



# 中华人民共和国国家标准

GB/T 13715—92

---

## 信息处理用现代汉语分词规范

Contemporary Chinese language word segmentation specification  
for information processing

1992-10-04 发布

1993-06-01 实施

---

国家技术监督局 发布

(京)新登字 023 号

中 华 人 民 共 和 国  
国 家 标 准  
信息处理用现代汉语分词规范  
GB/T 13715—92

\*

中国标准出版社出版发行  
北京西城区复兴门外三里河北街 16 号  
邮政编码: 100045

<http://www.spc.net.cn>

电话: 63787337、63787447

1993 年 3 月第一版 2005 年 12 月电子版制作

\*

书号: 155066 · 1-9287

版权专有 侵权必究  
举报电话: (010) 68533533

# 中华人民共和国国家标准

## 信息处理用现代汉语分词规范

GB/T 13715—92

Contemporary Chinese language word segmentation specification  
for information processing

### 1 主题内容与适用范围

#### 1.1 主题内容

本规范规定了现代汉语的分词原则,以满足信息处理的需要。它对汉语信息处理的规范化,对各种汉语信息处理系统之间的兼容性有重要的作用。

#### 1.2 适用范围

本规范适用于汉语信息处理各领域,其他行业和有关学科可以参考使用。

汉语信息处理各领域可以根据其专门需求,进一步补充和细化本规范的规定。

### 2 引用标准

GB 12200 汉语信息处理词汇

### 3 术语

以下术语引自 GB 12200。

#### 3.1 汉语信息处理

用计算机对汉语的音、形、义等信息进行的处理。

#### 3.2 词

最小的能独立运用的语言单位。

#### 3.3 词组

由两个或两个以上的词,按一定的语法规则组成,表达一定意义的语言单位。

#### 3.4 分词单位

汉语信息处理使用的、具有确定的语义或语法功能的基本单位。它包括本规范的规则限定的词和词组。

#### 3.5 汉语分词

从信息处理需要出发,按照特定的规范,对汉语按分词单位进行划分的过程。

### 4 概述

本规范以信息处理应用为目的,根据现代汉语的特点及规律,规定现代汉语的分词原则。

本规范用下划线“\_\_\_\_\_”作为分词单位标记。

4.1 空格或标点符号是计算机中分词单位的分隔标记。作为分隔标记的标点符号有:句号、逗号、顿号、分号、冒号、问号、叹号、引号、括号、破折号、省略号、书名号、间隔号、连接号及符号“/”等。

4.2 二字或三字词,以及结合紧密、使用稳定的二定或三字词组,一律为分词单位。例如: