



# 中华人民共和国国家标准

GB/T 26235—2010

---

## 信息技术 信息处理用蒙古文词语标记

Information technology—Mongolian word and expression marks  
for information processing

2011-01-14 发布

2011-05-01 实施

---

中华人民共和国国家质量监督检验检疫总局  
中国国家标准化管理委员会 发布

## 目 次

前言 .....	Ⅲ
1 范围 .....	1
2 术语和定义 .....	1
3 标记的范围 .....	2
4 标记的级别及命名方法 .....	2
5 词语标记代码 .....	2

## 前 言

本标准是依据《现代蒙古语》(1964)等语法著作为基础,规定了信息处理用的蒙古文词语标记。

本标准由全国信息技术标准化技术委员会提出并归口。

本标准起草单位:中国电子技术标准化研究所、内蒙古大学、内蒙古师范大学、内蒙古教育出版社、内蒙古自治区蒙古语文工作委员会、北京北大方正电子有限公司、潍坊北大青鸟华光照排有限公司。

本标准起草人:那顺乌日图、何正安、青格乐图、确精扎布、六十三、淑琴、贺喜格都仁、唐英敏、吕建春、阿荣塔娜。

# 信息技术 信息处理用蒙古文词语标记

## 1 范围

本标准规定了信息处理用蒙古文词语标记,只包括某种语文现象或单位的标记符号。  
本标准适用于蒙古文的各种语料库、词汇集-词典、语文知识库等。

## 2 术语和定义

下列术语和定义适用于本标准。

### 2.1

#### 蒙古文信息处理 **Mongolian information processing**

用计算机对蒙古文形、音、义等信息进行输入、排序、存储、输出、统计、提取等。

### 2.2

#### 标记单位 **tagging unit**

蒙古文信息处理中使用的词汇、词法、正字法及其他单位,如单词、复合词、固定词、习用语、成语、缩略语、构形附加成分、连接元音、字母、数字、标点符号等。

### 2.3

#### 词类 **word class**

词的语法分类,根据词法形态、句法功能、语义特征划分出来的类。

### 2.4

#### 不确定词 **ambiguous word**

无法确定词性的词或词素,包括:

- a) 无法确定词性的同形兼类词(如没有上下文环境的形式);
- b) 不能单独使用的复合词成分,如 üy(üy olan), tomsi(tog\_a tomsi ügey), say(gajar say ügey), xu(xu luubang);
- c) 形容词最高级形式:ub(ub ulagan), xab(xab xar\_a)等。

### 2.5

#### 复合词 **compound word**

由两个实词(有时也包括辅助意义的词)组成,语法和词汇上都构成一个单位的固定短语。

### 2.6

#### 固定词 **fixed word**

由一个实词和一个虚词组成,连用成型且未转义的固定短语。

### 2.7

#### 习用语 **habitual usage**

由一个转义形容词和一个实词组成,主谓结构且表达一个词汇意义的固定短语。

### 2.8

#### 成语 **idiom**

由3个或3个以上词组成,充当一个句子成分的固定短语。

### 2.9

#### 缩略语 **abbreviation**

专有名词或常用语的简略形式。