

図書・文献整理の新方式—100進法PCS

宍戸 修

概 要

UDCなど従来の図書・文献の整理方法は不便な点が多く、また機械化に向かないので、その欠点を改め、機械化した新しい方式を開発した。UDCでは知識を10個ずつの群に分割するのに対して、この方法では100個ずつに分割し、これに0から99までの数字をコードとして与え、PCS機械によって検索する。この方式は100進法なので分類に無理がなく、余裕があり、また使用者が自分に都合のよい分類を自由にこの方式に取り入れることができる。また機械検索であるがPCSなので、比較的経費がかからず、企業体や機関の研究所などで能率よく使用することができる。

1. まえがき

図書・文献などを整理するには書名、表題、著者名、発行年月日、などのような書誌的事項によって整理する方法もあるが、内容分類によって整理する方法が便利が多い。ことに科学技術文献などの場合は、この方法が必要である。この内容を分類する方法は、小規模なものでは、各人、各社または各機関が、自分に都合のよいように、自由に作って、自由に使用しているものが無数にある訳であるが、相当広く使われて権威あるものには、DC, UDC, NDC, Sfb/UDC, 米国議会図書館分類法、特許分類、その他がある。

しかしわれわれが企業体の研究所において、技術文献や技術図書の整理の実際に携わってみると、これらの方法はいずれも一長一短で、不便や不都合を感じるが多い。

ことにUDC分類法は国際十進分類法と訳され、知識の普遍的な分類法として、現在最も完備したもののように言われているが、あらゆる知識の部門を、わずか10個の部門に分割するところに、根本的な無理があり、その他種々の欠点があり、またこの分類法は電子計算機による機械化に不便なこともあって、技術革新の現代には適合しない面が多くなってきた。

試みにUDCのおもな長所と短所をあげてみると、

UDCの長所

- (1) 記号に数字を用いること。
- (2) 10進法であること。
- (3) 無限に細分可能であること。
- (4) 人類の全知識を普遍的に分類できること。
- (5) 国際的にかなり用いられていること（欧米、ソ連など）。

UDCの短所

- 1) 10分割では分類数が不足で不便である。知識のすべてが入り切らない。

- 2) 分類の考え方が古くて、新しい学問知識が入り切らない。
- 3) 分類が欧米中心主義で、日本には不便である。
- 4) 分類が国際的に固定されているので、各企業体または専門分野の特殊性を盛り込むことが困難で、自由さが無い。
- 5) 関連項目の継ぎ方（コロンの他）や補助標数が複雑である。
- 6) 詳細分類の日本語版がない。
- 7) 分類の採用に個人差がはなはだしくなり混乱する。
- 8) 標数が長くなって困る。（図書に書けない）
- 9) 機械化がむづかしい。（電子計算機に乗り難い）
- 10) 分類が複雑なので専門家を要するが、UDCの専門家が少ない。
- 11) 余りに繁雑困難なので、担当者がいやがり逃げる。
- 12) 建築に対する考え方が日本と違うので、建築の部門が3つに分れて標数が分散する。

建築学（設計・建物の種類）	72
建築業（一般構造・材料・ 施工・設備）	69
建設工学（構造力学）	624/628

2. 目的

UDC 其他在来の分類法は、以上のように不便、不具合な点が多いので、その欠点をできるだけ除き、しかも検索を機械化できる方式として、新たに開発したものがこの100進法PCS (Punch Card System) 方式である。

3. 方法

3.1. 分類法について

知識を分類するに際して、UDCでは10個に分割するのに対して、本法ではこれを100個に分割し、それをさらに細分するにも100個づつに分割し、以下同様とする。この分割されたものに、それぞれ0から99までの番号数字を記号として与える。(図-1)

この分類数字に対応する分類項目は、各社または各

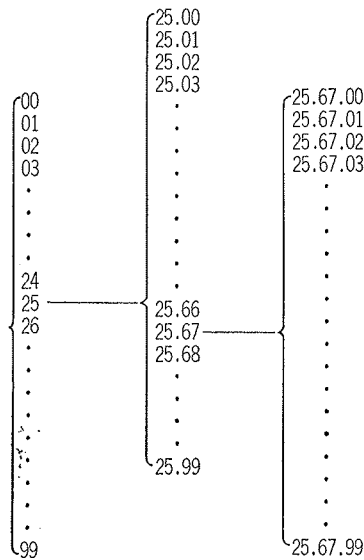


図-1 分類方式

個人などその使用者が、自己に都合のよい分類に応じて自由に選択できることとし、また100個の分類は必ずしもその全部を使う必要はなく、使用者が必要とす

る範囲内だけに止めることができる。(図-2)

要するに00.00.00.00……のように2桁づつの分類となる。もし化学元素などのように100個以上必要な場合は、その必要な部分だけ1000個に分割することも可能である。すなわちその部分だけ3桁となる。たとえば00.000.00.00……となる。またこの場合1000個に分割せず、同族体などをまとめて、分類を100個以下になるようにし、その同族体をさらに100個づつに分割する方法もとることができる。

