

● 洪 漪

## 我国信息网络建设中的检索语言问题

**ABSTRACT** The construction of information network requires a retrieval language with uniform compatibility to bring about the drawing up of retrieval language, the automatic management and the gradual increase of natural — language factors. The machine — readable system should be carried out as soon as possible, and the function of national descriptor base should be brought into full play so as to integrate with the natural language and construct a integrated retrieval language system. 2 refs.

**SUBJECT TERMS** Information network constructions — Relations — Retrieval language — Applications

**CLASS NUMBER** G254.0

信息网络建设是一项牵涉面广、耗费巨大的系统工程,包括文献资源建设、信息标准化建设、网络组织的科学化建设以及人员队伍建设等方面,其中检索语言在网络中起着语言保障的作用,占有举足轻重的地位。近些年来,我国检索语言无论在理论上还是在实践上都紧跟世界最新进展,发展很快,然而检索语言的研究和创制中还很少顾及信息网络建设的需要,长此以往必将导致将来联网的障碍。因此,有必要对信息网络建设中涉及的检索语言问题作更深入的研究,针对我国检索语言现状以及网络化对检索语言的特殊要求,明确检索语言在网络中应起的作用,探讨检索语言在信息网络化进程中的发展道路。

### 1 信息网络对检索语言的要求

信息网络是一个对信息资料进行存贮、加工、利用的协作系统。它是以分工合作为基础,使信息广为流通的整体性、开放性体系,相对于单个的信息系统而言,能满足更大范围的信息需要。信息网络的建立除了必要的通信手段和技术装备外,还需其他各方面的整体配合,而统一、兼容的检索语言则是信息网络建设的首要条件,它关系着网络内信息标准化的

程度、数据库的质量乃至整个网络的效率。

一般来说,检索语言是为某个系统特定的需要而编制的。不同部门、不同系统、不同专业领域所用的检索语言之间存在着很大的差异。随着检索语言语种和数量的增多,这种差异有逐渐增大的趋势。这在当前对检索工作或许并没多大影响,但对联机网络化的发展则会造成严重阻碍。系统间的语言障碍使各系统难于沟通、联结,使现有数据库不能充分地利用,从而影响到网络化的进程,所以标准化、兼容化就成为信息网络对检索语言首先提出的要求。检索语言的标准化和兼容化是为达到检索语言的通用性所采取的不同方法,其目的都是力求达到各种检索语言结构和词汇上最大程度的一致性,使不同系统的信息资源得以共享。标准化带有一定的强制性,而兼容化重在检索语言的协调转换,是在不变动各系统现有体系的基础上进行的。

现代信息网络是以计算机的应用为核心的,是联机检索发展的高级阶段。没有计算机技术的广泛应用,也就没有脉络贯通、应变快速的网络体系,与此相应,要求实现检索语言编制、管理的自动化。分类表、词表的机编化、机读化除了可以提高分类表、词表的质量、满足文献标引和检索不断增长的需要

外,还可为联机检索提供方便,有利于词汇的自动转换,对检索语言的统一和兼容大有益处。

随着联机检索的普及并向网络化方向发展,收罗概念有限、结构复杂的检索语言越来越难以适应数量众多、要求各异的联机终端用户的检索要求。这就要求检索语言逐步增加自然语言因素,使检索语言更易为普通检索者所接受,使他们在不熟悉检索语言结构的情况下也能编写出较好的检索提问式,迅速准确地查找到自己所需要的信息资料。计算机性能的不断改善使自然语言检索带来的不利因素会逐渐减少,联机网络中自然语言与规范语言的结合使用势在必行。自然语言不依附于特定的数据库,几乎可适用于一个联机网络中所有数据库,采用自然语言进行标引和检索基本上可解决检索语言的兼容问题。当然,这是以计算机信息检索的高度发展为前提的。

