

# 情報資料의 利用頻度順에 의한 排架法

崔 仁 壽\*

圖書館이나 情報團體에서 情報資料를 書架에 排列시키는 方法에는 여러 가지가 있으나 가장 보편적인 경우를 예로 들어 설명하면 特殊系統이 아닌 一般圖書館은 듀이十進分類法을, 科學技術系統의 資料를 다루는 特殊圖書館은 國際十進分類法을 사용하므로, 特殊圖書館 中 貸出이 主 業務인 貸出圖書館은 表題의 알파벳順으로 각각 情報資料를 書架에 排列시킬 것이다.

調査檢索시스템이 날로 복잡해지는 오늘날에 있어서 필요한 情報資料를 아무 지장없이 신속히 書架에서 찾아볼 수 있어야 함은 매우 중요하다. 이렇게 하자면 될 수 있는 한 情報資料에 接近하는데 걸리는 時間, 즉 接近時間(access time)을 최대한으로 줄여야 한다.

本稿에서는 이 接近時間을 최대한으로 줄이는 情報資料의 排架法을 論하고, 또 이 排架法의 妥當性を 종래의 主題別, 알파벳順 配置와 비교해 論하고 있다. 마지막으로 이 理論을 바탕으로 KORSTIC의 경우를 實例로 들어 가능성 있는 가장 실제적인 情報資料의 排架法을 提案한다.

## 1. Zipf의 法則

G. K. Zipf<sup>1)</sup>는 英語와 같은 言語를 사용할 때 어떤 單語의 使用頻度分布가 어떻게 되어 있는가를 다음과 같은 公式으로 나타낸 바 있다.

$$f \times r = K \dots\dots\dots (1)$$

여기서  $r$ 은 어떤 單語의 使用順位,  $f$ 는 이 單語의 使用頻度數,  $K$ 는 常數이다. 즉 가장 빈번히

사용되는 單語의 頻度數는  $f_1$ 이 되며; 이때 이 單語의 使用順位  $r$ 은 1이 된다. 다음 두번째로 빈번히 사용되는 單語의 頻度數는  $f_2$ 가 되며 이때의  $r$ 은 2가 된다.

$$f_1 \times 1 = f_2 \times 2 = f_3 \times 3 = \dots\dots\dots = f_n \times n = K$$

다시 말하면

$$(\text{單語의 使用頻度數}) \times (\text{이 單語의 使用順位}) = (\text{一定})$$

하다는 뜻이다.

이와 같은 言語에 관한 Zipf의 法則을 情報資料에 適用시킴이 妥當한가를 T. P. Fleming과 F. G. Kilgour<sup>2)</sup>의 데이터를 이용하여 살펴보려 한다. 이들은 大規模 圖書館에서의 雜誌類 利用頻度數와 그 順位를 調査한 바 있다(그림1참조).

그림 1에서 가로軸은 雜誌類의 利用順位( $r$ ), 세로軸은 雜誌類의 利用頻度數( $f$ )이다.  $r \times f = \text{一定}$ 이라는 Zipf의 法則을 調査하기 위하여 그림 1의  $1 \times f_1, 2 \times f_2, 3 \times f_3$  등을 계산해보면  $r > 10$ 에서  $r \times f$ 는 거의 一定함을 알 수 있다.

