

汉字信息处理系统资料汇编

(苏 州 会 议)

中国仪器仪表学会

汉字信息处理系统研究会

一九八一年

目 录

关于报纸自动编排系统的设想.....	叶心力 王振锋 张书杰 (1)
谈谈我馆对汉字信息处理的一些希望和要求	北京图书馆小计算机组 (9)
汉字拼音键盘的设计.....	杭州自动化研究所 盛 谏 (15)
中文信息处理系统的实现.....	上海计算技术研究所 王景寅 (25)
管理汉字终端网络系统设计中的几个问题	一机部机械工业自动化所 周小川 (29)
汉字相关压缩技术的研究.....	中国印刷科学技术研究所 周小鹤 (36)
上海市汉字信息处理系统	上海市汉字信息处理系统领导小组办公室 (48)
汉字显示终端的实践和改进方案	一机部上海工业自动化仪表研究所 章万理 周祝龄 王双球 成伟明 (52)
MONOTYPE——3000中文激光照排系统简介	上海印刷技术研究所 张佑陵 (71)
磁盘存储器的发展和应用.....	上海电表厂 (79)
双拼对字表.....	一机部自动化研究所 抚良文 (92)
汉字信息处理系统.....	一机部自动化所 刘卫民 (107)

关于报纸自动编排系统的设想

叶心力、王振铎、张书杰

报纸自动编排是当前报纸印刷技术的主要发展趋势，是衡量一个国家报纸印刷水平的重要标志，七十年代以来，一些技术先进的国家已先后采用了新闻自动编排技术。根据我国情况，发展自动编排系统取代目前的手工操作，以适应新闻事业的发展需要也是势在必行。报纸的出版印刷与一般书刊、杂志出版印刷有明显差异，其主要特点如下：

1、稿件繁多，种类复杂。报社每日要收到大量的新闻稿件、电讯、报导、图片、传真照片、读者来信、来稿、广告等等，需要及时处理，其文字处理量和管理工作极为繁重。

2、政策性强，影响面大（尤其在我国的更是如此）。因此对新闻稿件可靠性、准确性要求严格，以致稿件、图片删改反复次数很多。

3、时间性强，审稿定稿时间集中、短促，加之发行量大，有些大量报纸还有外地印刷问题，外地印刷点的样张及纸型需及时发出，给编辑排版工作造成困难。

鉴于以上特点，为了提高报纸质量，缩短出版时间，必须采用先进技术以改变目前的落后状态。我们认为，如何在短促时间内分析处理大量的稿件，迅速准确地进行编辑排版是报纸编排系统的核心问题，而不仅仅是要用自动照排机照排以代替铅字的问题。为此，我们根据国外一些报馆的自动编排系统的情况和我国报社的实际工作状况，拟订了一个报纸自动编排系统，现介绍如下：

一、系统的基本要求

1、所设计的编排系统可代替现行的通讯接收、稿件的管理编辑、校对和排版功能，并进一步改进通讯、编辑、印刷部门的工作方式和方法。

2、选择适合的输入方式（包括输入编码、键盘功能），使稿件的传输、编排管理工作能高效能地进行，而且易于训练，操作方便。

3、有足够的稿件存贮容量，并能快速而方便的检索稿件。

4、准确、快速、灵活，以期能在最短时间内产生正确的修改版以便付印，而且具有不低于铅字浇注法的灵活性。

5、系统可靠，维护简便。应配有能力较强的故障诊断系统。

二、系统的功能

为完成上述任务,该系统应是一个大型多级系统,有较强的软件功能,编排语言力求简明易懂,最好采用汉语拼音或直接用中文信息,系统应有充分扩充余地。

1、新闻稿件的接收处理

报社每日从各种渠道汇集的稿件多达几十万字以上,对这些稿件应分别处理。

①从通讯社来的电讯稿,经通讯控制计算机作汉字编码处理后送磁盘存放。同时打印复制送编辑部门审查,填写登记表后交回通讯计算机打字员。如果该稿录用,由通讯计算机将该表和相应于磁盘中的稿件一并送到主计算机有关稿件库存放。

通讯计算机仅起暂存作用,一旦稿件送主机后即自行清除。

②非电讯稿:如记者采访报导,特约文章、资料、来信来稿等,采用脱机输入。稿件先由编辑部门审阅,同样填写稿件登记表,而后一并交穿孔员按汉字编码穿成纸带或存入软盘,经检查无误后,送主机系统的纸带输入机或软盘输入机,以存入有关稿件库。

③稿件登记和分类稿件库

稿件登记表一般应填写编号、类别、标题、作者、来源、收到日期、批准人、处理意见(保留、刊登、清除)、字数、优先数、内容摘要等。

分类稿件库分待登稿件库和备用稿件库,备用稿件库指短期不用的稿件,一般存放在磁带中。对于大型报纸而言,备用版面至少应有50版左右。

稿件管理系统应根据登记表的要求按类别、优先数等将稿件分别放入相应的稿件库中,并能根据登记录的编号、作者、标题等进行检索调用。

2、编辑校改系统

一旦稿件被送交待登稿件库,不论是复制的,还是待加工的稿件,都要经过编辑校改处理,以完成编辑人员所应做的一切工作。因此系统应具有以下功能:

