

情報檢索效率에 관한 研究

尹 龜 鎬*

<目 次>

- | | |
|---------------|---------------|
| 1. 緒 論 | 5. 檢索效率과 不確實性 |
| 2. 檢索效率의 評價基準 | 6. 檢索效率의 價値性 |
| 3. 情報의 適合性 | 7. 檢索效率의 支配要因 |
| 4. 再現率과 精度率 | 8. 結 論 |

1. 緒 論

檢索效率(Retrieval Effectiveness)은 情報檢索시스템評價의 第一基準으로서 情報檢索(IR : Information Retrieval)이라는 말이 널리 쓰이기 始作한 1960年代 初부터 많은 研究와 試驗의 核心領域이 되었다.

第二次 世界大戰以後 科學技術의 急速한 發展으로 인한 情報量의 幾何級數的 增加와 수 많은 情報利用者들의 多樣化된 情報要求는 既存의 圖書館 및 情報시스템에 많은 變化를 불러 일으켰다. 소위 “情報化時代”에 대처할 수 있는 새로운 情報處理技術의 開發을 要求하였다.

1950年代 初 Taube, Mooers 등에 의한 새로운 索引作成法의 試圖과 때마침 實用化되기 始作한 電子計算機의 利用은 情報檢索시스템의 高度의 機械化 내지는 自動化를 前提로 한 보다 效率的인 情報시스템 構築을 促進함으로써 시스템의 評價는 必然的으로 수행되어야 했다.

1953年 지금의 DDC(美國의 Defense Documentation Center)의 前身인 ASTIA (Armed Services Technical Information Agency)가 試圖한 最初의 評價試驗¹⁾이후 情報檢索시스템의 評價를 위한 많은 現場試驗과 研究試驗

* 啓明大學校圖書館學科 助教授

1) 이 評價試驗은 ASTIA가 開發한 主題名標目表에 의한 在來式 主題名目錄과 Mortimer Taube博士가 提案한 유니덱시스템간의 索引시스템의 性能을 比較 研究한 것이다.

이 계속되어 왔다.²⁾

특히 두번에 걸쳐 수행된 Cleverdon의 “Cranfield Project”와 Lancaster의 “MEDLARS Evaluation” 및 Salton의 “SMART System” 등은 代表的 評價試驗으로서 索引言語의 性能과 檢索效率의 測定에 관한 여러 가지 理論을 提示하였고 많은 事實을 確認시켜 주었다.

그러나 情報檢索시스템 評價에 있어 가장 重要한 問題는 適用된 評價基準와 評價方法의 客觀的 妥當性 및 數量的 測定の 正確性이라고 본다. 오늘 날 情報檢索시스템 評價에 가장 널리 適用되고 있는 評價基準(檢索效率, 檢索時間, 檢索費用)은 어떤 측면은 이들이 비교적 용이하나 어떤 측면은 이들이 매우 어렵거나 거의 不可能하기 때문에 뚜렷한 意見의 一致를 보지 못하고 論難의 對象이 되고 있다.

특히 情報의 適合性이라는 概念이 介入되는 檢索效率은 利用者의 質問(情報要求)에 대한 文獻의 適合性判定이 다분히 主觀的일 수 밖에 없을 뿐만아니라 여러가지 變數要因에 의해 영향을 받기 때문에 正確한 數量的 測定이 매우 어렵다는 점에서 研究와 試驗의 焦點이 되고 있다.

따라서 本稿에서는 情報檢索시스템 評價에 관한 先行研究를 考察하여 첫째, 檢索效率의 評價基準 및 測定技法, 情報의 適合性判定 및 檢索效率의 評價에 관한 最適理論과 方法을 抽出하여 보고, 둘째 檢索效率의 여러 가지 支配要因을 分析하여 檢索效率向上을 위한 諸 方案을 提示하여 보고자 한다.

아직 우리나라 圖書館 및 情報시스템의 現場評價試驗이 수행된 일이 별로 없음을 감안할 때 本研究가 앞으로 반드시 수행되어야 할 情報檢索시스템 評價에 다소나마 도움이 될 수 있으리라고 믿는 바이다.

2. 檢索效率의 評價基準

檢索效率은 “檢索에 의하여 얻어진 情報가 利用者의 要求에 一致하는 比

2) 情報檢索시스템 評價에 관한 先行研究는 A. C. Foskett : *The Subject Approach to Information*, 3rd ed. London, Clive Bingley, 1977. Chapter 28(p. 441-463)과 F. W. Lancaster : *Information Retrieval Systems ; Characters, Testing and Evaluation*, 2nd ed. New York, John Wiley, 1979. Chapter 19 (p. 273-278).

率”이라고 定義된다.³⁾ 즉, 利用者의 情報要求에 應答하기 위해 시스템이 관련 있는 情報 파일(file)에서 利用者의 要求를 만족시켜 줄 수 있는 情報를 探索해 내는 시스템의 能力이라고 할 수 있다.

그러므로 檢索效率은 實際적으로는 시스템의 축적된 文獻파일(Document File)에서 利用者의 情報要求에 알맞는 適合文獻(Relevant Documents)과 알맞지 않은 非適合文獻(Irrelevant Documents)의 檢索比率로서 決定되는 것이다.