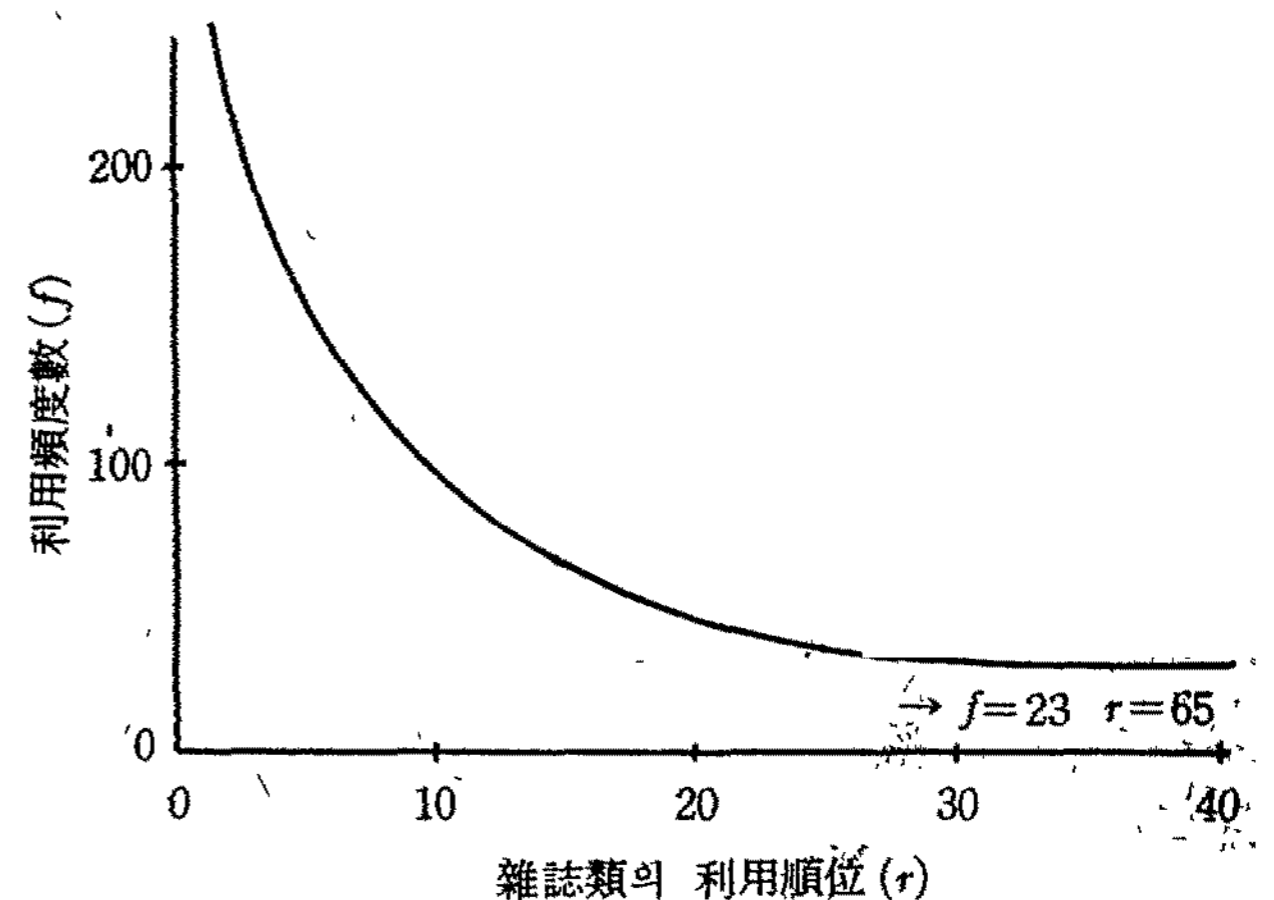


그림 1. 雜誌類의 利用頻度와 그 順位

\*KORSTIC 情報處理部

따라서  $r$ 이 10을 넘어설 때에는 Zipf의 法則이 有用하며,  $r < 10$ 일 때에는

$$r \times f = 1200 \{1 - (0.8)^r\}$$

이라는 式이 더 妥當할 것 같다. 또 Zipf의 法則은 단지  $r$ 值가 작을 때에만 오차가 있는데 현재까지는 이 오차의 根本原因이 매우 많이 요구되는 資料에 대해서 複寫서비스를 충분히 해주지 못하는 데에 있을 것이라고 알려져 있다.

Zipf의 法則은 利用頻도가 가장 높은 처음 몇 개의 資料에만 成立되지 않으므로 이 法則을 情報資料에도 適用시킬 수 있다. 따라서 本稿를 통해서  $r \times f = K$ 라는 公式을 두루 사용하고자 한다.

## 2. 接近距離 (access distance)에 重點을 둔 情報資料의 排架

情報資料를 어떻게 書架에 排列시키면 利用者 (開架式인 경우) 혹은 貸出擔當者가 最小의 接近時間으로 자기가 바라는 資料를 書架에서 골라낼 수 있을 것인가 하는 것이 本項에서 다룰 問題點이다. 이 問題에 관한 分명한 解答은 가장 빈번히 이용되는 資料를 接近點 (access point)에서 가장 가까운 자리에 配置시키는 것이다. 따라서 本項에서는 이러한 排架法이 接近時間을 最小化시킨다는 것을 數學的으로 전개해 나가겠다<sup>3)</sup>

資料  $b_i (i=1, 2, \dots, N)$ 의 利用頻度數를  $f_i$ 라 하고,  $b_i$ 의 接近距離가  $l_i$ 가 되게끔 書架에  $b_i$ 를 配置시켰다 하자. 만약 資料  $b_i$ 를  $f_1 > f_2 > f_3 \dots > f_i \dots > f_N$ 이 되게끔 書架에 配置시켰다고 하면  $N$ 개의 資料를 골라내는데 요구되는 全体 接近距離  $D$ 는 다음과 같은 式으로 表示된다.

$$D = \sum_{i=1}^N l_i f_i \dots \dots \dots (2)$$

(2)式에서 만약 資料를 그 利用頻도가 가장 높은 것을 接近點에서 가장 가까운 곳에, 가장 利用頻도가 낮은 것을 接近點에서 가장 먼 곳에, 즉  $l_1 < l_2 \dots < l_i \dots < l_N$ 이 되게 配置시키면  $D$ 는 最小가 된다. 이 理由는 다음과 같다. 만약 두 가지 資料  $b_i, b_j$ 를 서로 位置 變更시키면 이로 인한  $D$ 의 變化  $Q$ 는

$$Q = (l_i f_i + l_j f_j) - (l_j f_i + l_i f_j) = -(l_i - l_j) (f_i - f_j) \text{가 된다.}$$

假定에서  $f_i > f_j, l_i < l_j$ 이므로  $Q \geq 0$ 이 된다.  $Q$ 가 0인 경우는  $f_i = f_j$  혹은  $l_i = l_j$ 인 때이다. 따라서 資料를  $f_1 > f_2 \dots > f_N, l_1 < l_2 \dots < l_N$ 이 되게 書架에 配置시키면  $D$ 가 最小가 되는 것이다.

