



# 中华人民共和国国家标准

GB/T 18790—2010  
代替 GB/T 18790—2002

---

## 联机手写汉字识别系统技术要求与 测试规程

Requirements and test procedure of  
on-line handwriting Chinese character recognition system

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

## 前 言

本标准代替 GB/T 18790—2002《联机手写汉字识别技术要求与测试规程》。

本标准与 GB/T 18790—2002 的主要差别如下：

- 扩大了识别字符集的范围；
- 增加了识别系统应具备的功能；
- 删除了 GB/T 18790—2002 的附录 C。

本标准的附录 A 和附录 B 是规范性附录。

本标准由全国信息技术标准化技术委员会提出并归口。

本标准主要起草单位：汉王科技股份有限公司、中国电子技术标准化研究所。

本标准主要起草人：刘迎建、陈静、钮兴昱、王欣、李鑫梅。

本标准所代替标准的历次版本发布情况为：

- GB/T 18790—2002。

# 联机手写汉字识别系统技术要求与 测试规程

## 1 范围

本标准规定了联机手写汉字识别系统的技术要求和测试规程。

本标准适用于微型计算机、手持式信息处理设备 etc 数字化设备中使用的联机手写汉字识别系统。

## 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 18030—2005 信息技术 中文编码字符集

## 3 术语和定义

下列术语和定义适用于本标准。

### 3.1

**联机手写数据采集设备 on-line handwriting device**

可将人们书写的汉字及字符轨迹实时转换成坐标点序列并形成电子数据的设备，如手写板、触摸屏、鼠标等。

### 3.2

**联机手写汉字识别系统 on-line handwriting Chinese character recognition system**

运行于微型计算机、手持式信息设备等具有计算能力的系统中，可将联机手写数据采集设备采集的汉字手写数据进行处理与辨识，输出相应的汉字标准字符的软件系统。

### 3.3

**样本 sample**

一个完整的、由联机手写数据采集设备采集的、符合附录 B 规定数据格式的数据称为一个样本。

### 3.4

**样本文件 sample file**

由多个样本按附录 B 规定的文件格式组成的文件。

### 3.5

**样本库 library of sample files**

多个同类型样本文件组成的文件集合。

### 3.6

**工整样本库 library of printed-handwriting sample**

由书写者使用联机手写数据采集设备、在工整书写提示下自然书写的汉字样本组成的样本库。

### 3.7

**乱笔顺样本库 library of random-stroke-order sample**

以随机方式打乱了笔划顺序的汉字样本组成的样本库。

### 3.8

#### 自由手写体样本库 library of cursive-handwriting sample

由书写者使用联机手写数据采集设备、在自由书写提示下按照自身书写习惯自然书写的汉字样本组成的样本库。

### 3.9

#### 平均首选识别率 average accuracy of hit-in 1

首选正确的样本数与样本库中样本总数的比值。

注：测试过程中，如识别系统对被测样本给出的首个候选字是正确结果，则该样本为首选正确的样本。

### 3.10

#### 平均十选识别率 average accuracy of hit-in 10

十选正确的样本数与样本库中样本总数的比值。

注：测试过程中，如识别系统对被测样本给出的前10个候选字中包含正确结果，则该样本为十选正确的样本。

## 4 技术要求

### 4.1 识别字符集的范围

#### 4.1.1 汉字字符

联机手写识别系统识别的汉字字符集应为 GB 18030—2005 中的全部汉字，或至少包括 GB 18030—2005 强制性部分中的全部汉字。

#### 4.1.2 非汉字字符

联机手写汉字识别系统识别的非汉字字符应包括数字、拉丁字母、西文标点和符号、中文标点和符号，以及手写控制符，见附录 A。

### 4.2 识别率

#### 4.2.1 对工整样本库的识别要求

- a) 对 GB 18030—2005 中强制性部分汉字的平均首选识别率应大于 95%；
- b) 对附录 A 中的非汉字字符的平均首选识别率应大于 80%；
- c) 对支持 GB 18030—2005 中的全部汉字字符的产品，平均首选识别率应大于 85%。