이와 같은 檢索效率의 評價는 利用者의 情報要求에 대해 檢索시스템이 探索할 시스템의 蓄積情報(Database)를 適合情報과 非適合情報로 區分하여 나타난 探索結果가 數值로서 表示되어야 하기 때문에 圖表 1과 같은 “2×2테이블”(Two-by-two table ; 一名 分割表(Contingency table)라고도 함)이 널리 使用되고 있다.

〈圖表 1〉 蓄積情報의 探索結果

	<table border="1"> <tr> <td></td> <td>適合情報</td> <td>非適合情報</td> <td>全體情報</td> </tr> <tr> <td>檢索된 情報</td> <td>a(적중)</td> <td>b(잡음)</td> <td>a+b</td> </tr> <tr> <td>檢索되지 않은 情報</td> <td>c(누락)</td> <td>d(제거)</td> <td>c+d</td> </tr> <tr> <td>全體情報</td> <td>a+c</td> <td>b+d</td> <td>a+b+c+d</td> </tr> </table>		適合情報	非適合情報	全體情報	檢索된 情報	a(적중)	b(잡음)	a+b	檢索되지 않은 情報	c(누락)	d(제거)	c+d	全體情報	a+c	b+d	a+b+c+d
	適合情報	非適合情報	全體情報														
檢索된 情報	a(적중)	b(잡음)	a+b														
檢索되지 않은 情報	c(누락)	d(제거)	c+d														
全體情報	a+c	b+d	a+b+c+d														

圖表 1로 부터 아래와 같은 檢索效率의 여러가지 測定公式이 提案되었으나 오늘날까지 가장 널리 適用되고 있는 檢索效率의 評價基準은 再現率과 精度率이다.

$$\text{再現率}^{4)} = \frac{a}{a+c} \times 100\% ; \frac{\text{檢索된 適合情報}}{\text{全體의 適合情報}} \times 100\%$$

3) 司空哲, 情報檢索論, 서울, 亞細亞文化社, 1977, p.160.

4) Kent등은 “Recall Factor”로, Cleverdon등은 “Recall Ratio”로, Swets는 “Conditional Probability of a Hit”로, Goffman과 Newill은 “Sensitivity”로 命名하였다. 韓國語로는 再現率이 使用되고 있다

$$\text{精度率}^{5)} = \frac{a}{a+b} \times 100\% ; \frac{\text{檢索된 適合情報}}{\text{檢索된 全體情報}} \times 100\%$$

$$\text{非適合率}^{6)} = \frac{b}{b+d} \times 100\% ; \frac{\text{檢索된 非適合情報}}{\text{全體의 非適合情報}} \times 100\%$$

$$\text{適合率}^{7)} = \frac{a+c}{a+b+c+d} \times 100\% ; \frac{\text{適合情報}}{\text{全體情報}} \times 100\%$$

$$\text{雜音率}^{8)} = \frac{b}{a+b} \times 100\% ; \frac{\text{檢索된 非適合情報}}{\text{檢索된 全體情報}} \times 100\%$$

$$\text{漏落率}^{9)} = \frac{c}{a+c} \times 100\% ; \frac{\text{檢索되지 않은 適合情報}}{\text{全體의 適合情報}} \times 100\%$$

$$\text{除去率}^{10)} = \frac{d}{b+d} \times 100\% ; \frac{\text{檢索되지 않은 非適合情報}}{\text{全體의 非適合情報}} \times 100\%$$

再現率과 精度率에 의한 檢索效率의 評價基準은 1955年 Kent등에 의해 提案¹¹⁾된 것이나 “Cranfield Project” 이후 一般的으로 “Cranfield Measures”라고 알려져 널리 쓰이기 始作하였다.¹²⁾

再現率과 精度率에 의한 檢索效率의 測定은 시스템의 蓄積情報量($N=a+b+c+d$)과 그 안에 있는 適合情報($a+c$)에 左右된다고 볼 수 있다. 따라서 이들의 比率를 나타내는 適合率에 의해 再現率과 精度率 및 非適合率이 影響을 받게 될 것이다. 이는 Fairthorne에 의해 최초로 暗示되었으나¹³⁾ 뒤에

5) Kent등은 “Pertinency Factor”로 Cleverdon등은 “Precision Ratio” 또는 “Relevance Ratio”로 命名하였으며, 간혹 “Acceptance Rate”라는 用語도 使用되고 있다. 韓國語로는 精度率, 正確率등이 使用되고 있다.

6) Swets는 “Conditional Probability of a False Drop”로 Cleverdon등은 “Fallout Ratio”로 命名하였으며, 간혹 “Discard”라는 用語도 使用되고 있다.

7) Cleverdon등이 “Generality”로 命名하였다.

8) “Noise Factor”로 使用되고 있다.

9) Fairthorne은 “Snobbery Ratio”로, Swets는 “Conditional Probability of a Miss”로 命名하였다.

10) Goffman과 Newill은 “Specificity”로, Swets는 “Conditional Probability of a Correct Rejection”로, Vickery는 “Dodged”로 命名하였다.

11) Kent, A. et al. “Machine Literature Searching”. *American Documentation*, 6 (1955), pp.93-101.

12) Cleverdon, C. W. “Progress in Documentation: Evaluation Tests of Information Retrieval Systems.” *Journal of Documentation*, 26(1970), pp.55-67.

13) Fairthorne, R. A. “Basic Parameters of Retrieval Tests.” *Proceedings of the American Documentation Institute*, 1(1964), pp.343-345.

Keen 등에 의해 明白하게 提示되었다.¹⁴⁾

이들은 適合率(G)이 檢索效率測定에 介入됨으로써 再現率(R), 精度率(P), 非適合率(F) 사이에 다음과 같은 汎函數관계가 성립된다고 하였다.

$$P = \frac{R \times G}{R \times G + F(1 - G)}$$

따라서 R, P, F, G 는 相互獨立인 것이 아니고 어느 것이든 세가지의 값이 決定되면 다른 하나의 값은 自動적으로 決定된다는 것이다. 이는 事實은 確率에 관한 Bayes의 定理(Bayes Theorem)를 解說한 것이라고 Van Rijsbergen 은 말하고 있다.¹⁵⁾

다음으로 適合率(G)이 再現率(R), 精度率(P) 및 非適合率(F)에 어떻게 영향을 미치는가를 살펴 본다. 利用者の 情報要求에 대해 시스템이 探索할 蓄積情報(N)가 一定한데 適合率(G)이 상승하면(또는 適合率(G)이 一定하고 蓄積情報(N)가 감소하면) 檢索된 適合情報量은 平均적으로 增加될 것이다. 따라서 適合率(G)의 상승은 再現率(R), 精度率(P), 非適合率(F)에 다음 그림과 같은 영향을 미치게 될 것이다.

$$R = \begin{array}{c} \uparrow \\ - \\ \uparrow \end{array}, \quad P = \begin{array}{c} \uparrow \\ - \\ \rightarrow \end{array}, \quad F = \begin{array}{c} \rightarrow \\ - \\ \rightarrow \end{array}$$

위 그림에서 分母와 分子가 같은 方向으로 움직이는 再現率과 非適合率は 適合率이 變化하더라도 論理的으로 대략 一定한 값을 유지한다고 보겠으나 精度率は 適合率의 變化에 따라 직접적으로 값이 달라질 것이다.

이러한 觀點에서 檢索效率의 評價基準으로 널리 適用된 再現率과 精度率에 의한 效率測定이 一般的으로 바람직하지 못하다는 反論이 상당히 提起되었으며 相對적으로 再現率과 非適合率에 의한 測定이 適合率과 蓄積情報量の 變化에 별로 영향을 받지 않는다는 점에서 새로운 評價基準으로 提案되

14) Cleverdon, C. W., Mills, J., and Keen, M. "Factors Determining the Performance of Indexing Systems." Vol. I. "Design." Vol. II. "Test Results." College of Aeronautics, Cranfield, England.

15) Van Rijsbergen, C. J. "Retrieval Effectiveness." In: *Progress in Communication Sciences*, ed. by M. J. Voigt and G. J. Hanneman. Norwood (N. J.), ALEX Pub. Co., 1979. p. 95.

었다.¹⁶⁾

그러나 實際로는 蓄積情報量의 減小나 혹은 適合率의 增加에도 불구하고 精度率이 比較的 一定한데 反해 非適合率이 分明히 減小한다고 지적하고 再現率과 精度率, 또는 再現率과 非適合率에 의한 어느 檢索效率評價基準도 모든 상황에 만족스러운 것은 아니라고 Salton은 말하고 있다.¹⁷⁾

한편, 再現率과 精度率에 의한 檢索效率의 評價는 利用者 指向的이라고 볼 수 있다. 왜냐하면 이 基準은 시스템의 適合文獻의 檢索機能만을 나타내는 것으로 利用者는 一般的으로 適合情報의 檢索을 極大化하는데 關心이 높기 때문이다.

그러나 非適合率은 非適合情報의 排除能力(非檢索能力)의 測定으로 蓄積情報안에 있는 非適合情報量에 의해 영향을 받는다. 따라서 再現率과 非適合率에 의한 檢索效率의 評價는 適合文獻의 檢索機能과 非適合文獻의 檢索·排除機能을 나타내기 때문에 시스템 指向的이라고 할 수 있다.

두가지 評價基準이 모두 特殊한 環境이나 目的을 위해서 長短點을 가지고 있기 때문에 어느 基準이 보다 合理的이라고 단정할 수는 없다. 그러나 檢索시스템의 궁극적 목표가 利用者를 위해서 存在한다고 본다면 利用者 指向的인 檢索效率의 向上이 보다 바람직하다는 觀點에서 아마도 再現率과 精度率에 의한 評價基準이 훨씬 보편적으로 適用되고 있다고 볼 수 있을 것이다.

한편, 이와 같은 傳統的 評價基準들이 두가지 以上の 값을 적용하여 시스템의 檢索效率를 比較해야 하기 때문에 最近에는 Swets¹⁸⁾, Cooper¹⁹⁾, Robertson²⁰⁾ 등의 單一價에 의한 檢索效率의 評價基準(Single valued measures,

16) *Ibid.*, pp. 96-97.

17) Salton, G. "The Generality Effect and the Retrieval Evaluation for Large Collections," *Journal of the American Society for Information Science*, 23(1972), pp. 11-22.

18) Swets, J.A. "Effectiveness of Information Retrieval Methods," *American Documentation*, 20(1969), pp. 72-89.

19) Cooper, W.S. "Expected Search Length: A Single Measure of Retrieval Effectiveness based on Weak Ordering Action of Retrieval Systems," *Journal of the American Society for Information Science*, 19(1968), pp. 30-41.

20) Robertson, S.E. and Teather, D. "A Statistical Analysis of Retrieval Tests :

또는 Single number measures)이 提案되어 研究되고 있으나 여러가지 理論上的 問題로 아직은 보편적으로 活用되지 못하고 있다.

3. 情報의 適合性

情報의 適合性(Relevance 또는 Aboutness)이란 利用者の 特殊한 情報要求를 만족시켜 줄 수 있는 情報의 特性으로서 이미 앞에서 본 바와 같이 檢索效率測定의 바로미터가 되는 것이다.

情報檢索시스템의 檢索效率이 利用者の 特定情報要求를 만족시켜 줄 수 있는 適合情報(實際적으로는 適合情報가 들어 있는 適合文獻)의 檢索能力과 利用者の 特定情報要求에 否適合한 非適合情報(非適合文獻)의 檢索排除能力으로 測定되기 때문에 利用者の 情報要求에 대해 시스템이 探索할 情報의 적합성여부의 판단 즉, “適合性決定”(Relevance Decision)과 檢索된 情報의 “適合性的 程度”(Degree of relevance)는 檢索效率의 測定 및 評價의 열쇠가 된다.

A. 適合性決定

情報의 適合性이라는 概念은 評價尺度가 모호하여 그 決定이 매우 어려운 問題로서 學者들의 많은 關心과 論難의 對象이 되고 있다.

Rees와 Saracevic은 適合성에 關係 다음과 같이 質問하였다.²¹⁾

첫째, 適合성이 檢索시스템의 性能測定을 위한 적절한 基準인가?

둘째, 만일 적절한 基準이라면 適合성이란 무엇을 意味하는가?

셋째, 도대체 適合성이 實際로 使用될 수 있는가? 測定될 수 있는가? 어떤 制限이나 條件下에서 適用될 수 있는가?

Taube도 適合性的 概念에 대해 반박을 하였으며²²⁾ Rees는 또한 文獻과 利用者の 情報要求間的 適合성이란 때에 따라 또 個人에 따라 다른 것이지

a Bayesian Approach.” *Journal of Documentation*, 30(1974), pp. 273-282.

21) Rees, A.M. and Saracevic, T. “The Measurability of Relevance.” *Proceedings of the American Documentation Institute*, 3 (1966), pp. 225-234.

22) Taube, M. “A Note on the Pseudo-Mathematics of Relevance.” *American Documentation*, 16(1965), pp. 62-72.

항상 固定되어 있는 것이 아니라고 하였다.²³⁾

Lancaster는 지금까지 檢索시스템評價를 위한 많은 研究가 情報의 適合性에 의한 檢索效率의 測定을 試圖하였으나 情報의 適合性은 情報의 効用性(Pertinence)과 그 概念이 區分되어야 하며 진정한 意味의 檢索效率測定을 위해서는 두가지 評價尺度가 모두 適用되어야 한다고 主張하고 다음과 같이 適合性과 効用性에 대한 概念定義를 내렸다.²⁴⁾

適合性：“document-request statement of users” 사이의 適合性判定(Relevance judgments)으로서 一般的으로 探索主題에 關係 상당한 知識이 있는 情報專門家(Information specialists)나 혹은 主題專門家(Subject specialists)들에 의한 그룹査定으로 決定된다.

効用性：“document-information needs of users” 사이의 價値性判定(Value judgments)으로서 반드시 利用者自身에 의한 査定으로 決定된다.

上記 概念定義를 分析해 보면 情報의 適合性은 利用者の 情報要求(申請書에 의한 表現이나 口頭表現)에 대해 시스템의 情報專門家나 혹은 主題專門家들이 原文獻(Source Document)이나 혹은 代用物(Surrogates: 書誌, 抄錄, 拔萃 등)로서 判斷하는 情報의 特性이라고 볼 수 있으며 情報의 効用性은 原文獻이 情報要求者の 必要정보를 充足시켜 준 程度 즉, 情報의 實用性 혹은 價値性으로 判斷되는 情報의 特性이라고 볼 수 있다.

여기서 우리는 適合性이나 効用性的 概念自體의 重要性보다는 누구에 의해서 이들이 判定되느냐가 重要な 것임을 알 수 있다.

適合性은 情報專門家나 主題專門家들에 의해 査定되므로 시스템爲主의 判定基準이라고 볼 수 있으며 또한 그 判定이 물론 主觀的이지만 어느 정도 客觀的이라고 볼 수 있다. 反面, 効用性은 情報要求者自身에 의해 査定되므로 利用者爲主의 判定基準이라고 볼 수 있으며 그 判定은 다분히 主觀的이라고 볼 수 있을 것이다.

23) Rees, A. M. "The Relevance of Relevance to the Testing and Evaluation of Document Retrieval Systems," *ASLIB Proceedings*, 18(1966), pp. 316-326.

24) Lancaster, F. W. *Information Retrieval Systems: Characteristics, Testing and Evaluation*, 2nd ed. New York, John Wiley, 1979. pp. 261-262.

또한 情報의 効用性은 특별한 時點에서 利用者의 情報要求를 만족시켜 주는 文獻의 價値이므로 同一한 適合情報가 수록된 文獻이라고 하더라도 처음에 利用한 文獻과 마지막에 利用한 文獻과의 適合情報의 價値는 다를 수 있으며, 같은 文獻이라고 해도 時間의 經過로 인한 차이도 있을 수 있다. 따라서 情報의 効用性決定은 適合性決定보다 훨씬 可變的이고 一時的인 것이라고 볼 수 있다.

이러한 觀點에서 많은 사람들이 檢索效率測定基準으로 情報의 効用性을 適用하는데는 問題가 있다고 보고 檢索效率는 어디까지나 情報要求에 대한 文獻의 適合性으로 評價될 수 밖에 없다고 主張하고 適合性決定에 영향을 미치는 여러가지 要因들을 研究하였다.

SDC(System Development Corporation)社에서 수행된 研究에서 Cuadra와 Katter는 適合性決定이 ① 特定한 査定者들의 技術과 態度, ② 使用되는 文獻, ③ 特定情報要求申請書, ④ 査定을 위한 指導와 道具準備, ⑤ 査定에 適用된 適合性的의 概念과 定義, ⑥ 査定表示에 使用되는 等級範圍나 手段등에 의해 영향을 받는다고 報告하였다.²⁵⁾

WRU(Western Reserve University)에서 수행된 研究에서 Rees는 ① 研究의 段階, ② 文獻, ③ 文獻申請書, ④ 適合性査定者등에 의해 適合性決定이 영향을 받는다고 報告하였다.²⁶⁾

檢索시스템의 궁극적 목표가 利用者指向的이라는 觀點과 또한 利用者의 情報要求는 利用者以外에는 아무도 正確하게 把握할 수 없기 때문에 特定文獻이 情報要求를 만족시켜 줄 수 있는지 없는지는 利用者만이 가장 正確하게 決定할 수 있다는 觀點에서 진정한 意味의 檢索效率는 情報의 適合성과 效率성이 共に 適用되어 評價되어야 한다는 主張이 있다. Kemp, Foskett, Cooper등도 이 點에서는 Lancaster와 同一한 見解라고 본다.

Kemp는 시스템評價의 어떤 目的을 위해서는 適合性決定만으로 充分하나

25) Cuadra, C. A. and Katter, R. V. "The Relevance of Relevance Assessment. *Proceedings of the American Documentation Institute*, 4 (1967), pp.95-99.

26) Rees, A.M. "The Relevance of Relevance to the Testing and Evaluation of Document Retrieval Systems." *op. cit.*

다른 어떤 目的을 위해서는 効用性決定이 必要하다고 하고 適合性決定은 大衆的이고 客觀的(public and objective)인 것이며 効用性決定은 個人的이며 主觀的(private and subjective)인 것이라고 하였다.²⁷⁾

그러나 Kemp의 理論에서 適合性決定이 “客觀的”이라고 한 點에는 異議가 있다고 본다. 물론 適合性이 數名의 査定者에 의해 決定된다는 條件이라면 경우에 따라서는 “어느 程度의 客觀的”이라는 表現은 이해될 수 있다. 그러나 아무런 制限도 없이 適合性決定이 客觀的이라고 한다면 아무리 많은 사람이 多様な 情報要求에 대한 多様な 文獻의 適合性査定을 한다고 해도 適合性決定에 있어 完全한 一致가 이루어져야 할 것이다. 그러나 實際로 이와 같은 完全一致는 매우 어렵다고 본다.

Foskett도 適合性和 効用性에 關係 文獻을 “relevant document”와 “pertinent document”로 區分하여 Lancaster와 비슷한 概念定義를 하고, 언제나 그렇지는 않지만 간혹 “pertinent document”는 또한 “relevant document”이고 “relevant document”는 또한 “pertinent document”라고 하였다.²⁸⁾

Cooper는 “relevance”란 用語 대신 “logical relevance” (혹은 “topicality”), “pertinence” 대신 “utility”란 用語를 使用하여 Lancaster와 비슷한 概念定義를 하였다.²⁹⁾ 그는 또 “利用者”가 보기를 願하는 것은 단지 適合文獻이 아니고 實際로는 効用성이 높은 文獻이다”라고 指摘하고 情報檢索시스템은 “utility”로서 評價되어야 한다고 主張하고 있다.³⁰⁾

情報의 適合성이든 効用성이든 모두 그 概念定義가 數學的으로 分明하게 表現될 수 없으며 또한 그들의 決定이 相對的이며 主觀的일 수 밖에 없다는 데 問題가 있다. 그러나 檢索效率의 測定對象이 情報라는 特性에서 시스템의 特殊한 目的을 위해서는 모두가 使用될 수 있는 基準이라고 생각한다.

27) Kemp, D.A. “Relevance, Pertinence and Information System Development.” *Information Storage and Retrieval*, 10(1974), pp.37-47.

28) Foskett, D.J. “A Note on the Concept of Relevance.” *Information Storage and Retrieval*, 8(1972), pp.77-78.

29) Cooper, W.S. “A Definition of Relevance for Information Retrieval.” *Information Storage and Retrieval*, 7(1971), pp.19-37.

30) Cooper, W.S. “On Selecting a Measure of Retrieval Effectiveness.” *Journal of the American Society for Information Science*, 24(1973), pp.87-100.

다만 現在까지 適合性에 관한 研究와 試驗이 보다 많이 수행되었으며 檢索效率의 測定基準으로 情報의 適合性이 一般的으로 널리 適用되었다는 사실 뿐이다.

한편, 情報시스템 設計者들이 情報의 適合性 判斷能力을 向上시키기 위해 電子計算機를 使用할 수 있는 方法을 研究檢討하여 왔지만 情報의 適合性決定은 시스템이나 利用者를 막론하고 다분히 主觀的일뿐 어떤 數值的이거나 客觀的인 基準이 없기 때문에 별로 成果를 거두지 못하고 있는 實情이다.

물론 Stiles의 適合性算定技法³¹⁾이나 Garfield의 引用文獻索引法에 의한 適合情報의 提示方法³²⁾등 상당수의 適合性 測定技法이 研究開發되었으나 대부분의 檢索시스템이 이와 같은 技法들을 使用하지 않고 오히려 情報要求者에게 適合性的 決定을 맡기고 있으며 가끔 시스템의 性能測定을 위한 수단으로서 情報要求者의 適合性判定을 通報해 줄 것을 要求하고 있다.³³⁾ 이는 情報의 適合性判定은 最終的으로는 利用者의 主觀的 判斷에 의해 決定되기 때문이라고 본다.

適合性決定에 관한 오늘날의 경향은 探索者가 探索을 하는 동안에 適合性 判斷을 보다 容易하게 할 수 있도록 해주는 것이라고 Meadow는 言及하고 있다.³⁴⁾

그러나 이 경우 探索者가 情報要求者 自身이라면 몰라도 시스템에 探索을 委託하는 委任探索의 경우라면 適合性決定은 결국 시스템과 利用者사이의 合意에 의하여 決定될 것이다. 따라서 Lancaster는 “再現率이나 精度率은 결국 시스템의 適合性豫測과 利用者의 適合性判定사이의 一致의 程度를 測定하는 것이라고 하였다.”³⁵⁾

31) Stiles, H. E. "The Association Factor in Information Retrieval." *Journal of the Association for Computing Machinery*, 8(1961), pp. 271-279.

32) Garfield, E. "Citation Indexing, a Natural Science Literature Retrieval System for the Social Sciences." *American Behavioral Scientist*, 5(1964), pp. 58-61.

33) Meadow, C. T. *The Analysis of Information Systems*, 2nd ed. Los Angeles, Melville Pub. Co., 1973. p. 169.

34) *ibid.*

35) Lancaster, F. W. *op cit.* p. 113.

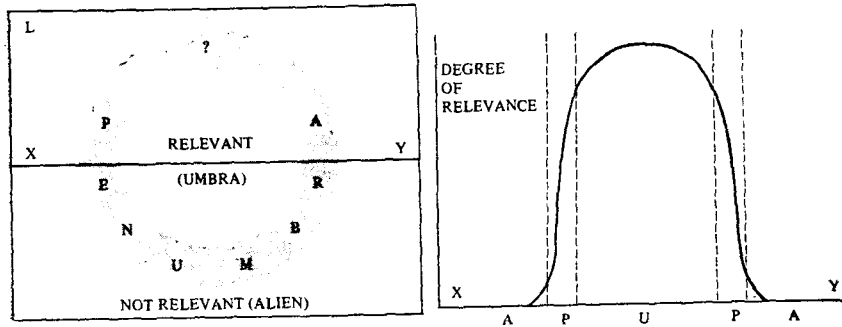
B. 適合性的 程度

利用者の 情報要求에 대해 檢索된 情報를 適合情報과 非適合情報로 兩區分하는데도 問題가 있을 뿐 아니라 또한 適合情報라 해도 어느 程度로 適合하느냐 하는 適合性的 程度는 利用者の 適合性基準에 따라 달라질 것이다.

Lancaster는 適合性を 價値있는 것과 없는 것으로 區分하였으나 King은 適合한 것, 適合하지 않으나 다른 側面에서 有用한 것, 適合하지 않은 것으로 三分하였고, Barber등은 適合한 것, 適合與否를 알 수 없는 것, 適合하지 않으나 다른 側面에서 有用한 것, 適合하지 않은 것으로 四分하고 있다.³⁶⁾

그러나 이들을 綜合하여 分析해 보면 適合한 情報, 適合하지 않은 情報, 그리고 適合할지도 모를 情報로 區分될 수 있어 결국 Ranganathan의 “APU-PA” 理論과 같은 結果가 된다고 Foskett은 指摘하고 있다.³⁷⁾

〈圖表 2〉 Ranganathan의 適合의 程度



圖表 2에서 U는 “umbra” 즉, 분명히 適合한 情報를 말하며, A는 “alien” 즉, 분명히 適合하지 않은 情報를 말하며, P는 “penumbra” 즉, 희미한 부분으로서 분명하게 A라고는 規定할 수 없는 情報라고 하였다.