이제까지의 理論을 바탕으로 지금부터는 實例를 들어가면서 排架法을 살펴보고자 한다. 첫째 가장 간단한 例는 그림 2와 같이 1개의 近 書架와 接近點  $A$ 가 이 書架의 한쪽 端에 있는 경우이다. 물론 이 때에 資料는 그 利用頻도가 가장 높은 것부터 차례로 接近點  $A$ 에 가깝게 配置시킨다.

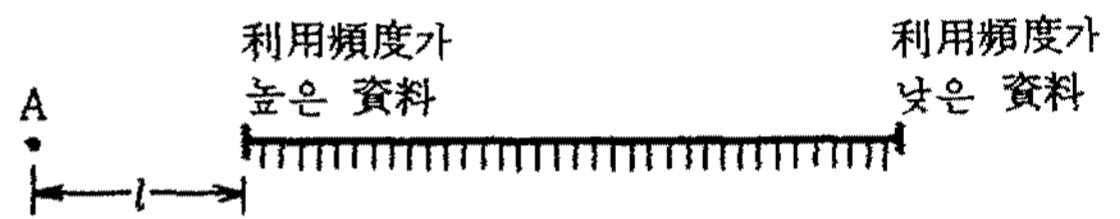


그림 2. 線形 書架—一端 接近

만약 資料의 平均두께를  $p$ 라 하고 接近點  $A$ 로부터 書架의 처음 部分까지의 距離를  $l$ 이라고 하면 주어진 時間內에 利用者 혹은 貸出擔當者가 資料를 찾기 위해 書架를 왕래한 距離  $D_2$ 는 다음과 같다.

$$D_2 = \sum (l + ip) f_i = l \sum f_i + p \sum i f_i \dots \dots \dots (3)$$

둘째 例는 그림 2와는 달리 接近點  $A$ 를 書架의 中心位置에 두는 경우이다(그림 3 참조). 이 때의 接近距離  $D_3$ 는 式(4)로 表示된다.

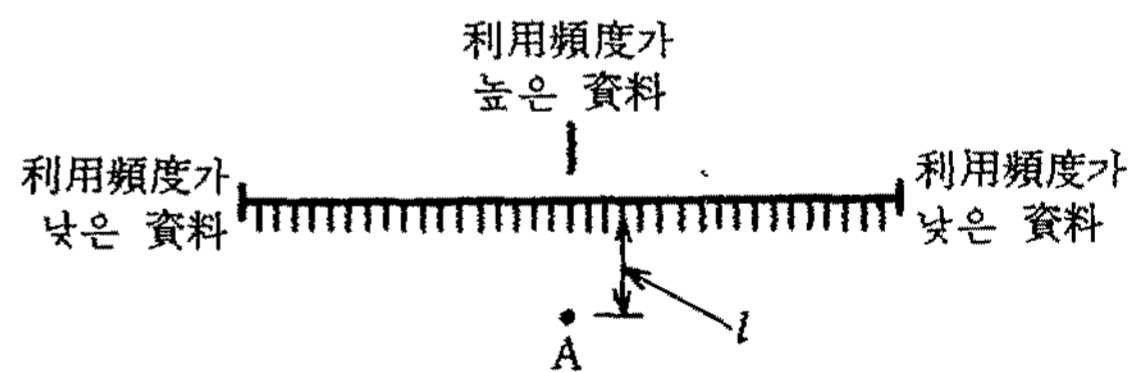


그림 3. 線形 書架—中央 接近

$$D_3 = l \sum f_i + \frac{p}{2} \sum i f_i \dots \dots \dots (4)$$

그림 3의 變型으로서 그림 4 및 그림 5를 들 수 있다. 물론 그림 4 및 그림 5의 接近距離  $D_4$  및  $D_5$ 는  $D_3$ 와 같다.

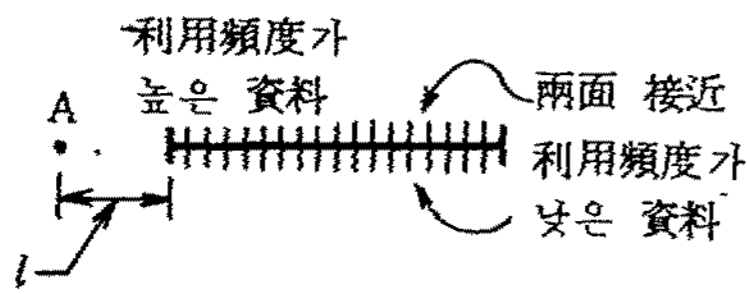


그림 4. 線形書架—兩面 接近

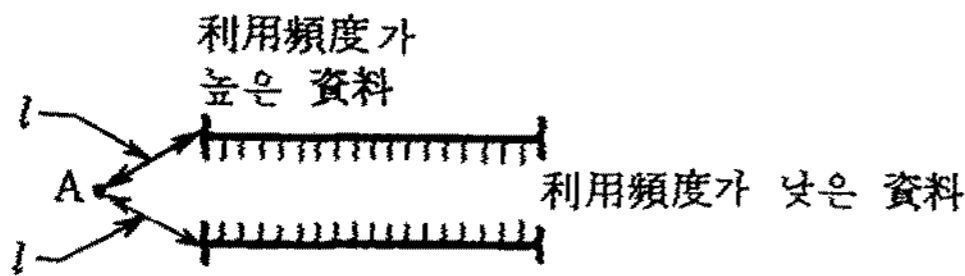


그림 5. 線形書架(廊下式)—一側 接近

작은 數値의  $l$ 과 資料數  $N$ 이 클 때에는 式(3)과 式(4)에서  $l \sum f_i$ 는 무시되므로  $D_3 \approx \frac{1}{2} D_2$ 가 된다.

機械化된 圖書館에서는 이제까지 言及한 單一段 線形書架가 매력적일지는 모르지만 이와 같은 單一段 線形書架는 보통 圖書館에서는 現實性이 없다. 따라서 셋째 例로 이제까지와는 달리 單一段을 쓰지 않고 多段書架 ( $S$ 段)를 들 수 있다. 이 때에는 일반적으로 書架의 길이 數値가 書架의 個數  $S$ 보다 매우 크기 때문에 垂直方向의 接近時間은 무시할 수 있다.  $S$ 段 書架를 式(3), (4)에 適用시키면 이는 각각 式(5), (6)과 같이 된다.

$$D_2 = l \sum f_i + \frac{p}{S} \sum if_i \dots \dots \dots (5)$$

$$D_3 = l \sum f_i + \frac{p}{2S} \sum if_i \dots \dots \dots (6)$$

그림 5와 같은 形態는 단차 書庫形態가 긴 廊下와 같이 되지 않으면 안될만큼 面積의 제한을 받지 않는한 거의 채택되지 않는다. 넷째 例로 그림 5를 약간 變更시켜 接近點  $A$ 를 中央에 놓는 경우를 들 수 있다(그림 6 참조).

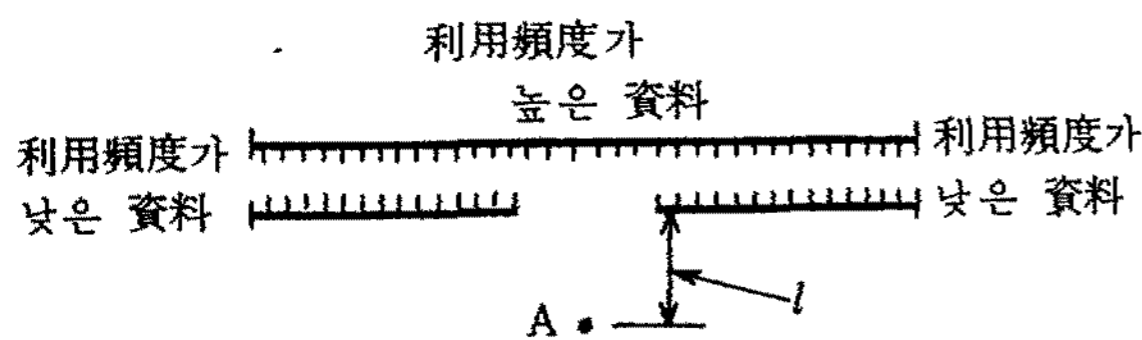


그림 6. 線形書架(廊下式)—中央 接近

그림 6에서의 接近距離  $D_6$ 은 다음과 같다.

$$D_6 = l \sum f_i + \frac{p}{4S} \sum if_i \dots \dots \dots (7)$$

式(7)은 式(6)을 발전시킨 것이다. 式(7)을 더욱 발전시킨 것이 다섯째 例가 된다(그림 7, 8 참조).

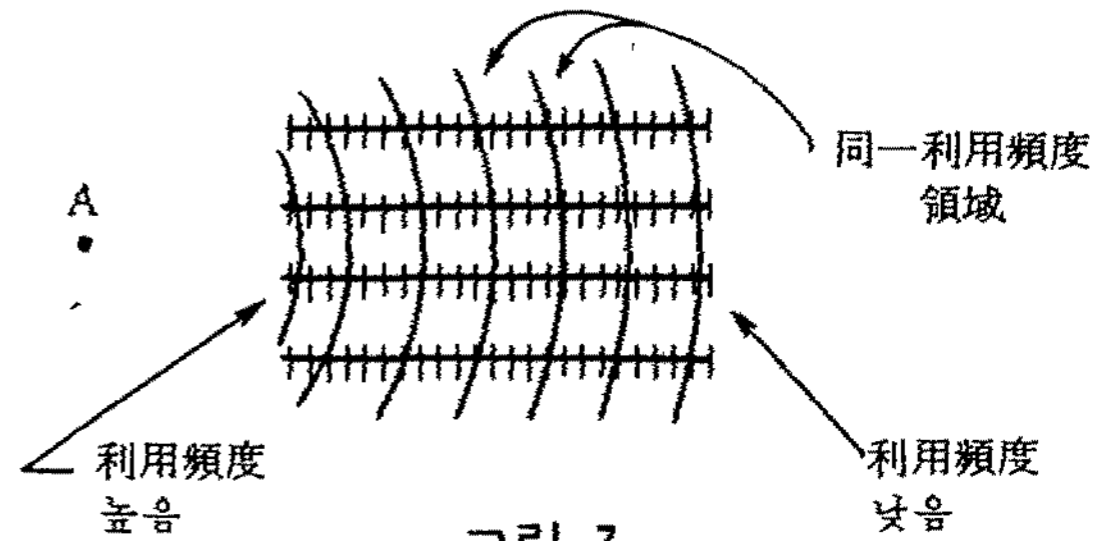


그림 7.

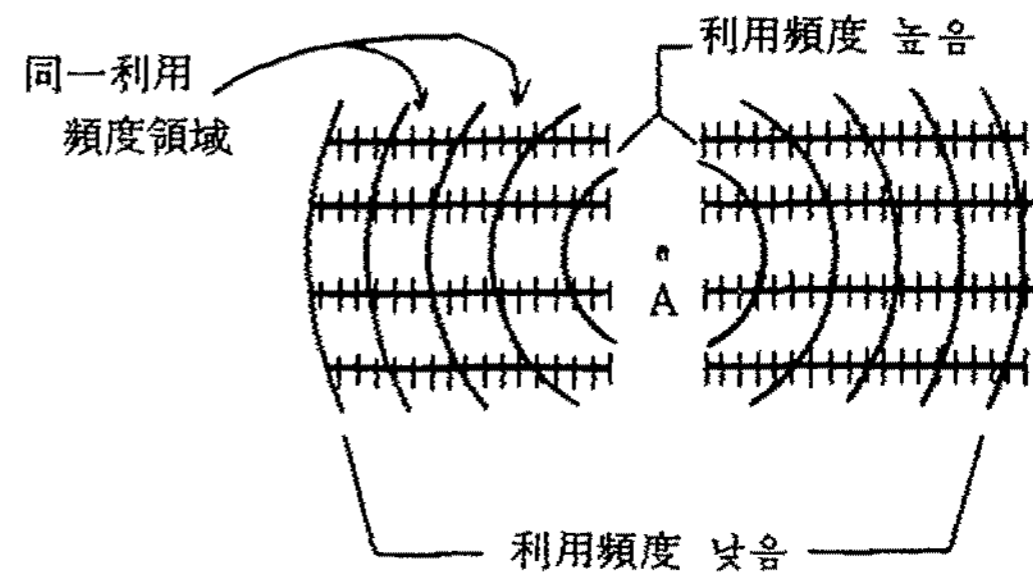


그림 8.

그림 7, 8과 같은 시스템은 매우 복잡하다. 왜냐하면 同一 利用頻도 領域이 그림에서 보다시피 卍 둘레로 형성되어 있기 때문에 이 領域에 정확히 資料를 配置시키기 힘들기 때문이다. 그러나 全体 接近距離를 計算해 보면 이제까지 記述한 어떤 시스템보다 우수한 結果를 얻는다. 그림 7 및 그림 8의 接近距離를 각각  $D_7$  및  $D_8$ 이라 하면  $D_7, D_8, D_6$ 와의 關係는 다음과 같다.

$$D_8 \approx \frac{1}{2} D_7 \approx \frac{1}{T} D_6 \dots \dots \dots (8)$$

여기서  $T$ 는 그림 8에서 書架 列의 個數이다. 前述한 다섯가지 方法外에 몇가지를 더 살펴보면 다음 그림과 같다.

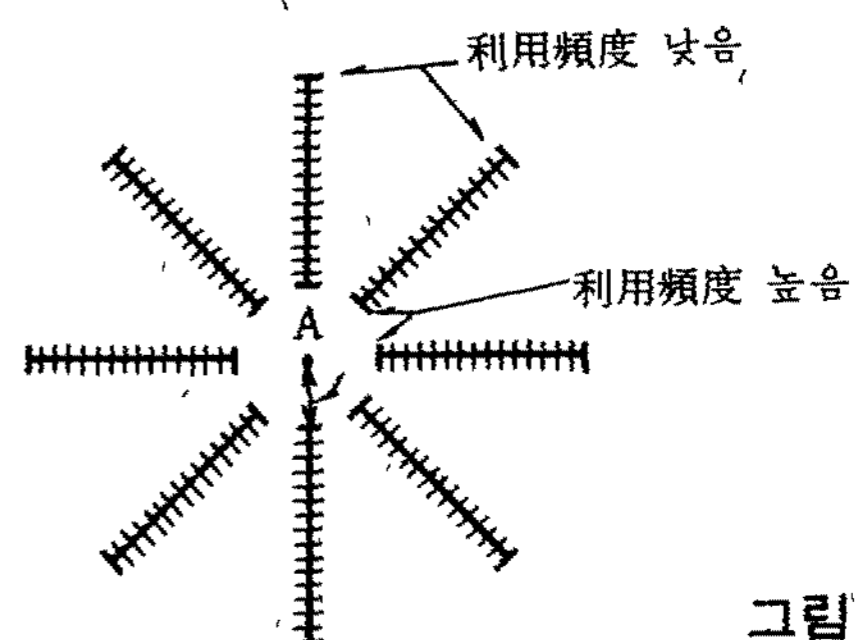


그림 9.

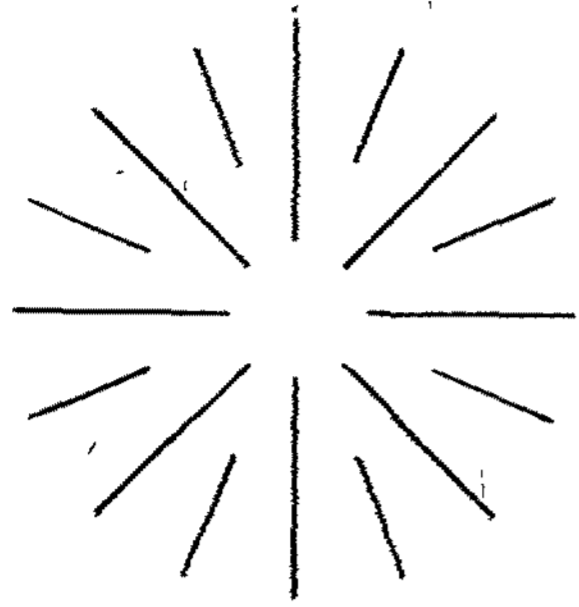


그림 10.

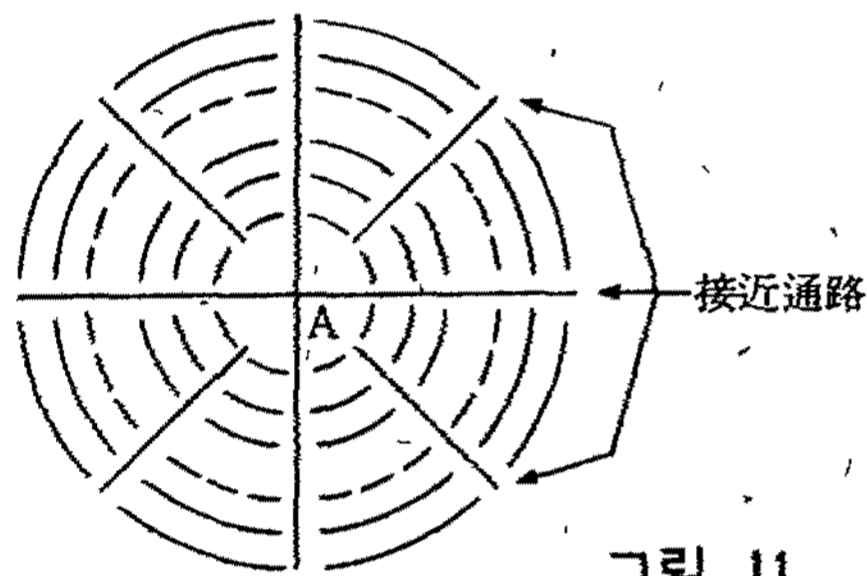


그림 11.

### 3. 理論의 妥当性

資料를 利用頻度順으로 排列하는 것이 圖書館 傳統시스템인 主題別, 著者別, 알파벳順 排列보다 얼마나 더 能率的인가를 살펴보기로 하자.

그림 5와 같은 排架法을 例로 들면,

$$D_5 = l \sum f_i + \frac{p}{2S} \sum i f_i \text{가 된다.}$$

利用頻度順이 전혀 고려되지 않는 主題別 혹은 著者別 排列에서의 接近距離  $D_5^1$ 은 다음과 같다.

$$D_5^1 = \sum \left( 1 + \frac{Np}{2 \times 2S} \right) f_i$$

$$= l \sum f_i + \frac{p}{2S} \sum \frac{N}{2} f_i$$

$N$ 이 크고  $l$ 이 무시될 때에는 Zipf의 法則  $f_i = K/i$ 가 成立된다. 따라서

$$D_5 = \frac{p}{2S} \sum K i / i = \frac{KNp}{2S}$$

$$D_5^1 = \frac{p}{2S} \sum \frac{KN}{2i} = \frac{KNp}{4S} \sum \frac{1}{i} \approx \frac{KNp}{4S} \log_e N$$

즉  $\frac{D_5^1}{D_5} = \frac{1}{2} \log_e N \dots\dots\dots (9)$

이 된다. 만약  $N$ 이 10,000이라면 式(9)에서  $D_5^1 = 4.6D_5$ 가 된다. 즉 그림 5의 경우에는 利用頻度別로 書架에 情報資料를 配置하는 것이 主

題別, 著者別과 같은 方法으로 配置시키는 것보다  $N$ 이 10,000인 경우 接近距離(혹은 接近時間)가 4.6배나 단축된다는 것이다. 資料의 數  $N$ 이 增加하면 할수록 利用頻度順으로 配置시키는 것이 他 경우보다 점점 더 能率的으로 될 것이다.

### 4. KORSTIC의 排架法에 관한 提案

接近距離에 重點을 두어 情報資料를 書架에 排列시키는 方法은 2項에서 記述한 것 외에도 매우 많이 있다. 그러나 前述한 理論에 근거를 들 때에 KORSTIC에 가장 적절한 시스템은 그림 8과 같은 시스템을 약간 變更시킨 것일 것이다.