この方法による分類の桁数が、どのくらいになるかを見るために、100進法とUDCとの分類例を二三比較して示すと次の通りである。(ただし100進法のは分類数字が未定なので桁数だけを示す)

分類項目	UDC	100進法
電気工学	621.3	00. (2桁)
電灯電気照明装置	621.32	00.00 (4桁)
放電灯	621.327	00.00.00 (6桁)
水銀灯	621.327.534	00.00.00.00 (8桁)
機械工学	621.	00 (2桁)
物上げおよび荷役機械	621.86	00.00 (4桁)
コンベヤ	621.867	00.00.00 (6桁)
ローラコンベヤ	621.867.6	00.00.00.00 (8桁)
建築工学(建設工学)	624/628	00 (2桁)
力学	624.04	00.00 (4桁)
弾性	539.3	00.00.00 (6桁)
二次元弾性	?	00.00.00.00 (8桁)

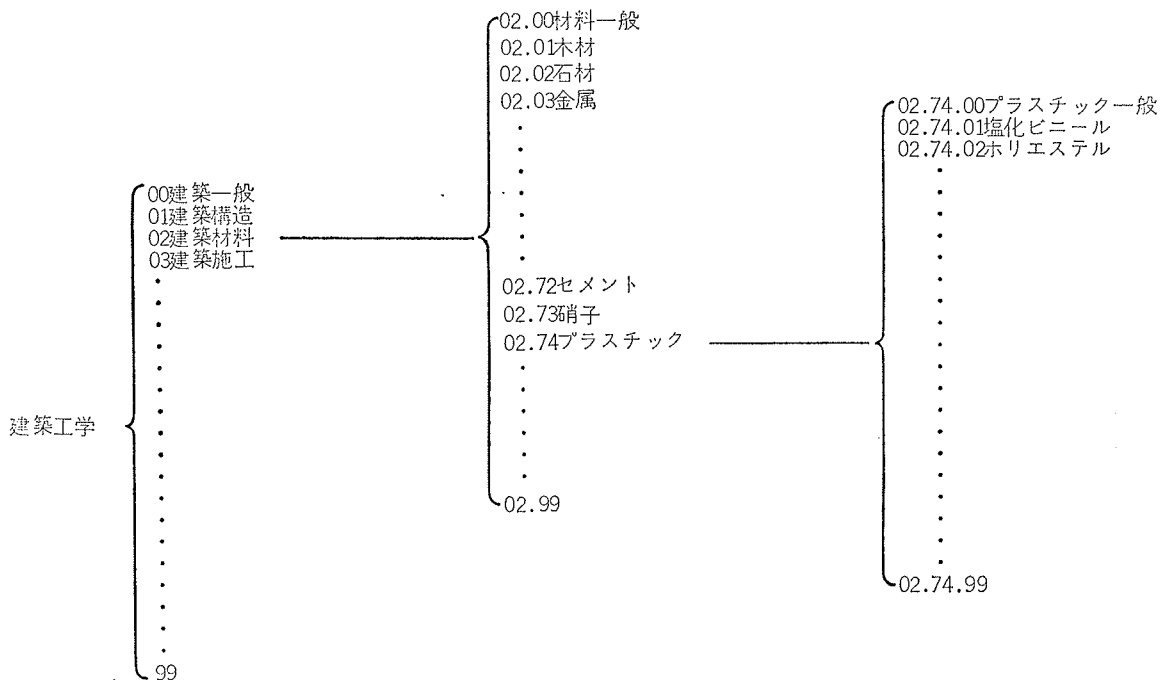


図-2 分類例

土 木 工 学	624/628	00	(2 桁)
土 木 施 工	?	00.00	(4 桁)
く い 基 礎	624.154	00.00.00	(6 桁)
ジェットく	624.154. x	00.00.00.00	(8 桁)

これで見ると、分類の桁数はほとんど同程度であるが、この方式は企業体、機関などの専門分野の特殊性を盛り込むことが主眼であるから、最初の 2 桁は省略できる場合が多い。すなわち桁数を 2 桁減少することが可能である。

3.2. 検索法について

前記分類数字をコードとして、80 欄または 90 欄カードにパンチして、PCS 機械にかけて検索するものとする。

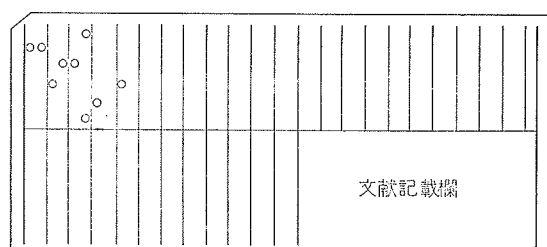


図-3 カードの一例

たとえば 90 欄カードを用いるとすれば (図-3) のように、下欄の半分を文献の表題、執筆者、掲載誌、発表年その他の記載欄とし、それ以外を分類欄とし、分類欄は 2 コラムごと (必要に応じて 3 コラムでもよい) に区分して分類コードをパンチする。