①稿件流通:通过编辑校改终端,编辑能灵活地调动稿件到显示屏幕,或从一个库转到另一个库,从一种状态转到另一种状态,稿件调动变更后,系统能修改登记录,产生最新登记录。

②为实现终端修改功能,该终端键盘应具备完善的游标键、传送功能键、修改编辑功能键,以对字符、字段进行增、删、改、调整等操作控制。编辑校改完成后,该稿件被送回稿件库以便排版使用。

3、版面设计系统

报纸版面格式多种多样,灵活性强,标准与非标准格式有几百种之多,编辑人员根据新闻的重要性、文章的长短等还可以随机修改。为适应高峰期间和临时修改版面的需要,拼一个版面的时间不应超过20分钟。因此,应有较强的版面编排功能,主要有:

①版面编辑能将所需稿件调到版面编辑终端屏幕指定位置。正文版面编排软件应按指定

的字体、字号、栏数、字距、横竖等格式要求将正文正确排列。屏幕显示时，每个汉字只用几个光点表示。

②标题处理软件根据编辑确定的位置大小、体字、字号、格式等标题说明对标题、付标题、作者名等进行处理并显示，如有可能应以显示字库中的汉字将标题显示出来，大字标题以方格代替，中间有汉字。

③图片广告处理，根据编辑确定的图片广告，从图片广告库调出图片广告到指定位置，但屏幕上不显示图片广告内容，只以边框代替，框内显示图片名、广告名（不作照排用），同时要按有关规则对图片标题、说明、作者名等进行处理。当图片比例不合适时，可以通过图片管理系统作修改。

④花边、边框等处理时在屏幕上只显示直线。

⑤诗歌、表格、英文、数字、特殊符号应有专用模拟处理。

⑥版面编辑系统应有足够的灵活性。例如：在指定位置放不下时应有转版处理功能。

版面编排完成后，该版面内容和照排信息应送入主机专用磁盘存放，等待照排。按八版计算，该磁盘应能容纳10万汉字。同时，应转入复制流程，制出大样以供校对。

4、广告编排管理系统

考虑到今后分类广告数量日益增加的情况，并根据我国分类广告特点和发展趋势，在拟定分类广告系统方案时，我们认为在可以预见的将来，为广告用户配备终端是不现实的，用户难以在终端实现输入、修改、清除、复制等操作。我们仍沿用现行的方式方法来接收广告，但要求接收以后的工作由系统进行处理，以改变目前的手工方式，适应广告的发展需要。其主要功能要求是：

①通过广告输入终端将广告文件和内容打入软盘或穿孔，并打出校样，经检查校对后通过广告处理终端送入主机的分类广告数据库存放。

广告合同签订后，广告接收员应制定出该广告的广告文件，至应包括以下内容：

总编号、合同号、单位、联系人、电话号、账号、地址、接收人、批准人、广告类型（如文娱、体育、工商、国外、其它等）、尺寸系列、格式、登记日期、刊登天数、优先数、内容摘要等。

所有广告均按广告文件存放，该管理系统根据类型是优先数、日期等因素自动按顺序排队。

②该广告管理系统可通过广告编辑终端对广告进行编排操作处理，该终端应具有以下功能：

a、广告编辑可以把所需的任意广告从分类广告数据库调用到广告编辑部终端屏幕，也可以按优先数顺序调用。

b、对调到屏幕的广告进行删、增、替换、移动等修改操作，并确定字体、字号、行距、格式等。

c、广告清除后，系统应对现存广告重新排队，以改变见根日期。

d、当检查出错误需要用户更改时，系统自动打印显示用户名、电话、地址、错误内容等，以便通知用户更改，该广告随之转入错误处理工作队列。

e、修改完成后,在该系统控制下进入复制流程进行复制,并送回分类广告库。

f、该系统能根据分配的版面按优先数等调出一次照排的广告到广告编辑终端,以进行版面编排设计。完成后应将该次照排信息和内容送主机专用磁盘存放,以便照排。

根据上述要求,该广告编辑终端应具有完善的控制字符运动的游标键和修改、编辑功能键。

③该系统能根据广告尺寸系列、天数、优先数等进行自动定位、记账,对长期用户可以记月账。系统可自动打出账单或月账单,提供会计部门作为结算和开发票的依据。

根据结账情况,由会计部门提供数据给广告管理系统,以便对广告作信用检查,如产生信用危机,该项广告将被送入信用危机队列处理。当进行信用检查时,系统在终端屏幕显示用户名,电话,应交款项,拖欠日期等,并打印欠款通知单。

④产生管理日报表,供广告接收员在接收安排新广告时的参数。该报表应有以下内容:每日接收的分类广告数总尺寸和累计数总尺寸;当日刊登广告数;取消广告数和待登广告数尺寸。

4、图片处理系统

解决报纸的自动编排问题,除了解决新闻稿件和广告编排外,还必须考虑到图片的处理。否则整个报纸的编排仍然达不到及时、准确的要求。为此,在图片处理方面我们要求做到:

