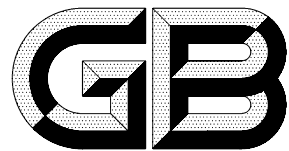


ICS 35.240.30
L 67



中华人民共和国国家标准

GB/T 18790—2002

联机手写汉字识别技术要求 与测试规程

Requirements and test procedure of
on-line handwriting Chinese ideogram recognition

2002-07-18 发布

2002-12-01 实施

中华人民共和国
国家质量监督检验检疫总局 发布

前 言

本标准规定了联机手写汉字识别系统的汉字识别技术要求、测试规程。该标准的制定和实施将规范联机手写汉字识别系统的研究、开发和应用。

本标准的附录 A 和附录 B 是标准的附录,附录 C 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:中国电子技术标准化研究所、北京汉王科技有限公司、清华大学电子工程系。

本标准主要起草人:刘迎建、王立建、张立清、刘长松、钮兴昱、王宝艾。

中华人民共和国国家标准

联机手写汉字识别技术要求与测试规程

GB/T 18790—2002

Requirements and test procedure of on-line handwriting Chinese ideogram recognition

1 范围

- 1.1 本标准规定了联机手写汉字识别系统的汉字识别技术要求和测试规程。
- 1.2 本标准适用于微型计算机、手持式信息处理设备和数字化电器配置的联机手写汉字识别系统。

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,所有版本均为有效。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 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 10646:1993)
- GB 18030—2000 信息技术 信息交换用汉字编码字符集 基本集的扩充

3 定义

本标准采用下列定义。

- 3.1 联机手写数据采集设备 on-line handwriting data capture device
是指鼠标、手写板、触摸屏等具备实时地将人们书写的汉字及字符轨迹转换成坐标点序列、形成电子数据的设备。
- 3.2 联机手写汉字识别系统 on-line handwriting Chinese ideogram recognition system
是指使用微型计算机或具有计算能力的系统,对联机手写电子数据采集设备采集的手写电子数据进行处理与辨识,获得相应的标准内码的系统。
- 3.3 样本 sample
一个完整的、由联机手写电子数据采集设备采集的、符合第4章中的汉字或字符的电子数据称为一个样本。
- 3.4 样本文件 sample file
是指由多个样本按附录B数据格式组成的文件,称为样本文件。
- 3.5 样本库 library of sample file
多个同类型样本文件组成的文件集合称为样本库。

4 识别字符集的范围

联机手写汉字识别系统识别的最小字符集应是GB 2312中全部汉字字符(包括偏旁部首)以及附录A中的非汉字字符。字符扩展时,联机手写汉字识别系统应识别GB 12345,或GB 18030,或GB