1975年 한해동안 KORSTIC 利用者들이 複寫해 간 文獻(特許 제외) 중에서 가장 많이 利用된 150種의 雜誌目錄은 表 1과 같다.

表 1을 보면 順位 1位인 'New Food Industry'는 1975年 한해동안 355번 이용되었고, 150位인 'Adhesive Age'는 32번 이용되었다. 이처럼 1位와 150位 雜誌의 利用頻度數에는 큰 差異가 있으므로 表 1의 150種 雜誌만을 골라 이를 KORSTIC 全体文獻의 샘플로 삼아도 무리가 없을 것으로 생각한다.

表 1의 150種 雜誌를 편의상 數學·物理, 電氣·電子, 機械·金屬·鑛山·建設, 化學·化學工業, 生物·農業·藥學의 5分野로 나누었다. 또 雜誌利用順位 區間을 6等分하여 各 區間에 25種씩 들어가게 하여 上記 5分野別로 種數 및 利用頻度數를 정리한 것은 表 2와 같다. 表 2에 의하면 1~25位에서는 生物·農業·藥學分野가 化學·化學工業分野보다 利用效率이 높으나, 26~50位에서는 이와 反對가 된다. 150種 全体에 대해 알기 쉽게 나타낸 것이 表 3이다.

表 3에 의하면 上位 150種 雜誌중에 化學·化學工業分野의 雜誌가 37.33%에 해당되는 56種이나 포함되어 있으며 그 利用百分率は 36.57%가 된다. 生物·農業·藥學分野는 약간 뒤떨어지나 化學·化學工業分野와 거의 비슷하다. 結論적으로 KORSTIC所藏文獻의 分野別 有用性 順位는 化學·化學工業, 生物·農業·藥學, 機械·金屬

表 1. 利用頻度順 雜誌目錄

| 順位  | 雜誌名   | 分野       | 利用頻度數 |
|-----|---|----------|-------|
| 1   | New Food Industry                             | 生物・農業・藥學 | 355   |
| 2   | 食品工業  | "        | 209   |
| 3   | Journal of Farm Economics                     | "        | 170   |
| 4   | Journal of Food Science                       | "        | 149   |
| 5   | Analytical Chemistry                          | 化學・化學工業  | 145   |
| 6   | Journal of Pharmaceutical Science             | 生物・農業・藥學 | 140   |
| 7   | Journal of the American Chemical Society      | 化學・化學工業  | 133   |
| 8   | 食品と科學   | 生物・農業・藥學 | 131   |
| 9   | 日本釀造協會誌                                       | 生物・農業・藥學 | 127   |
| 10  | 油化學   | 化學・化學工業  | 126   |
| 11  | Agricultural and Biological Chemistry         | 生物・農業・藥學 | 119   |
| 12  | 工業化學雜誌  | 化學・化學工業  | 119   |
| 13  | Food Technology                               | 生物・農業・藥學 | 115   |
| 14  | 化學工場  | 化學・化學工業  | 102   |
| 15  | 油脂  | "        | 93    |
| 16  | Journal of the American Oil Chemists' Society | "        | 90    |
| 17  | Journal of Organic Chemistry                  | "        | 88    |
| 18  | Rubber Age                                    | "        | 87    |
| 19  | 高分子加工   | "        | 87    |
| 20  | Journal of Agricultural and Food Chemistry    | 生物・農業・藥學 | 85    |
| 21  | 醱酵協會誌   | "        | 84    |
| 22  | 缶詰時報  | "        | 84    |
| 23  | 高分子論文集  | 化學・化學工業  | 83    |
| 24  | Textile Research Journal                      | "        | 82    |
| 25  | 日本農藝化學會誌                                      | 生物・農業・藥學 | 82    |
| .   | .   | .        | .     |
| .   | .   | .        | .     |
| .   | .   | .        | .     |
| 150 | Adhesives Age                                 | 化學・化學工業  | 32    |

表 2. 分野別 利用分布

| 分野          | 1~25位 |       | 26~50位 |       | 51~75位 |       | 76~100位 |       | 101~125位 |       | 126~150位 |       |
|-------------|-------|-------|--------|-------|--------|-------|---------|-------|----------|-------|----------|-------|
|             | 種數    | 利用頻度數 | 種數     | 利用頻度數 | 種數     | 利用頻度數 | 種數      | 利用頻度數 | 種數       | 利用頻度數 | 種數       | 利用頻度數 |
| 數學・物理       | 0     | 0     | 3      | 225   | 1      | 62    | 0       | 0     | 0        | 0     | 3        | 105   |
| 電氣・電子       | 0     | 0     | 4      | 298   | 4      | 207   | 3       | 139   | 4        | 162   | 4        | 137   |
| 機械・金屬・鑛山・建設 | 0     | 0     | 2      | 135   | 6      | 330   | 7       | 319   | 7        | 280   | 3        | 98    |
| 化學・化學工業     | 12    | 1235  | 9      | 644   | 8      | 447   | 10      | 455   | 7        | 266   | 10       | 343   |
| 生物・農業・藥學    | 13    | 1850  | 7      | 504   | 6      | 365   | 5       | 232   | 7        | 261   | 5        | 172   |

· 鑛山·建設, 電氣·電子, 數學·物理分野의 順序이다.