① 输入采用图形数字化仪,将图片按规定系列确定坐标并予以数字化,然后经图形处理终端送交主计算机图片库储存、处理。另外,展示广告也按图片处理。

每张图片均应附有图片文件随同输入,以便随时调用。该件主要包括:编文号、类别、尺寸、来源、作者名、接收日期、批准人、内容说明等。

② 通过图形编辑终端可对图片编排处理,其主要功能要求是:

a、能按编号或作者名等,从图片库中调出图片或图片的某部分列终端屏幕。

b、能任意切割、分割图片,以去掉不需要的部分。

c、可按规定系列放大或缩小图片。

d、对美术字体可按比例缩小、放大、变斜、拉长、压扁等。

e、可将图片中某一图形移动到任一位置(展示广告中这种情况较多)。

f、将修改好的图片送回图片库。

g、可进行版面图片编排,完成的版面和照片信息可直接送主机有关磁盘存放,以便照排。

③ 产生管理报告,内容有当日收到各类图片数及目录;现存图片数及目录;今日刊登图片及取消图片数和目录。

5、照排输入

文章组版后,操作人员即可打入照排控制命令,通知照排系统开始照排,照排输出的结果直接关系到报纸的制版、印刷质量,因此,照排系统应具有:

①较大的数字化字模库。汉字应有四种字体(宋、仿宋、黑体、楷体),字号从6P~

15P。标题字需要字体更多（除上述4种外，再加上墨黑、长体、扁体、扁黑、长黑、棣体等），字号从5P~72P（关于字号可按国家出版总局颁布的照排级别标准）。此外，还有英、法、俄、德、日、西班牙、希腊等文字字母，以及阿拉伯数字、罗马数字、歌曲简谱和一些常用的科技符号。这样字模库总容量可在三十万字左右，迁到字库中没有的字，可随时写入，或与手动照排机配合使用。手动照排机主要供照排特大字号和一些特殊符号用。

② 全页或照排输出，输出的行宽要可变，最大行宽为400mm（整个版面为400mm×560mm），扫描线密度在25线/mm以上，照排速度不低于5分钟/版。

③ 能输出图形，也能输出图片和线画类。能输出软片，也能输出印象纸。

6、管理功能

主要为管理人员提供整个工作流程有关的数据，以便合理的管理调度系统。这些功能有分散在分系统之中。

① 日志（包括年、月、日，系统通讯、操作员通讯）。

② 终端记录：记录每一终端接收和发送的每一稿件的题目和时间。

③ 编辑部记录，该系统给每一个编辑部门保持一个记录，该记录按日常逻辑顺序记录每一稿件并分类。

工作流程见图1

三、系统组成

该系统分为三级，六部分。通讯处理系统，稿件校改、稿件输入系统，广告系统，图片处理系统为第一级，此四个系统能并行工作；第二级为主机系统，第三级为照排系统。（整个系统框图见图2）

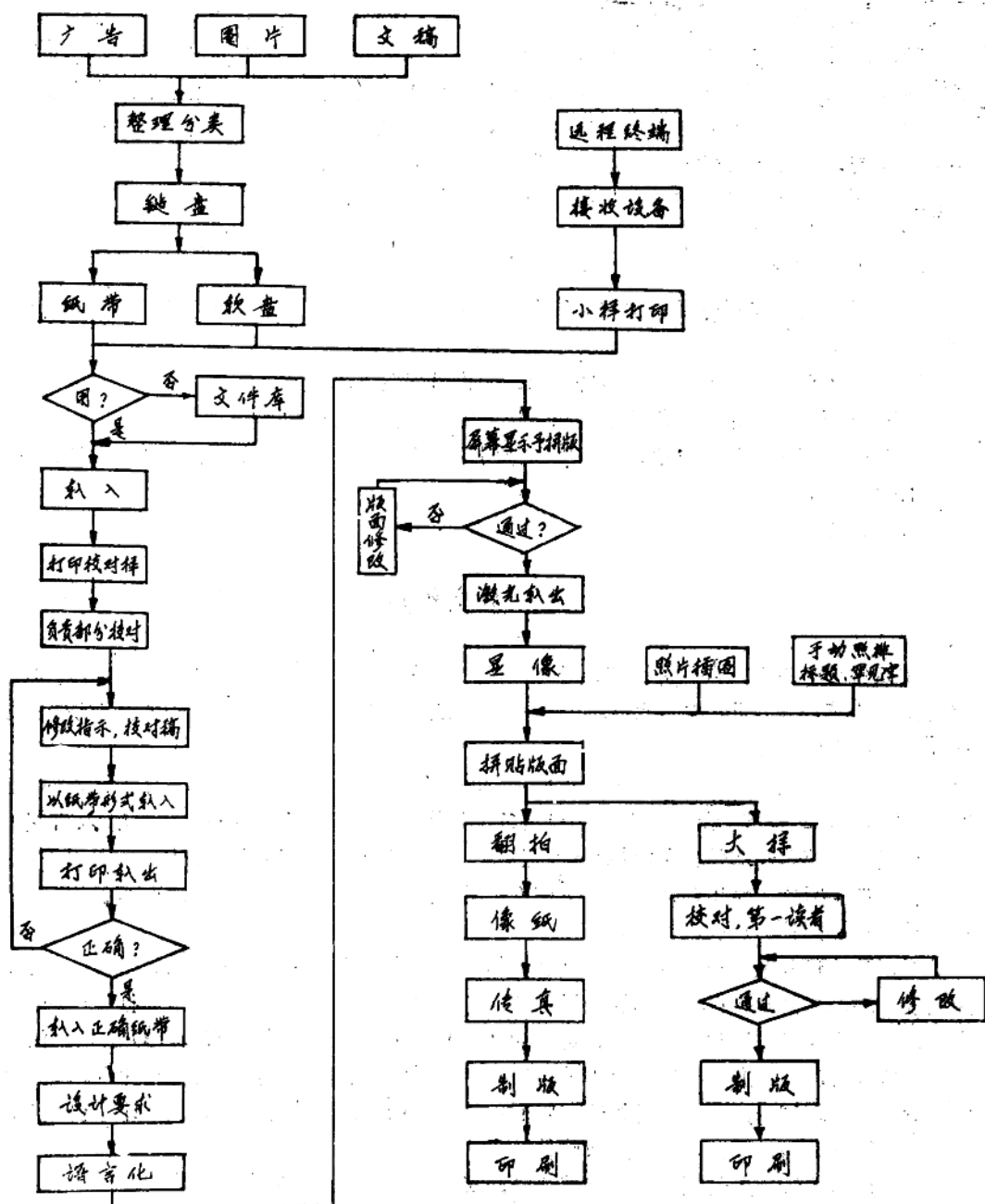
1、主机系统，由一台大中型计算机为核心，用以协调控制整个系统工作。该机内存应在2MB以上，有完善的外围机通讯接口设备。此外，应配备纸带输入机、软盘输入机各两台，用以输入稿件和软件。印字机2台，磁盘4台，每台容量在200MB以上，用以存放系统软件、稿件、图片、专用软件和文字字模（字模库容量要在500MB以上）。磁带机4台，用以存放稿件、图片、广告和软件。