## 2 我国现有检索语言的局限性

信息网络建设要求检索语言实现标准化、兼容化、计算机化和自然语言化,但我国检索语言的现状远远不能适应这些要求,主要表现在6个方面。

**2.1** 尽管走统一、兼容的发展道路已成为检索语言界的共识,但实际工作中大多数检索语言在编制使用上仍是各行其是,只顾满足本系统、本单位的当前需要,很少从全局考虑,更少考虑到将来网络化的需要。

**2.2** 检索语言的标准化工作已有一定进展,分类法系统中,《中图法》实际已成为全国统一的分类检索标准。主题法方面,也颁布了几项国家标准,如《汉语叙词表编制规则》、《文献叙词标引规则》等,对汉语叙词表的编制使用给予规范。然而,已编成的大多数分类表、叙词表是在没有标准的情况下完成的,这些类表、词表如何根据标准进行改造,是急需解决的问题。针对不同专业的特点,还应制定相应的细则与标准配套使用,使标准得到严格的贯彻实施。有关检索语言的标准中分类法和主题法分列,新型的一体化词表的编制使用无标准可循,也是不适应分类主题一体化发展的实际的。

**2.3** 兼容问题已成为检索语言研究热点,并已取得一定成果,如在分类法与主题法的兼容方面,《中国分类主题词表》即将出版,叙词语言方面,大型的国家叙词库正在建立。但总的来说目前我国检索语言的兼容还停留在较低水平,往往只是在选词、结构安

排、关系显示、标引规则等方面参考现存检索语言,带有自发性。对一些较为先进的兼容技术,如词汇的自动转换、中介词典的编制等没有系统研究和应用。

**2.4** 最近一段时间编制的词表大多采用计算机编制和管理,如《教育分面叙词词表》、《社会科学检索词表》等,但这毕竟是少数。现有检索语言中,机读型检索语言所占比例极低,尤其是分类语言基本还处于手工编制和管理阶段,这在相当程度上限制了检索语言功能的发挥。

**2.5** 由于我国信息检索自动化程度较低,适于手工检索的规范化语言占据主导地位,自然语言的应用很少。以自然语言为基础的主题检索语言迟迟达不到象分类语言那样的普及程度,采用自然语言检索的关键词法由于种种原因得不到广泛应用,汉语叙词表中词汇的等同率普遍较低,自由词的使用受到严格限制。凡此种种造成检索语言的易用性较低,并给各系统间的兼容带来困难。

**2.6** 与网络化趋势相适应,当前世界上多语种检索语言持续增长,而我国检索语言的绝大多数为单语种。这将不利于中外检索语言的兼容并使我国信息系统难以进入国际联机检索网络。

## 3 对策

### 3.1 尽早实现检索语言的机读化。

相对而言,汉语叙词语言的机读化比分类语言要容易一些。原因有二:一是新编词表多为计算机编制;二是在国家叙词库的第一期工程中,已将现有汉语叙词表全部输入计算机,这样经过必要的修订后,原有的印刷版词表可通过词库自动转化为机读型词表。

近年来计算机编制分面叙词表的实践表明,分类表也是可以利用计算机编制和管理的。关键在于如何对目前通用的体系分类法作适于计算机的改造,同时寻求分类语言与叙词语言在机检系统中的有机结合,从而充分发挥分类法的系统性特长和计算机的优异性能。分类法长期以来占据我国检索语言的统治地位,从某种意义上说,分类法在联机检索中的应用研究更具实用性。在分类语言的机读化改造中《中图法》理应首当其冲。

### 3.2 充分发挥国家叙词库的作用。

建立国家叙词库是为实现汉语叙词表的统一和兼容采取的一项重大措施。词库最终能否达到预期的目的还很难预测,但它在信息网络中的作用不可

低估。这就要求在建库初期充分考虑网络化建设的需要,使词库所收数据尽可能完备,结构尽可能合理,功能尽可能多样。词库的建立除了有利于汉语叙词表的机读化外,还可起到以下作用。