수 많은 各種 雜誌를 利用頻度가 높은 것에서 부터 낮은 順으로 書架에 排列시킨다는 것은 쉬운 일이 아니며, 또한 반납된 雜誌를 順位別로 다시 書架에 配置시키는 것도 쉽지 않은 일이다. 그러므로 現實性을 감안해 볼 때 前述한 다섯 分野만으로도 優先順位를 정하는 것이 좋을 것 같다. 또한 円둘레 모양으로 利用頻度領域을 區分한 그림 8의 복잡한 시스템을 그림 12와 같이 變更시키는 것이 現實에 비추어 적절할 것 같다.

表 3.

| 分 野         | 全体 種數       | 全体 利用頻度數     |
|-------------|-------------|--------------|
| 數學·物理       | 7(4.67)     | 392(4.23)    |
| 電氣·電子       | 19(12.67)   | 943(10.17)   |
| 機械·金屬·鑛山·建設 | 25(16.67)   | 1162(12.53)  |
| 化學·化學工業     | 56(37.33)   | 3390(36.57)  |
| 生物·農業·藥學    | 43(28.66)   | 3384(36.50)  |
| 合 計         | 150(100.00) | 9271(100.00) |

괄호 안은 百分率

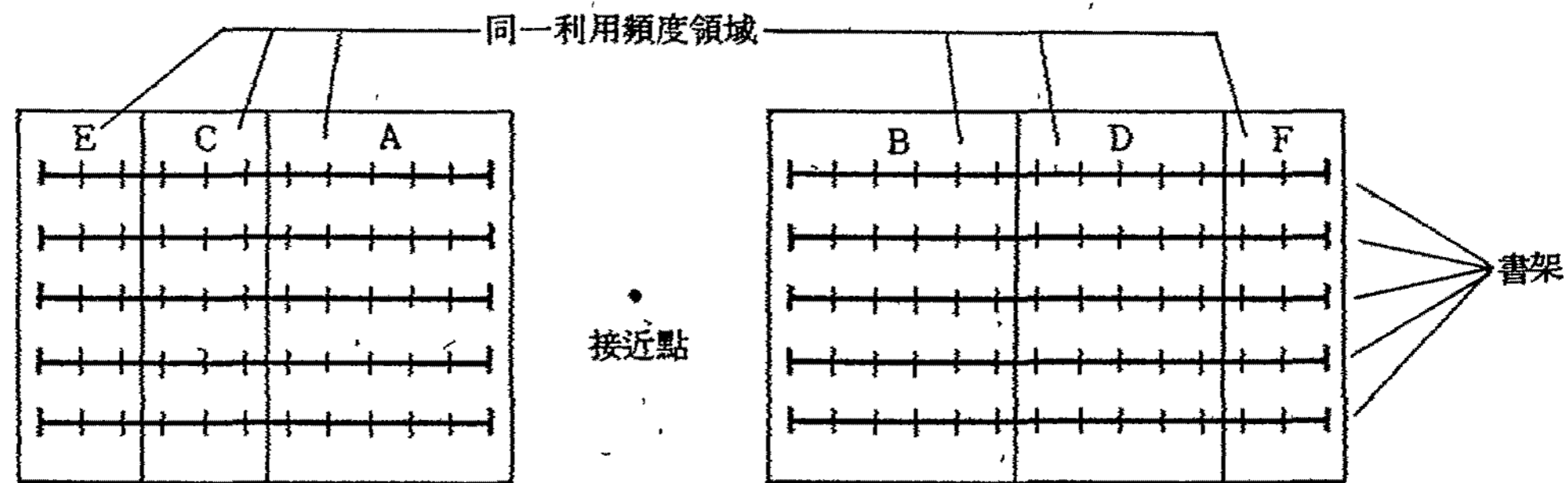


그림 12. KORSTIC의 書架配置

그림 12와 같은 雜誌의 排架法이 本稿를 통해 提案할 수 있는 KORSTIC의 排架法이다. A, B 두 領域중 어느 한 領域은 化學·化學工業分野의 雜誌가, 나머지 한 領域은 生物·農業·藥學分野의 雜誌가 차지하게끔 該當分野의 雜誌를 排架시키고 C, D 두 領域중 어느 한 領域은 機械·金屬·鑛山·建設分野의 雜誌가, 나머지 한 領域은 電氣·電子分野의 雜誌가 차지하게끔 該當分野의 雜誌를 排架시키며, 數學·物理分野의 雜誌를 E, F 두 領域중 어느 한 領域이나 혹은 E, F 두 領域 모두에 分散시켜 排架시키는 方法인 것이다.

參 考 文 獻

- 1) ZIPF, G. K. Human behaviour and the principle of least effort. Addison-Wesley, Mass., 1949.
- 2) FLEMING, T. P. & KILGOUR, F. G. Moderately and heavily used biomedical journals. Proc. 2nd Int. Congress on Med. Lib., 1963, pp. 234~241.
- 3) BOOTH, A. D. On the geometry of libraries. Journal of Documentation, v. 25, n. 1, 1969, pp. 28~42.