

●谢汉斌

## 国外对分类法在机检中应用的研究

随着计算机系统的发展,人们日益认识到文献记录文档及其相关的著者标准文档和主题词标准文档的重要性。在我国对分类法在机检中的研究近来引起了足够重视。而在国外,这方面的研究已有不短的历史了。

### 一、对 UDC 在机检中应用的研究

分类法在自动化系统中的应用可追溯到40年代末期。当时,采用的是穿孔卡片。第一个将分类法用作主题存取和浏览的联机检索系统是1961—1968年间建成的,承担研制任务的是美国的 Robert R. Freeman 和 Pauline Atherton。他们用《国际十进分类法》(UDC)的码号和英文类名,建了一个核科学文档。UDC的相关索引和范围注释也作为其中的一部分。他们的基本想法是:现在,人们一般认为使用自然语言比使用类号方便,但实际上类号在机检中有一定的优点。它不仅能唯一地表达一个概念,而且能反映概念之间的关系。若能将UDC的类号与自然语言中的语词直接挂上钩,那么,用户就不用直接接触UDC类号,而可以方便地使用自然语言,同时又获得了使用类号所具有的优点。

基于这一考虑,他们在存储UDC类表

时,不仅存储了类号,而且存储了反映类目关系的符号。但是,为了方便计算机处理,对其中一些作了变通处理。(见表1)。

表1 UDC符号的形式比较

| 序号 | 类型   | 名称   | 一般形式              |           | 机内形式                       |              |
|----|------|------|-------------------|-----------|----------------------------|--------------|
|    |      |      | 符号                | 示例        | 符号                         | 示例           |
| 1  | 内容分面 | 通用主题 | n                 | 551.524   | C <sub>n</sub>             | C551.524     |
| 2  | 形式分面 | 语种   | =n                | =30       | B <sub>n</sub>             | E30          |
| 3  | 形式分面 | 载体种类 | (O <sub>n</sub> ) | (084.3)   | F <sub>n</sub>             | F843         |
| 4  | 内容分面 | 地点   | (m <sub>n</sub> ) | (265)     | P <sub>m<sub>n</sub></sub> | P265         |
| 5  | 内容分面 | 种类   | (=n)              | (=30)     | R <sub>n</sub>             | R390         |
| 6  | 内容分面 | 时间   | "n"               | "473"     | T <sub>n</sub>             | T475         |
| 7  | 下位分面 | 观点   | OO <sub>n</sub>   | 55.002.2  | V <sub>n</sub>             | c55V22       |
| 8  | 下位分面 | 专用复分 | -n                | 62-451    | W <sub>n</sub>             | c62x187      |
| 9  | 下位分面 | 专用复分 | on                | 62.018.7  | X <sub>n</sub>             | c62x187      |
| 10 | 联系   | 句法联系 | n,n               | 546.32.13 | n'n                        | c54632Y13    |
| 11 | 联系   | 包含联系 | n/n               | 543/546   |                            | cb43<br>c546 |
| 12 | 联系   | 相关联系 | n:n               | 543;546   |                            |              |
| 13 | 联系   | 一般联系 | n+n               | 543+546   |                            |              |
| 14 | 联系   | 下位联系 | n(n)              | 543(546)  |                            |              |

UDC属于体系组配分类法范畴,它有了个明显的不足:1、用有限类号不能直接表达大量的概念;2、用线性方式表达非线性的关系易引起含义模糊;3、不同标引人员对同一内容可能给出不同的类号。