36) Barber, A.S. et al. “On-line Information Retrieval as a Scientific Tool.” *Information Storage and Retrieval*, 9(1973), p.435.

37) Foskett, A.C. *The Subject Approach to Information*, 3rd ed. Lond on, Clive Bingley, 1977. p.15.

그렇다면 P는 A라고는 確定될 수 없기 때문에 결국 “適合할지도 모를” 情報가 되는 것으로 實際의 情報活動에서 흔히 存在할 수 있는 것이다. 결국 P는 경우에 따라서는 U가 될 수 있기 때문에 蓄積情報에 P가 存在하는 限 絕對再現率이나 絕對精度率は 測定될 수 없는 것이다. 다만 利用者の 適合性基準에 따라 P의 適合性與否가 決定됨으로써 再現率이나 精度率에 測定되는 것이다.

4. 再現率과 精度率

檢索效率의 傳統的인 評價基準으로 가장 널리 使用되고 있는 것이 再現率과 精度率이다.

Atherton은 “再現은 利用者가 願하는 情報를 檢索하는 能力이고 精度는 利用者가 願하지 않은 情報의 檢索을 避하는 能力”이라고 하였으며,³⁸⁾ Lancaster는 이들을 시스템의 濾過裝置의 能力으로 表現하여 “再現率は 適合한 情報들을 추려내는 濾過裝置의 能力測定이고 精度率は 適合하지 않은 情報(雜音)를 걸러내는 濾過裝置의 能力測定”이라고 하였다.³⁹⁾

따라서 再現은 檢索된 情報 속에 利用者の 要求를 만족시켜 줄 수 있는 適合情報가 얼마나 많이 들어 있는지를 나타내며, 精度는 檢索된 情報속에 利用者の 要求에 適合하지 않은 情報 즉, 雜音(Noise)이 얼마나 많이 들어 있는지를 나타내는 것으로 이들을 百分率로 表示한 것이 再現率과 精度率이다.

Meadow는 檢索效率의 測定은 原則적으로 하나의 假說을 試驗하는 것으로서 모든 假說의 試驗에는 誤差가 發生할 수 있으며 그 發生要因으로 索引作成의 誤差, 質問作成의 誤差(利用者の 情報要求 表現上의 誤差), 質問이나 索引作成에 있어서 構文制限(Syntax Restrictions) 또는 語彙制限(Vocabulary Restrictions)등을 指摘하고 있다.⁴⁰⁾ 따라서 情報檢索시스템이 100%의 再

38) Atherton, P. *Handbook for Information Systems and Services*. Paris, Unesco, 1977. p.75.

39) Lancaster, F.W. *op. cit.* p.112.

40) Meadow, C.T. *op. cit.* p.169.

現率과 100%의 精度率을 얻기란 실질적으로 거의 不可能하다고 지적하고 Hoel의 統計學用語⁴¹⁾를 引用하여 다음과 같이 定義하였다.

第1種의 誤差: 假說이 맞는 것이 기각되는 것으로서 간혹 “Blank Sort”라고 불리운다. 이는 適合情報가 檢索되지 못한 경우이므로 第1種의 誤差는 再現率을 低下시킨다.

第2種의 誤差: 假說이 맞지 않는 것이 채택되는 것으로서 “False Drop”이라고도 알려져 있다. 이는 適合情報가 아닌 雜音이 檢索된 경우이므로 第2種의 誤差는 精度率을 低下시킨다.⁴²⁾

모든 檢索시스템이 願하는 것은 100%의 再現率과 100%의 精度率일 것이다. 완전무결한 探索이라면 蓄積情報중에서 모든 適合情報가 檢索될 수 있을 것이다. 이 경우는 圖表 1에서 $b=0$, $c=0$ 인 경우로서 시스템의 再現率과 精度率이 모두 100%이고 漏落率과 雜音率이 모두 0%이다. 다시 말해 시스템의 適合性豫測과 利用者의 適合性判定사이에 完全한 一致가 이루어진 경우이다.

그러나 이는 Meadow가 誤差의 發生可能性으로指摘한 것처럼 實際的인 情報檢索活動에서는 극히 드문 일이며 매우 어려운 일이라고 본다.

實際의 檢索活動에서는 圖表 1에서 보는 바와 같이 全適合情報와 $(a+c)$ 全檢索情報($a+b$) 사이에 部分的 一致가 이루어지는 경우가 거의 대부분이라고 볼 수 있다.

훌륭한 檢索시스템일수록 대부분의 適合情報(most, but not all)를 檢索하는 反面에 대부분의 非適合情報(most, but not all)를 檢索하지 않을 것이다. 이와 같은 두가지의 “MOBNA”⁴³⁾현상의 크기가 바로 再現率과 精度率인 것이다.

