	61	49	31	1	数字1
00110010	62	50	32	2	数字2
00110011	63	51	33	3	数字3
00110100	64	52	34	4	数字4
00110101	65	53	35	5	数字5
00110110	66	54	36	6	数字6
00110111	67	55	37	7	数字7
00111000	70	56	38	8	数字8
00111001	71	57	39	9	数字9
00111010	72	58	3A	:	冒号
00111011	73	59	3B	;	分号
00111100	74	60	3C	<	小于
00111101	75	61	3D	=	等号
00111110	76	62	3E	>	大于
00111111	77	63	3F	?	问号
01000000	100	64	40	@	电子邮件符号

base64

- Base64编码要求把3个8位字节（3*8=24）转化为4个6位的字节（4*6=24），之后在6位的前面补两个0，形成8位一个字节的形式。 如果剩下的字符不足3个字节，则用0填充，输出字符使用'='，因此编码后输出的文本末尾可能会出现1或2个'='。
- 为了保证所输出的编码位可读字符，Base64制定了一个编码表，以便进行统一转换。编码表的大小为2^6=64，这也是Base64名称的由来。

1.2 Unicode编码

- XML及其子集HTML采用UTF-8作为标准字集
- JAVA使用UTF-8

- BOM（Byte Order Mark），字节顺序标记，出现在文本文件头部，Unicode编码标准中用于标识文件是采用哪种格式的编码。

BOM

不同编码的字节顺序标记的表示

编辑

编码	表示(十六进制)	表示(十进制)
UTF-8	EF BB BF	239 187 191
UTF-16（大端序）	FE FF	254 255
UTF-16（小端序）	FF FE	255 254
UTF-32（大端序）	00 00 FE FF	0 0 254 255
UTF-32（小端序）	FF FE 00 00	255 254 0 0
UTF-7	2B 2F 76和以下的一个字节：[38 | 39 | 2B | 2F]	43 47 118和以下的一个字节：[56 | 57 | 43 | 47]
en:UTF-1	F7 64 4C	247 100 76
en:UTF-EBCDIC	DD 73 66 73	221 115 102 115
en:Standard Compression Scheme for Unicode	0E FE FF	14 254 255
en:BOCU-1	FB EE 28*及可能跟随着*FF	251 238 40*及可能跟随着*255
GB-18030	84 31 95 33	132 49 149 51

1.3 GBK

- GBK全称《汉字内码扩展规范》（GBK即“国标”、“扩展”汉语拼音的第一个字母，英文名称：Chinese Internal Code Specification），中华人民共和国全国信息技术标准化技术委员会1995年12月1日制订，国家技术监督局标准化司、电子工业部科技与质量监督司1995年12月15日联合以技监标函1995 229号文件的形式，将它确定为技术规范指导性文件。这一版的GBK规范为1.0版。
- GBK 亦采用双字节表示，总体编码范围为 8140-FEFE，首字节在 81-FE 之间，尾字节在 40-FE 之间，剔除 xx7F 一条线。总计 23940 个码位，共收入 21886 个汉字和图形符号，其中汉字（包括部首和构件）21003 个，图形符号 883 个。

url编码

- URL编码遵循下列规则： 每对name/value由&；符分开；每对来自表单的name/value由=符分开。如果用解码软件
- 户没有输入值给这个name，那么这个name还是出现，只是无值。任何特殊的字符（就是那些不是简单的七位ASCII，如汉字）将以百分符%用十六进制编码，当然也包括象=,&；，和 % 这些特殊的字符。其实url编码就是一个字符ascii码的十六进制。不过稍微有些变动，需要在前面加上“%”。比如“\”，它的ascii码是92，92的十六进制是5c，所以“\”的url编码就是%5c。那么汉字的url编码呢？很简单，看例子：“胡”的ascii码是-17670，十六进制是BAFA，url编码是“%BA%FA”。

2.加密算法

- 数据加密的基本过程就是对原来为明文的文件或数据按某种算法进行处理，使其成为不可读的一段代码，通常称为“密文”，使其只能在输入相应的密钥之后才能显示出本来内容，通过这样的途径来达到保护数据不被非法人窃取、阅读的目的。 该过程的逆过程为解密，即将该编码信息转化为其原来数据的过程。
- 加密技术通常分为两大类：“对称式”和“非对称式”。
- DES（Data Encryption Standard）
- 3DES（Triple DES）
- RC2和RC4
- IDEA（International Data Encryption Algorithm）
- RSA
- AES

3.签名算法（摘要算法）

- 数据摘要算法是密码学算法中非常重要的一个分支，它通过对所有数据提取指纹信息以实现数据签名、数据完整性校验等功能，由于其不可逆性，有时候会被用做敏感信息的加密。数据摘要算法也被称为哈希（Hash）算法、散列算法。
- 1、CRC8、CRC16、CRC32（通讯领域）
- 2、MD2 、MD4、MD5（密码，文件摘要）
- 3、SHA1、SHA256、SHA384、SHA512（ CA和数字证书）
- 4、RIPEMD、PANAMA、TIGER、ADLER32