针对这一情况,他们在AUDACIOUS系统进行试验(见表2)时,采用了倒排档的方

表2 机读版UDC检索过程

| 步 骤 | 屏 幕 内 容                                                                                                                                                                                                                                                                          | 涵 义                       |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| 1   | brows                                                                                                                                                                                                                                                                            | 用户输入表示准备查找                |
| 2   | BROWS MODE ENTERD                                                                                                                                                                                                                                                                | 系统响应,表示可以进行               |
| 3   | clear                                                                                                                                                                                                                                                                            | 将个人文档置空                   |
| 4   | COMMAND OKAY. PUT NEXT COMMAND AT TOP OF SCREEN                                                                                                                                                                                                                                  |                           |
| 5   | term Titani um                                                                                                                                                                                                                                                                   | 查找<br>Titanium            |
| 6   | u669295 669. 295 TITANIum<br>u661882 661. 882 TITANIum COMPOUND;DXIDES(WHITE),ETC.<br>u553494 553. 494 TITANIum,ZIRCONIum AND HAF NI uM ORES<br>u54682 546. 82 TITANIuM GROup<br>u54682i 546. 82i TITANIuM                                                                       | 类名中有<br>Titanium<br>的所有类目 |
| 7   | save<br>+u669295 669. 295 TITANIum<br>+u661882 661. 882 TITANIuM COMPOUND;OXIDES(WHITE)ETC                                                                                                                                                                                       | 将下述内容<br>存于个人文档           |
| 8   | VALUES SAVED                                                                                                                                                                                                                                                                     | 存贮完成                      |
| 9   | Sean +66188                                                                                                                                                                                                                                                                      | 显示类表<br>66188 片断          |
| 10  | u6618 661. 8 METALLIC COMPOUND MINERAL PIGMENT. CF. 669<br>u66188 661. 88 APOUPS IV AND V METALS. AS 546. 8<br>u661881 661. 881 TIN COMPOUND;STANNA TES.ETC<br>u661882 661. 882 TITANTUM COMPOUND;OXIDES(WHITE)ETC<br>u661886Y887 661. 8861. 887 ANTIMO NY AND BISMUTH COMPOUNDS |                           |
| 11  | dispiay                                                                                                                                                                                                                                                                          | 显示个人文档                    |
| 12  | 1. u669295<br>2. u661. 882                                                                                                                                                                                                                                                       |                           |
| 13  | I TEM CouNT=(1 'OR' 2)                                                                                                                                                                                                                                                           | 计算符合条<br>件的检索项            |
| 14  | I TEM CouNT=5<br>COMMAND OKAY<br>PuT NEXT COMMAND AT TOP OF SCREEN                                                                                                                                                                                                               | 共有 5 篇                    |
| 15  | find (1 'OR' 2)                                                                                                                                                                                                                                                                  | 检索                        |
| 16  | RESULTS OF RETRIEVAL<br>I TEM 1<br>.....<br>.....<br>END OF RE TRIEVAL<br>COMMAND OKAY<br>PuT NEXT COMMAND ATTOP OE SCREEN                                                                                                                                                       | 逐篇显示                      |

式,一方面回避了这些问题,另一方面也适应了机检的需要。

在大量试验的基础上,他们得出的结论认为:UDC在机检中的应用也许不如专为机检而编制的检索语言有效,但无论作批处理还是联机操作,并没有什么本质上的困难影响它在机检中的应用,而且由于现在有许多用UDC编的目录,有许多熟悉UDC的工作

人员,况且UDC已广泛用于国际交流,同时采用UDC也可节省另编检索语言的费用,因此UDC在机检中的应用不仅是必要的,也是可行的。

这项研究是40年代至70年代所作的最广泛,最深入的研究。虽然在其前后也有人作过理论和实际上的研究,但无论从广度或深度来看都相差甚远。然而,这项研究的不

足之处也是显而易见的。它没有将使用分类法的系统和不使用分类法的系统作横向的比较,以明确谁优谁劣;也没有探讨如何用分类法来弥补主题词法的不足。到了 80 年代,随着叙词法在机检中的广泛应用,探讨和解决这两个问题就成为更加紧迫的课题。

## 二、对 UDC 在机检中应用的研究

1981 年秋, Karen Markey 在 Forest 出版社,图书馆资源委员会 (CLR) 和 OCLC 的支持下,针对 Freeman 和 Atherton 的研究只是将 UDC 作为联机的入口词,而没有探讨分类法在联机状态下怎样成为主题存取、浏览和显示的工具,怎样拟定不适合使用主题词法的检索策略的这一不足,开展了对《杜威十进分类法》(DDC) 的机读版的研究,这是继 Freeman 和 Atherton 之后所进行的第二个研究高潮。

她采用 DDC 作联机检索研究,试图着重弄清以下 3 个问题:

- 1、DDC 是否适用于主题存取、浏览和显示;
- 2、应用 DDC 是否改进了机检系统的性能;
- 3、用户是否更愿意采用有 DDC 的联机公共目录。

这项研究所用的目录包括 4 个图书馆的文献记录,以联机目录的方式提供给用户查找。为了便于横向比较,这些目录分成两个部分:

- 1、Soc (Subject Online Catalog)。它和普通的联机目录一样(见表 3)。
- 2、Doc (Dewey on Online Catalog)。它除了有 SOC 中的内容外,还包括 DDC 的相关索引、类名和类目注释(见表 4)。

在这几种检索方式中,比较引人注意的

表 3 SDC 指令

| 指令   | 涵义                       |
|------|--------------------------|
| 1 SA | 查找有关的主题词,并能对检出的主题词作字顺浏览。 |
| 2 SD | 直接用主题词查找                 |
| 3 SC | 直接用分类号查找                 |

是 SC 和 SS。在 SOC 中使用 SC 时由用户输入类号直接进行检索,而在 DOC 中用户输入类号后系统还将反馈回有关信息,如上位类,下位类信息以及更详细了解类目和查找方法的指令等。

表 4 DOC 指令

| 指令   | 涵义                       |
|------|--------------------------|
| 1 SA | 查找有关的主题词,并能对检出的主题词作字顺浏览。 |
| 2 SD | 直接用主题词查找                 |
| 3 SC | 直接用分类号查找                 |
| 4 SS | 查找有关的主题词,并能查看主题词的范围注释。   |

使用 SS 时,系统将用户输入的词分别与 DDC 的类名、类目注释、相关索引和主题词中的头一个词进行匹配,并将匹配的有关类号按前 3 位是否相同结合成组,最后在屏幕上显示查找上位类、下位类、了解某一类号更详细情况和检索方法的指令,供用户选择。

对联机目录的试验分两种情况,一种是由用户自己选定课题,分别在 SOC 和 DOC 上检索,然后对结果作出评价;另一种是给用户指定课题,让他们在 SOC 或 DOC 上检索,然后对结果作出评价。总计在 SOC 上进行 342 次、在 DOC 上进行 338 次检索。结果是, SOC 的成功率为 90%, DOC 的成功率为 86%,另有 5% 的检索在 SOC 上失败了却在 DOC 中获得了成功,有 9% 的检索在 DOC 上失败了而在 SOC 上却获得了成功,还有 5% 的检索在 SOC 和 DOC 上都没有成功。

出乎我们意料的是,从发表的统计数据与许多图表来看,并不能从中找出 DOC 明显改善了系统功能的迹象,在性能上 DOC 与 SOC 不相上下,只是在 SD 指令中有点意外。

SD 指令在 DOC 中的应用次数多于 SOC, 成功率也高于 SOC。也就是说, DOC 能提高系统的用主题词直接检索的能力, 但有时也会检出太多的无关文献。大量调查表明, 工作人员有一半喜欢 SOC, 一半喜欢 DOC; 但用户却偏爱 SOC。

### 三、对 LCC 在机检中应用的研究

关于分类法在机检中应用的研究, 上面介绍的只是在世界范围内规模较大, 影响较为深远的两项试验。在这之后, 加拿大的 Nance J. Williamson 曾研究过《美国国会图书馆图书分类法》(LCC) 的机读版。从发表的文章来看, 她的指导思想有所提高, 遗憾的是至今尚未见到其研究成果。

### 四、结 论

各种检索语言在联机检索中应用得最成功的是叙词法。但是随着叙词法的广泛应用, 其局限性也日益暴露了出来。叙词法虽说也吸取了体系分类法的某些优点, 但其族性检索能力终究不及分类法, 分类语言与主

题词语言之间应该互相补充、互相完善, 而不能用某一种语言完全取代另一种语言。因此, 研究分类法在机检中的应用是必要的。特别在我国, 采用分类法的图书情报部门占绝对优势, 因此研究分类法在机检中的应用尤其必要。

从前面介绍的情况来看, 现有的这些研究成果是不能令人满意的。到底分类法应如何在机检中应用才能取得较好的检索效率呢? Robert R. Freeman 和 Pauline Atherton 认为, 应将 UDC 与详细的叙词表结合起来。Karen Markey 认为, 应通过指针把主题字顺索引和 SS 指令与 DDC 联系起来。

这些看法和建议, 无疑和分面叙词表的思想十分吻合, 虽说他们没有直接提出分面叙词表这样一个概念。但分面叙词表本身就是分类主题的一个有机整体, 它既通过主题词法因素增强了分类检索的能力, 又通过分类体系补偿了主题词检索的不足。从情报检索语言的理论来看, 它是一个很有前途的检索语言。Robert R. Freeman, Pauline Atherton 和 Karen Markey 在实践中的亲身体会与情报检索语言的最新成果如此吻合本身就是一个强有力的证据。从分类法在机检中应用的发展趋势看, 分类/主题一体化无疑将是其发展的基本方面。

(上接 62 页)

馆通讯》, 1987 年第 4 期, 1988 年第 1 期。

②戴维民: 我国分类主题一体化研究进展, 载《大学图书馆通讯》, 1987 年第 6 期。

③ Thesaurus guide: analytical directory of selected vocabularies for information retrieval. Prepared by G1D, 1985.

④(苏) 维列斯卡娅著 李修宁译: 在统一自动检索系统中主题和分类途径的相互关系, 未刊稿。

⑤ Foskett, A. C.: The Subject Approach to Information, 2nd ed. London, Bingley, 1977.

⑥(美) 兰开斯特, L. W.: 《情报检索词汇规范化》科技文献出版社, 1982 年, 第 74—77 页。

⑦ Lancater, L. W.: Vocabulary control for information retrieval, 2nd ed. Information Resources press 1986, PP89-103.

⑧侯汉清: 分类主题一体化的新进展——英国《基础叙词表》评介, 载《国外图书情报工作》1987 年第 1 期。

⑨ Aitchison, J.: Integration of The saari in the social Sciences, International classification, Vol. 8, NO. 2, 1981.

⑩戴维尼: 引用次序研究, 载《情报科学》1988 年, Vol. 9, No. 2.

⑪ Lancaster, L. W.: Thesaurus construction and use: a condensed course, Paris, Unesco, 1985.

**A breakthrough of thesaurus in traditional pattern — the theory and skill of drawing up “Educational Facet Thesaurus”** / Dai Weimin, Hou Hanqing // Bulletin of China Society of Library Science / China Society of Library Science. — 1989, 15 (2). —55~62,66

“Educational Facet Thesaurus” is the facet thesaurus which first draw up with the help of computer. It consist of facet classification scheme and alphabetical thesaurus. the former is used for classification indexing, building document files of classification retrieval and classification filing; the latter is used for building document files of subject retrieval. It is programmed by 10 steps: 1. overall planning; 2. fundamental frame designing; 3. collecting words; 4. facet analysing; 5. words contralling; 6. systematic arranging; 7. inputting data and editing; 8. computer automatic producing; 9. trying indexing; 10. modifying and finishing version. The publication of “Education Facet thesaurus” has shown us two points: 1. Facet thesaurus in largely superior to traditional ones. 2. The tendency of compiling thesaurus are becoming to be done with the aid of computer. With 2 figures, 2 tables and 11 references

Education Facet Thesaurus — China

Facet classification — Structure

Thesaurus — Compiling whith the aid of computer

G254

**国外对分类法在机检中应用的研究/谢汉斌//图书馆学通讯/中国图书馆学会. — 1989 15(2). —63~66**

国外对分类法在机检中应用的研究,始于40年代末期。主要有 Robert R. Freeman 和 Pauline Aterton 对 UDC 的研究、Karen Markey 对 DDC 的研究和 Nance J. 对 Lcc 的研究。作者认为,分类/主题一体化将是分类法在机检中应用的基本发展方向。表4。

图书分类法——机读化

情报检索语言——研究

G254.12

**Imagining of development strategy for library undertakings in colleges and universities** / Yan Dehai, Lin Suren // Bulletin of China Society of Library Science / Chinese Society of Library Science. — 1989, 15(2). —67~69

The development strategy for library undertakings in colleges and universities is mainly studying macroscopic management. The full text falls into five sections and makes the focal points of four enhancements. They are: strengthening entirety building; strengthening educational functions and informational functions; strengthening facilities and equipments building; strengthing professional staffs training. The are many good suggestions in the text. Among some of them are these: establishing central library for subject documents and information and storing library, forming editorial and publishing board, organizing research centre for soft science, building advisory group for equipments and automatic cooperation group and making work appraising plan, Author suggests that taking libraries in colleges and universities as strategic point and building libraries network by stages from provinces and municipalities to regions; form regions to the whole country.