また内容分類だけでなく、書誌的事項もコード化してパンチすることができる。

内容が多い分類項目にまたがるものは、クロスリファレンスができるように、分類コードのパンチ欄を 2 重または 3 重に設けて、多次元検索を行なえるようにする。

また分類項目の索引には、分類項目表のほかシソーラスあるいはコード索引辞書などを作成して用いることとする。

4. 結果

以上本方式の基本構造を述べたが、要するにこの方式によれば、内容分類を細密にできるだけでなく、使用者の要求に自由に適合させることができ、また検索の機械化が可能である。

この方式のおもな長点をあげてみると

- (1) UDC のように 10 分割でなく、100 分割なので分類に無理がなく、また余裕があること。
- (2) 各社、各個人など使用者が、自己に都合のよい分類を、自由にこの方式に取り入れることができること。
- (3) 80 欄または 90 欄カードを使用するので、ハンドソートパンチカードなどに比べれば、分類能

力が格段に高いこと。

- (4) PCS 機械によって機械検索するので、検索速度が早く、人力を軽減できるが、EDPS (Electronic Data Processing System) 方式のようにばく大な経費を要せず、企業体の研究所などには能率よく使用することができること。

5. 考察

この方式のうち 100 進法の利点についてはすでに述べたが、PCS 方式とコード化方式の得失について少し考察してみよう。

電子計算機としては PCS より EDPS の方が高級なので、情報処理方式として一般的には、PCS より EDPS の方がすぐれているとされるのが普通であるが、実際に文献検索の方法として使われているのを見ると、必ずしもそうとばかりは考えられない。

まず経費の点から言って、一企業体や一機関の研究所などにおいて、情報検索専用として高性能の EDPS 機械を導入するのが困難なことは自明であろう。たとえ経営、経理、技術計算等の用途と共用するとしても、わが国の実状では困難性が大きいし、共用の場合は機械待ちの問題もあり、また膨大なアウトプットの利用度の問題もある。

次にコード化方式についてみると、機械による情報検索の方式には

- (a) 情報の全文を蓄積し、そのまま全文を検索する方法
- (b) 情報の内容をコード化して、検索する方法
- (c) 情報の中のキーワードなどを手がかりとして検索する方法

の三つが考えられるが、(a) の方法は、これに適する新しい装置および方法が開発されない限り、技術文献のように長文のものには不適当であろう。(b) は最も普通の方法であり、(c) は電子計算機と結びついて EDPS 方式などに取り入れられている方法であるが、この両者を比較して見ると、キーワードを用いる方法の利点は

- (1) 処理作業を人手から解放して、機械に任せることが出来る。
- (2) 主題を分析し、内容を判断する専門家を必要としない。
- (3) 直接原表題 (または原文) の群が検索できる。
- (4) 文献目録や索引書作成に便利である。

一方この方法の欠点は

- 1) 専門家が内容判断を行なっていないので、ノイズが多く、価値の低いものも多数出てきて、利用上不便である。
- 2) アウトプットはそのキーワードに関する一群の文献で、その量が多く、その中から目的の文献を選ぶ労力と時間が大変である。

- 3) 主題分析による内容判断は不要としても、インプットの準備作業において、キーワードの変化（語尾変化、単複数、接頭語、接尾語、綴り、読み方の変化など）や同語異義、異語同義その他専門的判断を要することが多数あり、やはり判断作業は避けられない。
- 4) 国語が異なるものを、いっしょには取り扱えない。（日、英、独、仏、ソ、伊、西、中、その他）。（キーワードを国語ごとに何重にも作らねばならない）
- 5) 日本語は英文かローマ字書きに直さねばならないし、アウトプットも英文かローマ字で出てくるので、現在の日本では実用上不便である。
（カナモジの使える機械ができたとしても、似たような不便さがある。）

これに対し、コード化する方法の利点は

- (1) 内容が一度判断してあるので、ノイズが少なく、利用価値が高い。
- (2) コードの方がキーワードよりも内容を正確につかむことができる。
- (3) 国語が異なっても、いっしょに取り扱うことができる。
- (4) 日本語も英文やローマ字に直す必要がない。
- (5) アウトプットの量はキーワード法に比べて少ないので、その中から目的の文献を探す労力と時間は少なくてすむ。

この方法の欠点は

- 1) すべてを機械でやることができず、主題を分析し内容を判断し、コード化するための専門家を必要とする。
- 2) コード化する人と検索する人との主観、判断の相違によって、食い違いができる。
- 3) 直接に表題（または原文）が検索されず、一度中間媒体のコードが検索され、それを表題または原文に変換しなければならない。
- 4) 文献目録や索引書作成には不便。

以上いずれの方法も長短得失があるので、結局そのどこに重点を置くかと言う事になるが、キーワードを用いる方法も、現在のところまだ不完全で、文献目録の作成や速報には便利であっても、文献検索には満足な方法とは言いがたい。国語の障壁を乗り越えられない事や、日本語はローマ字に書き改めねばならないのも現状では実用性に乏しい。したがって検索されたものの利用価値から考えると、コード化する方式の方が有利と言えるのではあるまいか。