3.2.1 促进汉语叙词表的标准化。词库可为编制、修订叙词表提供标准化的数据,有利于汉语叙语表语汇的统一和规范。

3.2.2 作为一兼容中心,成为各种词表词汇协调转换的媒介。可利用词库中现成数据进行词汇的自动转换试验,还可为词库配备一中介词典,进行叙词的转换工作,使检索语言的兼容达到更高层次。

3.2.3 给词库中每个词配以外文等价词,据此编制多语种叙词表,与国外系统兼容。

3.2.4 需要时与《中图法》相联系,有利于分类主题一体化,并使《中图法》尽快用于计算机检索。

3.2.5 作为检索语言发展的基础,根据需要生成各种完全兼容的专业性词表或集成性词表。

3.3 与自然语言相结合,建立整体化检索语言体系。

鉴于我国计算机化程度不高,规范语言占据主导地位的现实,完全的自然语言检索在短期内还无法实现,将受控语言与自然语言结合使用就成为一种明智的选择,这在国外已形成一显著的趋势。受控语言与自然语言结合的方式有很多,其中将规范词、分类号、自由词合为一体的整体化检索语言是联机检索系统较为理想的语言,这种整体化检索语言的结构与功能如表所示。

整体化检索语言结构与功能

| 结构   | 受 控 语 言                     |                 |                | 自 然 语 言        |        |
|------|-----------------------------|-----------------|----------------|----------------|--------|
|      | 统一分类表                       | 叙词表             | 同义词表           | 关键词系统          | 自由词表   |
| 功能   | 分类标引、<br>分类检索工具             | 叙词标引、<br>主题检索工具 | 词汇控制           | 自由标引           | 辅助检索工具 |
| 应用范围 | 浏览性检索、<br>专题检索、<br>SDI 族性检索 | 特性检索            | 入口词表、<br>后控制词表 | 联机检索<br>SDI 检索 | 词表显示   |

在这里,几种检索语言是有机结合在一起的,分类表和叙词表之间用分类号相连,而同义词表则是联系受控语言 and 自然语言的纽带,它既可以作为叙词表的入口词表使用,又可作为关键词系统的后控

制词表使用。采用整体化检索语言可以同时用分类号、叙词、关键词标引一篇文献,组成同一词汇文档,各检索标识之间可通过计算机实行自动转换。由于整体化检索语言兼有分类语言、叙词语言、自然语言三者的长处,因此具有比单一的检索语言高得多的检索效率。

下面是一个利用整体化检索语言在联机检索系统中检索的例子:

检索课题:《红楼梦》研究

学科领域:文学

使用分类法:《中图法》

使用叙词表:《汉语主题词表》

检索过程:(1)同时采用叙词、自由词进行标引,标出检索标识:古典小说—文艺批评、古典小说—研究、红楼梦、曹雪芹。

(2)通过《中国分类主题词表》找到古典小说评论在《中图法》中的分类号 I207.41。

(3)输入类号 I207.41,由系统显示这一类号的相关类目结构:

|          |           |
|----------|-----------|
| I206     | 文学评论和研究   |
| I207     | 各体文学评论和研究 |
| I207.2   | 诗歌、韵文     |
| I207.4   | 小说        |
| I207.41  | 古代小说      |
| I207.411 | 《红楼梦》研究   |

检索者可在这个类目体系结构片断所提供的语言环境中,进行浏览性检索,扩大或缩小检索范围。

(4)用叙词、自由词进行更专指的检索。

总而言之,整体化检索语言集中体现了检索语言兼容化、计算机化、自然语言化的发展趋势,是符合信息网络化需要的。

参考文献

1 侯汉清、马张华.主题法导论.北京:北京大学出版社,1991  
2 洪漪.我国国家叙词库建设中几个问题的探讨.情报学刊,1993,6(3):209~211

洪 漪 1987 年武大图书情报学系毕业,现为武大图书情报学院讲师。通讯地址:武汉,邮编 430072

(来稿时间:1994-06-20。编发者:黄文田。)