2、稿件输入及编排系统。有小型计算机一台，用以控制各输入终端和编辑校改终端，版面设计终端的分时操作管理和汉字显示功能等。汉字发生器用以提供显示及打印用的文字点阵，每个字的点阵按20×20计算，容纳8000~9000个字需用500K字节以上的容量。汉字输入装置50台，编码方式可采用大中键盘结合型，用脱机输入方式，配备软盘和纸带机。汉字键盘显示终端25台，用于编排校改。版面设计终端10台，应带有光笔。汉字印字机2台，用以输出校样。磁盘一台，用以存放软件。

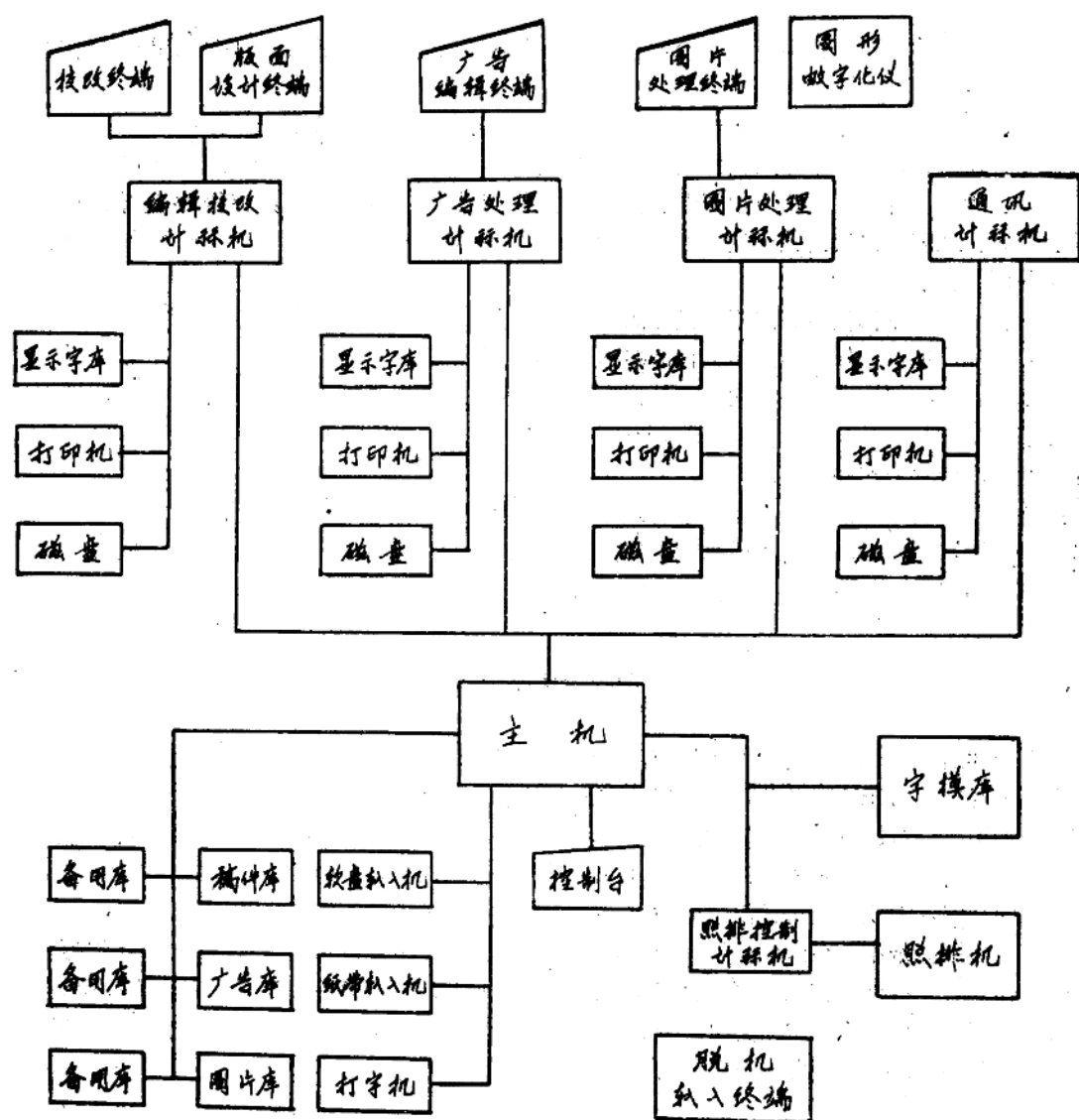
3、广告系统。广告版面设计终端2台，汉字印刷机一台，磁盘一台。

4、图片系统。图片处理终端2台，图形数字代仪一台，汉字印刷机一台，磁盘一台。

5、通讯处理系统。有小型计算机一台，用以控制远程数据传输、转换，汉字印字机二台，磁盘二台，每台容量50MB。此外，应为今后远程终端设置留有接口余地。



图一 工作流程图



图二 系统组成简图

6、照排系统：全页式激光照排机二台，照排控制小型计算机一台，其内存容量不小于64KB，以存放当前照排用的信息及控制软件。

四、主要软件

1、系统管理程序。用以协调处理整个系统的运行操作、分时管理、作业控制和系统监测。

2、输入处理程序。将汉字输入码变为内部码，并记入软盘或穿成纸带。

3、校改程序。用于输入稿件的显示、修改，并将修改好的稿件送有关磁盘存放。

4、版面设计程序。用以把图形、标题、稿件等调到版面指定区域，该程序对标题、行、段进行处理，以填满该区域。如不适合可以通过变换行距、转栏、转版、竖排等进行调整，使版面完善。

5、字模库维修程序。用于字模库的增、修、改、清除以及对盘外汉字的处理。

6、稿件库管理软件，用于稿件的传送、存放、清除和检索。

7、图片处理软件。包括图形数字化输入、调出、显示、修改、比例变换和图片库管理等程序。

8、广告管理软件。

9、照排控制软件。包括照排控制命令处理，版面设计信息的转换控制。

10、通讯处理程序。实现远程数据的传送控制，把交换码变为内部码，并存入磁盘。

11、输出处理程序，用于校样的打印输出。

五、结束语

对上述所设想的报纸自动编排系统，我们只是粗略地进行了描述，因为要实现这样的系统是不容易的。对于这样的系统，不但要有成套的硬设备，更要有完善的规模巨大的软件系统。我们不仅要解决汉字的处理和照排问题，更为艰巨的是要解决稿件的管理、编辑、排版等问题。在欧美一些国家的报馆，其编排系统的规模相当巨大。其原因就是具有处理各式各样稿件的功能。在我国，对于汉字的输入、存贮、再生等技术的研究已取得了很大进展，这给进一步研制报纸编排系统创造了一个良好的条件。但是，要完成这一系统的研制，还有大量的工作要做（特别是在稿件管理、编辑排版软件方面），其任务是相当艰巨的。我们希望有关领导部门，应组织各方面力量，分工协作，联合起来，早日研制成我国的报纸编排系统，尽快改变我国报纸行业的落后面貌。

谈谈我馆对汉字信息 处理的一些希望和要求

北京图书馆电子计算机组

大家知道,图书馆事业是文化科学事业的一部分,它在人类社会的发展中,起着十分重要的作用。革命导师列宁,甚至西方一些资产阶级政治家,都把图书馆事业的建设,看成是衡量一个国家文化水平的一个标志。作为图书馆,其主要功能表现在传递科学情报,为生产科研服务,社会教育和保存文化遗产等几个方面。在图书馆所收藏和浩如烟海的书刊资料中,既凝结着前人几千年来通过实践总结出来的宝贵经验,也反映着当今世界科学技术发展的现状。正因为如此,人们把科技情报和能源、材料并称为发展科学技术的三大支柱,把图书情报资料看成是无形的财富。诸位都是生产、科研等方面的技术人员,不少人也与图书馆有着不解之缘。对图书馆的作用一定深有体会。

图书馆事业的发展有助于科学技术的发展。科学技术的发展则对图书馆工作提出了更高的要求。由于图书资料的剧增,图书馆工作中原有的手工方式已远远不能适应大量增长的图书资料的搜集加工,传播等方面的需求。因而用现代化的手段来装备图书馆,特别是省市以上大型图书馆,已是迫在眉睫的事了。西方国家自50年代起,便开始将计算机应用于图书馆工作中,使图书馆的采购、编目、期刊管理、借阅、流通、参考咨询、文献检索等工作逐步实现自动化,提高了工作效力,加快了图书情报资料的传播,从而对人类社会的发展起了更大的促进作用。但在我国,图书馆工作的自动化,差不多还没有起步,即便采用计算机或是准备利用计算机的单位,大多数的还是局限于西文。因为汉字信息处理问题对大家来说还是一个难题。而中国的图书馆中的许许多多的工作环节,又无不牵扯到汉字信息的问题,如果计算机不能处理汉字信息,那么图书馆的自动化也就无从谈起。所以我们始终认为:汉字信息处理与计算机相结合,是我们图书馆实现自动化的一个基本条件。

下面我们根据北京图书馆的实际情况,向大家介绍一下图书馆工作中几个主要环节的自动化设想和对汉字信息处理的要求。仅供有关单位在设计汉字信息处理系统时参考。

北京图书馆是一个综合性的国家图书馆,是我国的国家书库。就其藏书量来说,已达一千万册,就其藏书内容来说,是古今中外,包罗万象,我国以及世界各国的政治经济、科学文化、社会历史、天文地理等方面无所不及。

作为图书馆来说,它所收藏的书刊资料是开展各项工作的基础,而揭示和利用馆藏书刊资料的主要工具则是图书目录。在图书馆图书的补充、保管、清查及外借流通、参考咨询等方面,目录是不可缺少的工具。因此,在图书馆的一系列工作中,目录工作是一个中心环节,也正因为如此,我们把目录工作的自动化作为我们图书馆自动化的首要目标。

一、编目工作

图书编目,便是给馆藏的书刊资料编制目录。编目工作大致分成两个部分,一是正确著录即按照一定的编目条例,将图书在各方面的特征(如书名、著者、出版的时间地点、提要

等)用文字表现出来,抄录到目录卡片上,二是组织目录,是将著录后的卡片依照一定的目录组织原则,组成一个目录体系,如分类目录、著者目录、书名目录等。这里我们介绍几个具有代表性的编目部门。

1、中文图书统一编目

统一编目,对国内出版的图书进行集中编目。其目的在于使著录工作趋向统一,推行标准化,也避免了全国各地成千上万个图书馆编目人员的重复劳动。

国家规定,凡国内出版社每出版一种新书,都要缴送我馆三本作为样书。我馆收到样书之后,则由中文统一编目组进行编目,编目人员按编目要求编写目录草片,经核对无误后交印刷厂印成目录卡片,向国内发行。

中文统一编目工作有下列特点:

①符号多。在 3×5 英寸(7.5×12.5 Cm)大小的目录卡片上,包括有书名、著者、提要、出版项、稽核项、分类号等内容。这样,在卡片上既要出现英文字母,又要出汉字,还要出现汉语拼音四声及其它符号。就汉字字体来说,有黑体和宋体,就大小来说有四号老五号、小五号和六号四种字号。这种混排的方式,给印刷排版、编辑校对的工作带来一定麻烦。一张卡片最多容纳为400个字符。

②周期短。目录卡片订户要求,一旦新书出版,向各地发行,就应该与此同时收到卡片。购到新书,配入相应卡片,使图书馆能够及时入库,及时流通。尽快为读者所用,从而避免新书在编目这个环节,滞留时间过长。现在由于这些工作均为手工进行,尤其排版、校对印刷等环节严重繁琐费时,很难满足订户在时间上的要求。

③编目工作日益加重。随着图书出版量越来越大,加在中文统一编目工作人员的工作愈来愈重。

④目录卡片的需求量不同。

2、中文资料的编目工作

资料即是指我馆通过征集、采购或其它途径而得到的非正式出版物,其编目工作与图书的编目工作大致相同,不同之处在于:目前尚未进行统一编目工作。因此:

①卡片需求量少,只供馆内所用。资料的目录草片,做成经核对无误后,并不进行排版印刷,而是刻版油印,一种资料印卡片五至七张。

②卡片上的著录事项少。每张卡片上字符个数最多为80—90个,字符种类包括汉字、英文字母、数字。

3、善本特藏

善本特藏资料是我们馆藏的一个重要组成部分。我国是一个文明古国。几千年来我国人民用汉字这种精炼简捷、寓意深长、丰富多采的表现形式,记载了中华民族的悠久历史和灿烂文化。善本特藏资料包括历代流传下来的写本、刻本、金石拓片、甲骨以及珍贵的革命文献等。由于很多善本特藏具有“古”的味道,因此,汉字的繁、多等特点在这里表现无遗。

①繁。在这里工作所接触到的几乎全部都是繁体字。作为表现善本特藏的目录卡片,必

须能够表现原文本的内容,再现当时的文字历史情况。因此,对善本特藏资料进行编目时,不能将繁体字改写成简化字,譬如“豐”不能写成“丰”。“舊”不能写成“旧”。善本部门特别注意强调要用原字,既不能丢掉一点一捺、随意挪动偏旁的位置,也不可互相代替使用。如“峰”与“峯”、“略”与“畧”等等。因为鉴别一书版本的异同,除了其它的条件之外,某些字形的异同,也是一个不可忽视的条件。如现在通用的“稿”字,古书中有三种写法,即“藁”、“藁”、“稿”,假若同一种书其它条件基本相同,而这几个字各异,那我们就可以据此来断定不同版本。

