

ICS 35.240.20
L 67



中华人民共和国国家标准

GB/T 17961—2000

印刷体汉字识别系统要求与测试方法

Requirements and test methods for identify
system of printed Chinese character

2000-01-03 发布

2000-08-01 实施

国家质量技术监督局 发布

前 言

本标准是非键盘输入标准,它规定印刷体汉字识别系统的功能、技术要求、测试方法,该标准的制定和实施将正确引导印刷体汉字识别系统的研制与开发。

本标准的附录 A 是提示的附录。

本标准由中华人民共和国信息产业部提出。

本标准由中国电子技术标准化研究所归口。

本标准起草单位:“八六三”计算机智能主题专家组。

本标准主要起草人:顾小凤、丁晓青、刘迎建、钱跃良。

**Requirements and test methods for identify
system of printed Chinese character**

1 范围

本标准规定了印刷体汉字识别系统的功能、技术要求和测试方法。
本标准适用于印刷体汉字识别系统,也适用于其他有关系统。

2 引用标准

下列标准所包含的条文,通过在本标准中引用而构成为本标准的条文。本标准出版时,所示版本均为有效。所有标准都会被修订,使用本标准的各方应探讨使用下列标准最新版本的可能性。

- GB 2312—1980 信息交换用汉字编码字符集 基本集
- GB/T 7589—1987 信息交换用汉字编码字符集 第二辅助集
- GB/T 7590—1987 信息交换用汉字编码字符集 第四辅助集
- GB/T 12345—1990 信息交换用汉字编码字符集 辅助集
- GB 13000.1—1993 信息技术 通用多八位编码字符集(UCS) 第1部分:体系结构
(idt ISO/IEC 10646.4:1993)
- GB 13131—1991 信息交换用汉字编码字符集 第三辅助集
- GB 13132—1991 信息交换用汉字编码字符集 第五辅助集

3 系统功能

3.1 基本功能

3.1.1 预处理

将普通纸媒体文本在自动或人工干预下通过光学输入设备将印刷信息转换成适合识别的灰度(含二值)图像。

3.1.2 识别

将上述图像所包括的印刷符号转换为 GB 2312、GB/T 7589、GB/T 7590、GB/T 12345、GB 13000.1、GB 13131 和 GB 13132 所规定的编码信息。

3.2 扩展功能

3.2.1 版面分析

应将版面自动分成块,并正确标明每个块的属性,对文字块还需标明各块之间连接关系的逻辑序号。

3.2.2 表格识别

应能正确识别表格线,并可将表格和文字建立对应关系。

4 要求

4.1 识别率

对于正式出版物及打印质量与其相当的打印文件的识别率应大于 95%。

4.2 误识率

对于正式出版物及打印质量与其相当的打印文件的误识率应小于 2%。

4.3 识别速度

使用 6.3 版本,识别速度应大于 10 字/秒。

5 测试方法

5.1 测试样张

测试时参照采用附录 A 给出的测试样张,在实际测试时,有关测试机构可依据样张的字体、字号和版面布局随机提供多种测试样张,测试样张至少应包含 103 个印刷符号。

5.2 识别率、误识率测试

5.2.1 测试设备

应采用通用操作系统及与其相适应的微型数字电子计算机。

5.2.2 测试样张

测试样张采用 5.1 规定的样张。

5.2.3 测试结果

识别率测试结果按式(1)计算:

$$\text{识别率} = (C/N) \times 100\% \quad \dots\dots\dots(1)$$

式中: C ——测试样张中被正确识别的印刷符号数;

N ——测试样张中印刷符号总数。

误识率测试结果按式(2)计算:

$$\text{误识率} = (E/N) \times 100\% \quad \dots\dots\dots(2)$$

式中: E ——测试样张中被错误识别的印刷符号数;

N ——测试样张中印刷符号总数。

测试结果应符合 4.1 和 4.2 的要求。

5.3 识别速度测试

5.3.1 测试设备

采用 5.2.1 规定的测试设备。

5.3.2 测试样张

测试样张采用 5.1 规定的样张。

5.3.3 测试结果

识别速度测试结果按式(3)计算:

$$\text{识别速度} = N/T \quad \dots\dots\dots(3)$$

式中: N ——测试样张中印刷符号总数;

T ——识别系统从开始读取测试数据至将识别结果记录到媒体上所用的时间。

识别速度应符合 4.3 的要求。

附录 A
(提示的附录)
测试样张

<p>客中作 李白</p> <p>兰陵美酒郁金香， 玉碗盛来琥珀光。 但使主人能醉客， 不知何处是他乡。</p>	<p>兰溪棹歌 戴叔伦</p> <p>凉月如眉挂柳湾， 越中山色镜中看。 兰溪三日桃花雨， 半夜鲤鱼来上滩。</p>	<p>春望 杜甫</p> <p>国破山河在， 城春草木深。 感时花溅泪， 恨别鸟惊心。 烽火连三月， 家书抵万金。 白头搔更短， 浑欲不胜簪。</p>
<p>宣州谢朓楼饯别校书叔云 李白</p> <p>？我去者， 昨日之日不可留。 乱我心者， 今日之日多烦忧。 长风万里送秋雁， 对此可以酣高楼。 蓬莱文章建安骨， 中间小谢又清发。 俱怀逸兴壮思飞， 欲上青天揽明月。 抽刀断水水更流， 举杯销愁愁更愁。 人生在世不称意， 明朝散发弄扁舟。</p>		
<p>将进酒 李白</p> <p>君不见黄河之水天上来， 奔流到海不复回。 君不见高堂明镜悲白发， 朝如青丝暮成雪。 人生得意须尽欢， 莫使金樽空对月。 天生我材必有用， 千金散尽还复来。 烹羊宰牛且为乐， 会须一饮三百杯。 岑夫子，丹丘生， 将进酒，君莫停。 与君歌一曲，</p> <p>请君为我倾耳听。 钟鼓馔玉不足贵， 但愿长醉不复醒。 古来圣贤皆寂寞， 惟有饮者留其名。 陈王昔时宴平乐， 斗酒十千恣欢谑。 主人何为言少钱， 径须沽取对君酌。 五花马，千金裘， 呼儿将出换美酒， 与尔同销万古愁。</p>		