#### 4.2.2 自由手写体库识别要求

对自由手写体样本库的平均十选识别率应大于 90%。

#### 4.2.3 对乱笔顺样本库的识别要求

对乱笔顺样本库的平均十选识别率应大于 80%。

### 4.3 功能要求

#### 4.3.1 候选字

联机手写识别系统应具备候选字功能。

#### 4.3.2 词组联想功能

联机手写识别系统应具备词组联想功能。

#### 4.3.3 用户笔迹自适应

联机手写识别系统应具备用户笔迹自适应功能。即，当用户确认正确的识别结果出现在候选字中（非首选）时，识别系统可以自动学习当前对应的用户笔迹，使得当用户以后再次输入同一笔迹时，正确结果可以出现在首选结果中。

#### 4.3.4 倾斜字识别

联机手写识别系统应具备倾斜字识别功能。即，给定屏幕内的垂直坐标为 0° 方向，识别系统应能正确识别 ±45° 以内倾斜方向书写的用户笔迹。

#### 4.3.5 混合识别

联机手写识别系统应具备识别中、英、数、符混合输入的功能。即，当用户手写输入汉字、英文字母、

数字、符号时,用户不需要设定或改变当前识别范围,识别系统就可以识别所有笔迹。

#### 4.3.6 简(体)繁(体)转换

联机手写识别系统应具备写繁出简的功能,即当用户书写笔迹为繁体字时,系统自动输出为对应的简体字识别结果;同时应提供选项设置,以实现写简出繁、写繁出繁。

### 5 测试规程

#### 5.1 测试样本库

测试样本库包括工整样本库、乱笔顺样本库和自由手写体样本库。

#### 5.2 测试样本库的建立

##### 5.2.1 工整样本库的建立

使用联机手写数据采集设备,采集由一定数量的书写者按照笔顺规范工整书写的汉字及符号的笔迹数据。

数据套数不应少于 30 套,书写者的数量不应少于 30 人,同一书写者不能重复书写相同字符。

应选取不同年龄、不同性别、不同文化程度的书写者。

每套数据的内容,应包含 GB 18030—2005 中全部或强制性部分包含的所有汉字,及附录 A 中列出的所有字符。

##### 5.2.2 笔顺样本库的建立

将工整样本库中的样本,以随机方式打乱笔顺后形成乱笔顺样本库。

##### 5.2.3 自由手写体样本库的建立

使用联机手写数据采集设备,采集由一定数量的书写者按照自身书写习惯书写的汉字及符号的笔迹数据。

数据套数不应少于 30 套,书写者的数量不应少于 30 人,同一书写者不能重复书写相同字符。

应选取不同年龄、不同性别、不同文化程度的书写者。

每套数据的内容,应包含 GB 18030—2005 强制性部分中双字节编码 2 区包含的所有汉字,及附录 A 中列出的所有字符。

#### 5.3 测试方法

##### 5.3.1 识别字符集的范围

使用被测联机手写识别系统,对全部工整样本进行识别,验证系统所能识别的汉字字符及非汉字字符,应符合 4.1 的要求。

##### 5.3.2 识别率

使用被测联机手写识别系统,分别对全部工整样本和乱笔顺样本进行识别,记录每个样本中每个字符的识别结果,并按下列公式计算平均首选识别率和平均十选识别率。

$$\text{平均首选识别率} = \text{CN}/\text{NN} \quad \dots\dots\dots(1)$$

式中:

NN——样本库中样本总数;

CN——经统计首选识别结果正确的样本数。

$$\text{平均十选识别率} = \text{CN}_{10}/\text{NN} \quad \dots\dots\dots(2)$$

式中:

NN——样本库中样本总数;

