

ICS 33.070

M 37

YD

中华人民共和国通信行业标准

YD/T 1902-2009

消息类业务内容分类技术要求

Technical Requirements for Content Classification of
Messaging Service

2009-06-15 发布

2009-09-01 实施

中华人民共和国工业和信息化部 发布

目 次

前 言	II
引 言	III
1 范围	1
2 规范性引用文件	1
3 术语、定义和缩略语	1
4 总体要求	4
5 分类编码	4
6 分类编码实施	6
附录A(规范性附录) 基于分类代码的消息过滤业务实施	7
附录B(资料性附录) 分类编码对消息容量的影响	8
参考文献	9

前 言

本标准是“基于用户设置规则的消息过滤业务”系列标准之一，该系列标准的名称预计如下：

1. YD/T 1774-2008 基于用户设置规则的消息过滤业务技术要求
2. YD/T 1775-2008 基于用户设置规则的短消息过滤系统技术要求
3. YD/T 1902-2009 消息类业务内容分类技术要求
4. 移动终端垃圾短信息过滤技术要求

本标准的附录A为规范性附录。

本标准的附录B为资料性附录。

本标准由中国通信标准化协会提出并归口。

本标准起草单位：工业和信息化部电信研究院、中国联合网络通信有限公司、中国移动通信集团公司、华为技术有限公司、中兴通讯股份有限公司、北京电信规划设计院有限公司、大唐电信科技产业集团、北京展讯高科通信技术有限公司、国家计算机网络应急技术处理协调中心

本标准起草人：落红卫、何桂立、张 薇、王新京、舒 敏、陈 萍、曾明发

引 言

群发消息类业务的广泛应用，不仅为用户提供了大量的有用信息，为政府部门提供了便利的宣传渠道；同时也为商家拓展了便捷的广告手段，为消息业务服务提供商带来了巨额的经济效益。但不可否认，目前群发消息通常没有确定的发送目标而多采取广播式的发送方式，在很大程度上给用户形成严重的骚扰。同时，用户也没有适合的方式来选择需要的消息，只能被动盲目地接收消息。为使用户可以有选择地接收自己希望得到的消息，而不是被迫接收消息，特制定本标准，目的是在群发消息发出之前，就在消息中加入消息类业务内容分类代码，进而用户在接收时可以选择接收需要的消息，同时屏蔽不需要的消息。

基于消息类业务内容分类代码的实施，不仅可以规范群发消息服务提供商相关行为，提升服务水平，还可以保护用户免受垃圾消息的骚扰和侵害。同时，可以提升行业主管部门专业性监管能力，配合相关的监管策略，去莠存良，使群发消息类业务得以健康发展。

消息类业务内容分类技术要求

1 范围

本标准描述了消息类业务内容分类总体技术要求，规范了消息类业务分类编码方案和分类代码表，并且对消息分类编码在消息类业务中具体实施进行了具体技术规范。

本标准主要适用于消息业务提供商提供的群发广告消息类业务，其他消息业务可参照执行。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准。然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 4754	国民经济行业分类与代码
YD/T 1774-2008	基于用户设置规则的消息过滤业务技术要求
YD/T 1775-2008	基于用户设置规则的短消息过滤系统技术要求