②字数多。由于善本大部是历代流传下来的资料,所以各种古体字、异体字、孤僻字等在这里充分得到了体现。许多的字也是“中华大字典”及“康熙字典”所不曾收录的。

③卡片上的内容多。除了通常的卡片内容之外,在卡片背后还要记录该书的特点,譬如,该书曾由谁做过序和跋,该书见诸于哪些书目,行款、书口、边栏如何。

由此看来,对于善本特藏这个部门,汉字信息处理设备的造字功能要特别强,现有的简化字恐怕只有极个别的可为它所用。因此造字与字的存贮等工作量都很大。

二、中文机读目录数据库

为达到图书情报资源的共享,我们开始了机读目录的准备和研制工作。机读目录源于美国国会图书馆在六十年代中期研制成功的MARC记录。机读目录也是在目录卡片的基础上发展的,它将传统的目录卡片的内容,按照相应的格式存贮到磁带上,这样便代替了手工检索的目录卡片。从而利用电子计算机存贮量大,检索速度快等特点,加快了目录情报的传播,扩大了图书的利用率。建立了机读目录数据库,图书馆的采购编目、借阅等繁重的手工工作便可由机器来代替。用它,我们还可以发行各种形式的目录,如磁带目录、书本式目录、卡片目录等。

在国内,图书资料的收藏量与国外先进国家相比,数量还是非常少的。收藏不多,但又没有一个统一的著录标准,图书馆界呈现一种杂乱无章,各自为政的状态。除此之外,国内也没有一个完善的联合目录报道工作的体系,各馆之间也不能互通有无,互相交流,达到图书情报资源的共享。这样既造成一些资料的重复收藏,也造成一些珍贵的资料收入之后便如石沉大海,无人问津,不能做到物尽其用。改变这种状况的一个方法是建立统一的图书编目中心,在计算机的基础上形成全国机读目录数据库。而作为国家图书馆的北京图书馆,理所当然的要承担大部分工作。

目前,北京图书馆与北京大学图书馆、北京大学图书馆学系、人民大学图书馆、中国科学院图书馆、中国图书进口公司等单位协同进行机读目录的研制工作,工作也只局限在西文范围上,中文机读目录数据库的建立还有待于我国汉字代码的标准统一和汉字信息处理系统的确定、研制成功之后,才能够进行。设计中的中文机读目录数据库将包括中文图书机读目录、中文资料机读目录、中文期刊机读目录、全国中文图书联合目录、全国善本书机读目录等。

在图书馆的其它工作中,也要涉及到汉字信息处理问题,如采购部门图书的财产登记。流通管理部门催还书单的打印,预约图书的登记,期刊的记到等等。但是据我们的了解,编

目是一个主要环节,无论是从它对汉字信息处理系统的要求上,还是同馆内其它工作的联系上,都是如此。因此,作为设想中的汉字信息处理系统,如果编目的问题得到解决,那么其它环节的问题也就迎刃而解了。

三、汉字信息的排版印刷问题

在我馆,除了上述工作中对汉字信息处理系统有所要求之外,还有一个排版印刷问题。北京图书馆内设有书目文献出版社,附属一个印刷厂。其工作性质与对汉字信息处理的要求与大多数出版社、印刷厂类似,在此我们就不作介绍了。

我馆印刷厂的主要任务是:

- 1、排印中外统一编目卡片
- 2、排印中外文书刊文摘动态、文献报道
- 3、排印中外文社会科学期刊篇名索引
- 4、排印馆藏专题书目
- 5、排印馆藏外文新书通报,全国外文新书联合通报及其索引本
- 6、排印定期与不定期刊物
- 7、排印专题联合目录
- 8、排印期刊联合目录
- 9、排印各种封面、报告会票、借书单等

将来新馆建成后的印刷厂预计全年的汉字排印量为2345万字,外文将达到100万英寸,用纸量达7—8千令。

综上所述,图书馆的每一个工作环节无不涉及到汉字信息的处理,可见汉字信息处理系统研制成功,将会给图书馆的工作带来巨大的变化,从而对我国的四个现代化工作也就做了非常重要的贡献。这里还值得向生产与研究单位进一言的是仅从图书馆界来说,1979年经计委统计全国公共图书馆共有1256个,国家文物局计委做出规划要求到1982年,至迟不超过1985年,全国县以上的图书馆要配齐。科学院系统的图书馆120多个,高等院校图书馆900多个,此外在数以十万计的国营工矿企业中,还存在有各种类型的图书馆,而所有这些图书馆,其工作方式大同小异,随着四个现代化的进展,也都有可能配备汉字信息处理系统,所以从将来的汉字信息处理系统产品的销售市场来看,其前景也是十分诱人的。

根据上面的介绍,从我馆汉字信息处理的需要出发,我们对设计中的汉字信息处理系统提出下列要求:

硬件设备:

- 1、由于在编目、采访、流通、期刊管理等部门设备的需要量较大,从使用方便与成本低等方面考虑,上述部门使用小键盘作为输入装置。
- 2、从传统习惯考虑,排版印刷部门使用大键盘作为输入装置。(小键盘也可考虑)
- 3、不论小键盘输入还是大键盘输入,都应配备萤光屏显示装置以便当场校对。
- 4、屏幕显示一顿应不少于400汉字,(可容纳一张卡片的内容),屏幕显示字符应包括汉字、英文字母、数字及其它特殊符号。