CN<sub>10</sub>——经统计前 10 选识别结果中包含正确结果的样本数。

##### 5.3.3 功能要求

采用被测联机手写识别系统进行实际输入,判定是否具备 4.3 要求的各项功能。

## 附录 A (规范性附录)

### 联机手写汉字识别系统的识别字符集中的非汉字字符集

联机手写汉字识别系统至少应识别如下非汉字字符：

#### A.1 数字

0 1 2 3 4 5 6 7 8 9

#### A.2 大写英文字符

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

#### A.3 小写英文字符

a b c d e f g h i j k l m n o p q r s t u v w x y z

#### A.4 西文标点、符号

! " # \$ % & ' ( ) \* + , - . / : ; < = > ? @ [ \ ] ^ \_ ` { | } ~ € £ ¢





#### A.5 中文标点、符号

！ ， ， 。 、 ： ； “ ” —……～ ( ) { } < > 《 》 ¥

#### A.6 手写控制符

手写控制符与内码对应关系见表 A.1。

表 A.1 手写控制符与内码对应关系

名称/Name	内码/Code	手写控制符/Figure
退格/Backspace	0x0008	
制表/Tab	0x0009	
回车/Return	0x000D	
空格/Space	0x0020	
撤销/Undo(Ctrl+Z)	0x001A	~Z
拷贝/Copy(Ctrl+C)	0x001B	~C
粘贴/Paste(Ctrl+V)	0x001C	~V
剪/Cut(Ctrl+X)	0x001D	~X
Tab、Return、Space 书写时均要求宽度大于高度。 “~”要求一笔写成，不能写成两笔的“八”字。 注：图中的箭头只表示书写方向，不需要书写。		

## 附录 B

## (规范性附录)

## 关于标准测试样本文件的规定

本附录规定了标准测试样本库中的样本文件的扩展名和文件格式。

本附录对于提供和使用联机手写汉字识别系统的各方提出共同遵从的约定。

## B.1 标准测试样本文件的扩展名

标准测试样本文件的扩展名统一为“.POT”。

## B.2 标准测试样本文件格式

文件格式:标准测试样本文件中存放的是连续的 POT 数据块;无其他额外的格式信息。

POT 数据块定义:每个 POT 数据块中包含一个手写汉字的字型采样数据和内码等信息,详见下表。

位移	内容
0 WORD	数据块长度,以字节(BYTE)为单位,假设为 n
2 DWORD	本数据块所代表的汉字内码,例如“啊”= 0x0000b0a1
6 WORD	手写样本的总笔划数,假设为 N
8 WORD	第一笔的第一个点的 X 方向坐标值
10 WORD	第一笔的第一个点的 Y 方向坐标值
...	
i WORD	0xFFFF,第一笔的结束标志
i+1 WORD	0x0000,第一笔的结束标志
...	
n-7 WORD	0xFFFF,第 N 笔的结束标志
n-5 WORD	0x0000,第 N 笔的结束标志
n-3 WORD	0xFFFF,本字的结束标志
n-1 WORD	0xFFFF,本字的结束标志

注:1 WORD=2 BYTE(低字节在前,高字节在后),例如:0x1234,在文件中的字节排列顺序为 34H,12H;

1 DWORD = 2 WORD(低字在前,高字在后),例如:0x12345678,在文件中的字节排列顺序为 78H,56H,34H,12H。

中 华 人 民 共 和 国  
国 家 标 准  
联机手写汉字识别系统技术要求与  
测试规程

GB/T 18790—2010

\*

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

网址 [www.spc.net.cn](http://www.spc.net.cn)

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷  
各地新华书店经销

\*

开本 880×1230 1/16 印张 0.75 字数 11 千字  
2011年5月第一版 2011年5月第一次印刷

\*

书号: 155066·1-42649 定价 16.00 元



GB/T 18790-2010

如有印装差错 由本社发行中心调换  
版权专有 侵权必究  
举报电话:(010)68533533