3 术语、定义和缩略语

下列术语、定义和缩略语适用于本标准。

3.1 术语和定义

3.1.1

编码 Coding

给事物或概念赋予代码的过程。

3.1.2

编码对象 Coded Objects

被编码的事物或概念。

3.1.3

编码对象集 Coded Set

一个元素集按一种编码方案映射到另一个元素集上。

3.1.4

编码方案 Coding Scheme

将一个集合中的元素映射到另一个集合中的元素的规则汇集。

3.1.5

代码 Code

表示特定事物或概念的一个或一组字符。

3.1.6

代码值 Code Value

将一种编码方案应用于编码对象集中的一个元素所产生的结果。

3.1.7

代码元素集 Code Element Set

将一种编码方案应用于编码对象集中的所有元素所产生的结果。

3.1.8

代码表 Code List

编码对象集和代码元素集的映射关系列表。

3.1.9

代码结构 Code Structure

代码字符排列的逻辑顺序。

3.1.10

代码长度 Code Length

一个代码中所包含的有效字符的个数。

3.1.11

等长码 Equal Length Code

在同一个代码体系中，所有编码对象的代码长度都相等。

3.1.12

短消息业务 Short Message Service

短消息业务是通信系统提供的通信终端之间，或者通信终端与其他短消息实体之间进行文字信息收发业务。

3.1.13

短消息业务中心 Short Message Service Center

在短消息业务网络中，完成短消息发送、接收、转发和存储等处理功能的系统。

3.1.14

多媒体消息 Multimedia Message

一种或者多种不同媒体类型元素组成的多媒体表述。它包括格式化文本、图片、图像、动画、声音、语音、视频剪辑等媒体类型的消息。

3.1.15

分类 Classification

按照选定的属性（或特征）区分分类对象，将具有某种共同属性（或特征）的分类对象集合在一起的过程。

3.1.16

行业 Industry

从事相同性质的经济活动的所有单位的集合。

3.1.17

互联网短消息网关 Internet Short Message Gateway

业务提供商与短消息业务中心之间的中介实体。互联网短消息网关一方面负责接收业务提供商发送给移动用户的短消息并且提交给短消息中心。另一方面，移动用户点播业务提供商的信息将由短消息中心通过互联网短消息网关发给业务提供商。另外，互联网短消息网关还应根据路由原则将业务提供商提交的信息转发到相应的互联网短消息网关。

3.1.18

类 Class

具有某种共同属性（或特征）的事物或概念的集合。

3.1.19

码位 Position

在字符串中，可由元素占据并用序号标识的部位。

3.1.20

上位类 Category in Higher Level

在线分类体系中，一个类目相对于由它直接划分出来的下一级类目而言，称为上位类。

3.1.21

受众 Receiver

消息的接受者和利用者，即消息类业务的服务对象。

3.1.22

数字代码 Numeric Code

有阿拉伯数字组成的代码。

3.1.23

线分类法 Method of Linear Classification

将分类对象按选定的若干属性（或特征），逐次地分为若干层级，每个层级又分为若干类目，同一分支的同层级类目之间构成并列关系，不同层级类目之间构成隶属关系。

3.1.24

消息 Message

消息是指包含信息的语言、文字和图像等。

3.1.25

消息类业务 Messaging Service

通信系统提供的通信终端之间，或者通信终端与其他短消息实体之间进行文本、图形、音频、视频信息收发的业务。本标准中主要指短消息业务和多媒体消息业务。

3.1.27

信息分类 Information classifying

把具有某种共同属性（或特征）的信息归并在一起，把具有不同属性（或特征）的信息区分开的过程。

3.1.28

字母代码 Alphabetic Code

有拉丁字母组成的代码。

3.2 缩略语

ISMG	Internet Short Message Gateway	互联网短消息网关
MC	Message Center	消息中心
MMS	Multimedia Message	多媒体消息
MS	Mobile Station	移动台
SME	Short Message Entity	短消息实体
SMIL	Synchronised Multimedia Integration Language	同步多媒体集成语言
SMS	Short Messaging Service	短消息业务
SP	Service Provider	服务提供商
WAP GW	WAP Gateway	WAP网关

4 总体要求

消息类业务最大的特点就是采用“主动发送，被动接收”的工作模式进行传输，用户并不知道发送者发送的是什么内容的消息，在这种情况下只有被迫接收到以后才能判断该消息是否是自己需要的。往往在这种情况下，用户已经受到垃圾消息的骚扰。为了便于用户有选择地接收需要的消息，有必要通过技术手段最大限度地解决消息识别问题。

根据消息类业务的特点，解决消息内容识别并且进行分类最有效的方法就是在消息源头使用合理有效的编码方案对要发送的消息进行分类编码，并且把相应代码添加到消息体前面作为消息内容分类依据，从而保证所有消息具备消息内容分类特征以备后期识别判定使用。

在现实生活中，消息类业务内容涉及社会生活的方方面面，制定出一套合理有效的编码方案并不容易。为了确保该编码方案可以真正满足消息过滤业务的需要，制定消息类内容分类编码方案应满足以下要求：

- 唯一性：每一个编码对象仅对应一个代码，一个代码只唯一表示一个编码对象；
- 合理性：有恰当的分类依据，代码结构与分类体系相适合；
- 扩充性：代码应可以满足不断编码对象扩充和修改的需求；
- 规范性：代码结构和编写格式必须一致。

另外，消息类业务中消息内容空间往往有一定限制，例如短消息业务内容代码空间只有140个字节。故此，消息类业务内容的分类代码要求尽量简短。同时消息类业务内容分类代码尽量使用等长代码，这样，消息过滤实体或者消息监听实体就可以在不分析消息体内容的情况下，直接按照字节位直接读取分类代码进而识别消息类别。

5 分类编码

5.1 分类编码方案

消息类业务内容包罗万象，涉及社会经济生活的方方面面，而国民经济行业分类是对社会经济活动进行的分类，是社会经济行业划分的基本依据。因此，基于国民经济行业的分类是最适合消息类内容分类的分类方法。在进行国民经济行业分类时候，通常采用经济活动的同质性原则进行行业类别的划分。即每一个行业类别都按照相同性质的经济活动归类，而不是单单依据行政事业编制、会计制度和部门管理归类。

根据国内外经济实际情况，国民经济行业分类适合采用线分类和分层编码的方法，划分为门类、大类、中类和小类四级。门类是国民经济行业分类中活动性质相近的经济部门的综合类别；大类构成国民

经济重要的经济部门；中类是活动性质相近的小类行业的综合类别；小类是国民经济行业分类的核心层，构成了全社会经济活动中可供观察和度量的最小的产业活动类别的全部内容。门类分类结构成树状结构。其中，门类采用英文字母代码，大类、中类和小类采用阿拉伯数字代码。分类信息采用等长编码，代码长度总长为5：门类字母代码码位为1，采用升序编码，顺序为A、B、C…；大类数字代码码位为2、3，采用升序编码，顺序为01、02、03…；中类数字代码码位为4，编码采用升序编码，顺序为1、2、3…；小类数字代码码位为5，采用升序编码，顺序为1、2、3…。其中：如果中类和小类需要设立带有“其他”的收容项，原则上规定收容项的代码尾数为“9”；如果大类或中类不再细分，则后面的代码补“0”直到第四位。相应编码结构如图1所示。

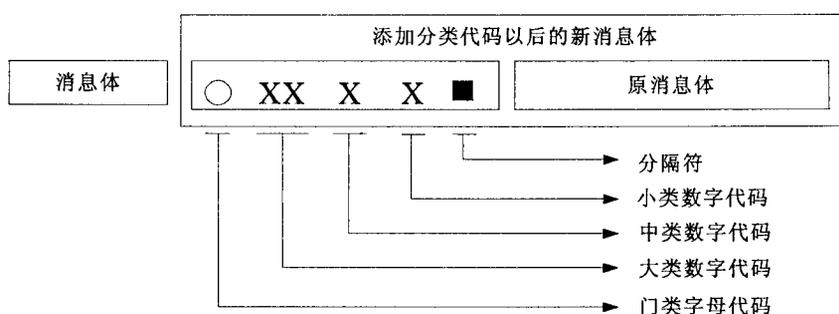


图1 分类编码结构

分类代码内代码紧密连接，无分隔符。分类代码与消息内容之间以空格作为分隔符。因此，分类代码加分隔符一共占用6个字符的长度。表1为门类代码，参见GB 4754-2002第5章。

表1 门类代码表

门类字母代码	门类名称
A	农、林、牧、渔业
B	采矿业
C	制造业
D	电力、燃气及水的生产和供应业
E	建筑业
F	交通运输、仓储和邮政业
G	信息传输、计算机服务和软件业
H	批发和零售业
I	住宿和餐饮业
J	金融业
K	房地产业
L	租赁和商务服务业
M	科学研究、技术服务和地质勘查业
N	水利、环境和公共设施管理业
O	居民服务和其他服务业
P	教育
Q	卫生、社会保障和社会福利业
R	文化、体育和娱乐业
S	公共管理和社会组织
T	国际组织

大类、中类和小类依据其归属的上位类而各有不同。图2是以住宿和餐饮业(I)为例进行的编码划分。

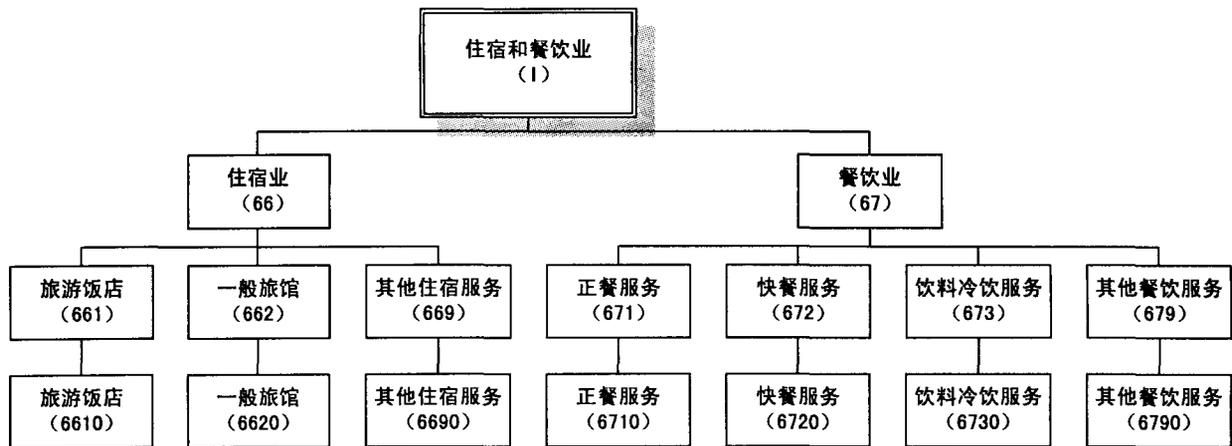


图2 编码划分示意 (住宿和餐饮业)

5.2 分类代码表

根据我国国民经济情况，国民经济生活可以分为20个门类，95个大类，396个中类，913个小类。具体分类代码表见GB 4754-2002第5章。

5.3 分类编码对消息容量的影响

参见附录B。

6 分类编码实施

6.1 对消息业务提供商的要求

对于所有群发消息，消息业务提供商应根据图1的消息类内容分类代码结构要求在消息体的前面添加消息类内容分类代码（见GB 4754-2002第5章）和分隔符，其中分类代码应和所要发送的内容相一致；对于添加消息类内容分类代码和分隔符以后，消息内容长度大于消息容量的消息，必须进行消息分割，并且要求在每一部分有相应的分类代码和分隔符。

6.2 对消息业务运营商的要求

消息业务运营商应具备验证消息内容与分类代码是否一致的能力。

消息业务运营商应具备一定手段，可以在群发消息类业务接入设备上（例如：ISMG、WAP GW等设备）上抽查或者实时监控群发消息，提取并分析消息的分类代码，验证消息内容是否与分类代码一致。分类代码验证示意图如图3所示：

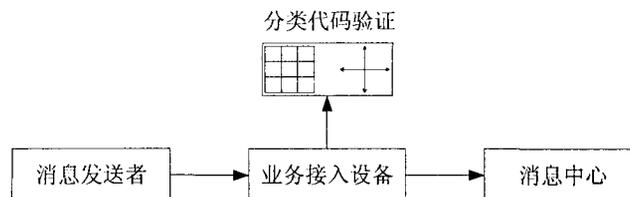


图3 分类代码验证示意图

对于不具备分类代码的消息或者分类代码与消息内容不一致的消息应拒绝发送。抽查验证或者实时监控既可以采用人工抽查方式，也可采用自动验证方式。

6.3 基于分类代码的消息过滤业务实施

见附录A。

附录 A

(规范性附录)

基于分类代码的消息过滤业务实施

本附录的目的是规范基于分类代码的消息过滤业务实施。

消息过滤系统应可以从消息提取分类代码，并且利用该分类代码对相应消息进行基于消息类型的消息过滤。相应基于分类代码的消息过滤既可以在前端业务接入设备（例如ISMG、WAP GW等设备）上实施；也可以在基于用户规则消息过滤平台上实施；还可以在用户终端上实施。同时，基于分类代码的消息过滤业务实施既可适用于公共过滤（基于公共过滤规则的消息过滤），也适用于个性过滤（基于用户规则的消息过滤）。

