

浅析手机气象短信息乱码的问题

杨荣芳^{1,2} 屈玉贵¹

(1. 中国科学技术大学电子工程与信息科学系,合肥 230026;

2. 河北省气象科技服务中心)

提 要: 随着 GSM (global system for mobile communication) 短信息服务 (short message service, SMS) 的推广和应用,手机短信已经成为气象信息服务传播的重要渠道。但是在为用户提供气象信息服务中,经常会出现用户手机收到的气象短信为乱码的现象。介绍了有关手机短信编码的规则,发生接收短信息乱码的原因,并结合实际应用提出了相关的解决方法。

关键词: 手机短信 短信编码 乱码

引 言

随着计算机技术、无线通信网络技术的不断发展,短信以其成本低特性,迅速发展为企业向客户提供服务的主要通道和手段,其惊人的发展速度,无疑创造了“拇指经济”的神话。气象短信服务在此背景下应运而生,很快发展成为短信行业的领头军。但是在为客户提供气象短信息服务的同时,也出现用户由于接收信息不正常而引发的各种投诉问题,如接收不到、接收信息显示不正常等。因此,保证客户短信息的接收质量成为短信稳定发展的关键问题。

通过对公众气象短信投诉的分析,影响气象短信服务质量的因素有很多,在保证气象短信正常准确发送的情况下,短信息乱码是用户投诉的原因之一。短信息出现乱码基本原理都是由于信息编码和解码不统一所

致。本文主要从手机用户发送和接收短信息乱码方面,介绍乱码产生的原因、表现形式及其解决方法,解读不同字符集的相互转换。

1 短信息编码的模式

手机发送短消息常用 Text 和 PDU (Protocol Data Unit, 协议数据单元) 模式。使用 Text 模式收发短信代码简单,实现起来十分容易,但 Text 模式只能发送普通的 ASCII 字符,最大的缺点是不能收发中文短信;而 PDU 模式不仅支持中文短信,也能发送英文短信。目前,气象短信息服务主要是以文本、数字或二进制数据为主,主要采用 PDU 模式,这种短消息的长度被限定在 140 字节之内,即 70 个中文字符。中文短消息的编码实现较简单,只需将 GB2312 的中文编码转换为代码页为 CP936 的 Unicode 编码即可。如表 1^[1] 所示。

表 1 中文编码的实现过程

字符	082312 编码	082312 编码 (BIN)	Unicode 编 码 代 码 页 为 Unicode 编码 CP936	
中	D6	1101 0110	0100 1110	4E
	D0	1101 0000	0010 1101	2D
文	CE	1100 1110	0110 0101	65
	C4	1100 0100	1000 0111	87
短	B6	1011 0110	0111 0111	77
	C4	1100 1111	0110 1101	6D
消	CF	1100 1111	0110 1101	6D
	FB	1111 1011	1000 1000	88
息	CF	1100 1111	0110 0000	60
	A2	1010 0000	0110 1111	6F

1.1 PDU 编码规则

PDU 模式收发短信可以使用 3 种编码：7bit、8bit 和 UCS2 编码^[2]。7bit 编码用于发送普通的 ASCII 字符，它将一串 7bit 的字符（最高位为 0）编码成 8bit 的数据，每 8 个字符可“压缩”成 7 个；8bit 编码通常用于发送数据消息，比如图片和铃声等；UCS2 编码用于发送 Unicode 字符。7bit 编码时，指原始短消息的字符个数，而不是编码后的字节数。如果用户信息（TP—UD）中存在一个头（基本参数的 TP—UDHI 为 1），在所有编码方式下，用户信息长度（TP—UDL）都等于头长度与编码后字节数之和。如果采用 GSM 03.42 所建议的压缩算法（TP—DCS 的高 3 位为 001），则该长度也是压缩编码后字节数或头长度与压缩编码后字节数之和。目前普遍使用的为文本短信息，所以采用的为 Unicode 字符中的 UCS2 编码。PDU 串的用户信息（TP—UD）段最大容量是 140 字节，所以在这种编码方式下，可以发送的短消息的最大字符数为 140 个。一般的 PDU 编码由 A—M 十三项组成。