5、大键盘盘面字应满足一般排版所需字量，盘面布局要合理，盘外字及字号（如四、五号等）字体（黑体、宋体等）要有处理措施。

6、输出设备又可分如下几类：

①目录卡片印字输出，并可拷贝多张，便于中文资料、特藏等部门使用。字体无特殊要求，但要清晰可辨。

②宽行印字输出，便于印制各种类型统计表格，如催还书单、工资清算单，包括横向、纵向印字。

③输出可以符合印刷要求的原件，应包括各种字体及不同型号的字，再经激光或其它光学制版的方法制成可供印刷的底板。

注：印字输出应采用普通纸。

7、字库包含的字数要多，且留有扩充的余地。字模信息应存贮于随机存贮介质中。

估算例：

一般编目检索用汉字以1万字计算，其容量为1.3MB（32×32点阵，每字有2字节代码）或0.7MB（24×24点阵，每字有2字节代码）。

排版印刷用汉字，以宋体、黑体各一万字计算，容量为25MB（100×100点阵，每字有2字节代码）。

以此巨大容量，似应采用信息压缩与字形变倍技术。

小键盘输入方法（编码方案）应满足下列要求：

①汉字与代码对应要尽可能统一

②编码的汉字量要大

③编码规则要简单明了、易学易记、使用方便

④编码要短，字——码转换要容易，操作速度快

⑤误码率低

软件要求：

一般功能：

①重码字处理（包括存放、调用、新造字的重码处理）

②编辑排版处理；可进行在一字符或任一字符行的增删改，上、下、左、右移动，行距、字距变化不同；

③扩充字的造字程序

④键盘输入程序

⑤信息压缩存贮程序

⑥字形变倍程序

⑦印字输出程序

专用功能：

①目录卡片编辑程序

②目录卡片文件的组织（排序）程序

③目录的检索程序（多种途径，如分类号、书名、著者、出版社、主题等。）

④目录输入打印程序

⑤图书流通管理程序（包括借书登记、预约、催书检查等）。

.....

通用功能：

①工资处理程序

②档案管理程序（如人事档案、医疗健康档案等）

③统计报表印程序

.....

由于我们接触不多，没有实际经验，还有许多意想不到的问题不能提及。简言之，对汉字信息处理系统的要求可以用十二个字概括：稳定可靠，经济实用，方便用户。

以上所谈要求仅是从分布的角度来论述（即各个部门都有若干输入、输出设备及处理机），这样如何同主计算机进行联系或将各部门设备作为终端，共用主机资源等等问题，还有待于在会下进一步交换意见。

汉字信息处理，错综复杂，涉及到计算机技术、激光技术、文字学、语言学等许多学科的内容，解决好这个问题，需要包括厂家、研究单位、用户在内的各个方面的共同协作，为了尽快地把汉字信息处理系统搞上去，提高国内计算机的应用水平，促进四个现代化建设，根据目前了解的情况，我们提出如下希望：

1、由专业、主管部门统一领导，在人力、物力、财力上统一规划安排，

2、将国内外各种类型的汉字信息处理系统进行比较，分析，综合，确定在我国可以实用普及的汉字信息处理系统方案，

3、汉字信息处理系统的研制应与计算机软件、硬件设备的研制结合起来。

4、在汉字编码“百花齐放”的基础上，进行筛选，分析，综合，集各家所长，制定几个公认是比较好的方案，尽快上机实验。

5、制定标准的计算机汉字信息交换代码，

6、大键盘、中键盘、小键盘各有利弊，受到不同用户的重视应该同时采用，满足不同要求，

7、国外现有不少厂家与用户合作，生产承包式产品，国内是否也根据各种不同的用户要求，采取这种方式。

汉字拼音键盘的设计

杭州自动化研究所 盛 谏

前 言

我国目前在电子计算机终端上普遍使用的键盘⁽¹⁾是国际通用的英文键盘,它已有一百多年历史。虽可采用各种汉字编码的办法输入汉字,但由于各字母的排列不符合汉字频度规律和构字规律,不仅难学、操作效率也大为降低,不能适应我国电子计算机事业发展的需要。

为了提高字母键盘输入汉字拼音的效率,1963年丁乙提出了根据Dvorak的合理化原则安排字母键盘布局的设想⁽²⁾。Dvorak的合理化原则⁽³⁾十分符合人们的经验,即:1)打字工作量应使左右手负担相当而右手略重;2)左右两手交替合作击键,避免单手连续击键,亦即打字要有节奏感,形成节奏流;3)最常用的字母放在最便于操作的中排字键上,并由最有力的手指来管理。各手指中以食指能力最强,中指、无名指其次,小指最弱;4)越排击键的动作应减到最少。

根据丁乙的思路,陈明远利用计算机统计汉语拼音字母的出现概率并兼顾英文字母的出现率,得到了一种字母键盘的排列,称为“DUIHGE”键盘⁽⁴⁾见图1

R	P	W	B	Y	J	S	L	C	F
.0074	.0033	.0109	.0172	.0288	.0234	.0292	.0190	.0131	.0082
D	U	I	H	G	E	A	N	O	Z
.0401	.0794	.01255	.0735	.0721	.0844	.0957	.1174	.0604	.0340
	V	K	M	NG		X	T	Q	
		.0061	.0125			.0162	.0162	.0104	
.0475	.0827	.1425	.2041			.2489	.1482	.0839	.0422
小指	四指	中指	食指			食指	中指	四指	小指

左手 0.4768

右手 0.5232

图1 DUIHGE键盘布置⁽⁴⁾ 括号内数字是该字母的出现率)