A.1 基于分类代码的消息过滤在业务接入设备的实施

业务接入设备应该在识别消息内容分类代码的基础上，可以对不希望发送或者限制发送的消息进行阻塞，并针对不同消息业务系统使用相应的通信协议（例如短消息中的SMPP协议）通知消息发送者该群发消息被阻塞，同时附加阻塞原因。另外，前端业务接入设备可选支持消息监听能力，对于公开许可类别的群发消息可以进行监听分析。

A.2 基于分类代码的消息过滤在基于用户过滤规则消息过滤平台的实施

在基于用户过滤规则消息过滤平台中，消息过滤模块要求具备基于消息内容分类代码的过滤功能，在用户规则管理模块要求具备基于分类代码过滤规则管理手段。

基于分类信息代码的消息过滤主要是针对消息的过滤手段。基于消息分类的过滤规则可以和基于地址、关键词和时间的过滤规则结合使用。对于消息，基于消息分类的过滤规则是消息过滤的第一步。如果属于过滤范围就直接进行消息过滤处理，如果不属于消息过滤范围，则进入后续的消息过滤流程（基于地址、关键词和时间的消息过滤流程）。

具体过滤业务技术描述见YD/T 1774-2008的第7章，用户规则管理协议参见YD/T 1774-2008的第8章。对于基于用户规则群发短消息过滤的技术实现参见YD/T 1775-2008。

A.3 基于分类代码的消息过滤在用户终端的实施

主要针对具备基于用户规则消息过滤功能软件的智能手机而定，相应软件需要具备相应过滤规则管理方式并且管理界面友好。至少具备以下管理功能：

- 增加允许或者拒绝基于内容分类代码的过滤规则；
- 删除允许或者拒绝基于内容分类代码的过滤规则；
- 查询根据内容分类代码的过滤掉的消息。

基于分类代码的消息过滤在用户终端的实现有两种方式：过滤规则管理在用户终端上实现，而消息过滤在基于用户设置规则的消息过滤平台实现；过滤规则管理和消息过滤都在用户终端实现。对于第2种方式，要求终端可以更新分类代码集，并且可以基于分类代码进行消息过滤。

附 录 B

(资料性附录)

分类编码对消息容量的影响

短消息主要传输文本消息，有三种编码方式（见ETSI GSM 03.40）：7-bit编码、8-bit编码、16-bit编码。7-bit编码、8-bit编码用于发送普通的ASCII字符；而16-bit编码用于发送Unicode字符。由于短消息通常利用信令信道传输，受限于信令信道传输容量，短消息内容最多为140个字节，因此在这三种编码方式下，可以发送的最大字符数分别是160、140、70。若要发送中文，必须采用Unicode编码方式。中文短消息通常采用Unicode编码，由于分类代码加分隔符占用6个字节，故此在同一个短消息内还可以编写67个中文字符或者134个英文字符。

多媒体消息可以传输文本、声音、视频等多种媒体，通常采用SMIL格式编码。由于多媒体消息采用业务信道传输，故此传输容量较大，多媒体消息内容最多为100K个字节。但是分类代码加分隔符占用6个字节，基本上可以忽略不计。

参 考 文 献

- YD/T 1094-2000 900/1800MHz TDMA数字蜂窝移动通信网扩展短消息实体与短消息业务中心间的接口协议规范
- YD/T 1291-2003 点对点短消息网间互通协议要求
- YD/T 1364-2005 点对点短消息网间互通设备技术要求
- 3GPP TS 23.040 (2004) 短消息业务的技术实现
- GB/T 7027 标准化工作导则 信息分类编码的基本原则和方法
- GB/T 10113 分类与编码通用术语
- ETSI GSM 03.38 (1996) 欧洲数字蜂窝通信系统 (阶段2+) : 字符和特定语言信息
- ETSI GSM 03.40 (1996) 欧洲数字蜂窝通信系统 (阶段2+) : 短消息业务技术实现
-