A:短信息中心地址长度,2 位十六进制数(1 字节)。

B:短信息中心号码类型,2 位十六进制数。

C:短信息中心号码,B+C 的长度将由 A 中的数据决定。

D:文件头字节,2 位十六进制数。

E:信息类型,2 位十六进制数。

F:被叫号码长度,2 位十六进制数。

G:被叫号码类型,2 位十六进制数,取值同 B。

H:被叫号码,长度由 F 中的数据决定。

I:协议标识,2 位十六进制数。

J:数据编码方案,2 位十六进制数。

K:有效期,2 位十六进制数。

L:用户数据长度,2 位十六进制数。

M:用户数据,其长度由 L 中的数据决定。J 中设定采用 UCS2 编码,这里是中英文的 Unicode 字符。

1.2 手机短信息出现乱码的现象及其原因

1.2.1 手机收到空的短信息

当发送方信息已经发出,而接收方手机没有收到或者收到为空信息。前者是由于短信在传输路径上发送失败,原因有很多,其中主要的有三种情况,第一种情况为手机用户在网关鉴权失败,此用户的正常服务受限制,不能为其进行短信服务;或者短信在传输时延时太长,超过了它存在的生命周期而被系统抛弃;最后一种情况如果手机用户不在服务区或者关机,也不能接收到短信。

当手机用户收到短信息,但信息内容为空,主要是因为发送方信息编码与接收方解码不一致造成,当接收信息为空时,一般是因为手机不支持此种字符无法显示,或者因为手机用户没有对文字信息字符显示进行设置。

1.2.2 手机收到短信息不完整

有时手机只收到部分短信息内容,这是由于短信息本身采用 PDU 编码模式造成。采用 PDU 编码模式,加之受到手机机型的限制,一般手机支持发送短信在 70 个汉字或

140 个英文字符(标点作为汉字字符),因此如果超出这个长度,短信内容将被自动截去超长的部分,只收到允许长度范围内的内容。

1.2.3 手机收到短信息为乱码

当手机收到短信息为无法辨认的字符时,一般意义称为乱码。这种乱码和收到空信息的基本原理时一样的,都是由于发送方信息编码与接收方解码不一致造成,使解码出来的短信息不正确,从而无法正常显示。

2 手机短信息乱码的解决方法

对于手机收到短信息乱码的问题,可以从发送方和接收方两方面进行解决。一方面,是手机本身造成的原因;另一方面,由于发送方的原因生成乱码。

2.1 手机本身造成的乱码

一种情况为:用户的手机本身有缺陷(非中国区销售的手机),由于手机汉字内码不一样,在接收中文短消息是会造成接收乱码;另一种情况可能是由于手机不支持中文短消息引起的。还有某些型号手机虽然能显示中文,但却无法显示中文标点,会出现手机收到空信息或者乱码。针对手机本身造成乱码的情况,如果在手机本身设置为支持全部字符情况下,收到的短信息仍为乱码,一般建议用户更换手机或者尽量不采用易出现乱码的字符。

2.2 发送方原因造成的乱码

发送方首先要将发送的信息长度控制在手机允许接收短信长度范围之内。一般小灵通手机的一条短信长度是 45 到 56 个中文字

左右;手机短消息最长为 160 个英文字母(包括数字)或者 70 个汉字字符或汉字英文混合字符,但不同型号的手机可能会有不同的限制,所以一般在短信息长度超过 140 字符数时,即超长短信,有些手机不支持长短信,超长部分信息会被系统自动截断,或者为乱码,造成用户接收信息不完整。

发送方在发送信息时,一定要选择正确的编码格式。由于气象短信息发送的为中文短信息,因此在发送信息时,一定要选择 UCS2 编码,即中文 Unicode 码。接收方不仅收到 UCS2 格式编码的 PDU 串,也有可能是 7bit 编码格式(TP-DCS 为 00)或 8bit 编码格式(TP-DCS 为 04)的 PDU 串,对于此情况,接收方一定要选择相应的算法进行解码。

3 小结

为了避免在发送气象短信息时,用户收到气象短信乱码的现象,就必须在发送短信时,采用正确的编码格式和规范的短信长度,尽量使用通用手机类型,以保证短信用户接收信息的准确性,从而使气象短信服务做得更好。

参考文献

- [1] 徐彦琦,刘雅晴,黄登山. 基于 GSM 短消息的编解码及编码实现[J]. 通信技术,2006-10-18:56-58.
- [2] 王晋海,刘光昌. 短信息服务 SMS 的开发[J]. 计算机工程与设计,2003-02-22:77-79
- [3] 亓莱滨.《Unicode 内码转换与汉字乱码》[M]. 电脑知识与技术,北京:158-160.