Cranfield Institute of Technology에서 創案한 파시트색인(Facet Index)과 WRU(Western Reserve University)의 金屬工學關係文獻 索引의 比較檢討

41) Hoel, P. G. *Introduction to Mathematical Statistics*, 4th ed. New York, John Wiley, 1971. p. 108.

42) Meadow, C. T. *op. cit.* pp. 165—168.

43) 筆者의 命名임.

를 內容으로 한 WRU 試驗⁴⁴⁾은 再現率과 精度率間의 反比例關係(Inverse Relationship)를 論證한 첫 試圖로서 그 後 많은 試驗과 研究에서 再確認되었다.⁴⁵⁾

Ziman은 이러한 再現과 精度間의 逆關係를 檢索의 “基本”法則(“Fundamental” Law of Retrieval)이라고 命名하였으며⁴⁶⁾ Cleverdon은 1972년에 그의 理論을 하나의 法則으로서 재정립하였다.⁴⁷⁾

Cleverdon이 재정립한 法則敘述에 대해 Brookes, Farradane, Moss등의 약간의 異義가 있었으나 Robertson은 “Cleverdon의 法則敘述은 合理的으로 解析될 수 있으며, 비록 反比例關係가 法則으로서 수락될 수 없을지 몰라도 그것은 分明히 檢索效率評價의 道具로서 使用되어야 한다. 만일 評價試驗에서 이 逆關係가 위반되면 그 試驗結果는 심히 의심스러울 수 밖에 없다”고 言及하고 있다.⁴⁸⁾ 또한 Van Rijsbergen도 이를 “檢索의 第1法則”(First Law of Retrieval)이라고 말하고 있다.⁴⁹⁾

이와 같은 再現率과 精度率과의 相反關係를 그래프로 表示한 것을 性能曲線(Performance Curve)이라고 하는데 이 性能曲線의 一般의인 型은 圖表 3과 같다.

再現率과 精度率 間의 性能曲線은 主題分野의 言語特性이나 利用者에 의한 情報의 適合性基準에 따라 差異가 있게 된다.

主題나 概念이 比較的 明確하게 記述될 수 있는 自然科學分野는 性能曲線이 圖表 3의 Y曲線과 같이 오른쪽으로 기울게 되며 抽象的인 記述이 많은 社會科學分野는 圖表 3의 X曲線처럼 왼쪽으로 기울게 된다. 또한 SDI

44) Aichison, J. and Cleverdon, C. W. *Report on a Test of the Index of Metallurgical Literature of Western Reserve University*, Cranfield, 1963.

45) Cleverdon, C. W. “Progress in Documentation :...” *op. cit.*

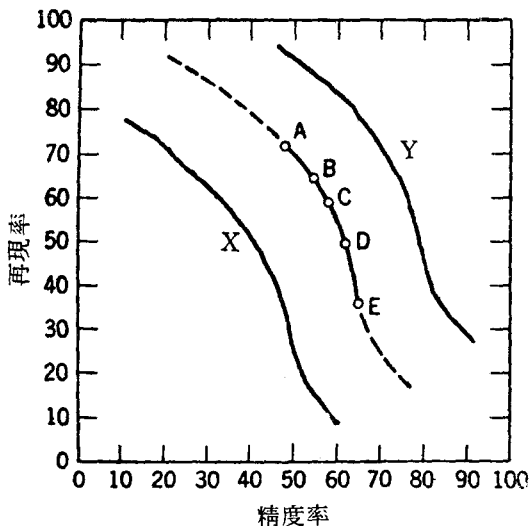
46) Ziman, J. H. “Information, Communication, Knowledge.” *Nature*, 224(1975), pp. 318-324.

47) Cleverdon, C. W. “On the Inverse Relationship of Recall and Precision.” *Journal of Documentation*, 28(1972), pp. 195-201.

48) Robertson, S. E. “Explicit and Implicit Variables in Information Retrieval (IR) Systems.” *Journal of the American Society for Information Science*, 26(1975), pp. 214-222.

49) Van Rijsbergen, C. J. *op. cit.* p. 114.

〈圖表 3〉 再現率과 精度率과의 性能曲線



(Selective Dissemination of Information) 서비스의 경우에는 情報의 適合性基準이 RS(Retrospective Search)서비스의 경우보다 낮게 되어 SDI시스템의 性能曲線은 오른쪽으로, RS시스템의 性能曲線은 왼쪽으로 기울게 된다.⁵⁰⁾

한편, 檢索된 情報가 再現率이 높을 때는(性能曲線上의 A點) 精度率은 反對로 낮고, 精度率이 높을 때는(性能曲線上의 E點) 再現率은 反對로 낮음을 볼 수 있는데 이러한 性能曲線의 反比例的 關係는 언제나 不變하는 絶對的인 것은 아니다. 이것은 어떠한 檢索시스템을 막론하고 시스템의 平均值的 性能曲線을 나타내는 것으로 경우에 따라서는 상당히 變化될 수 있는 것이다.

예를들어 Lancaster의 MEDLARS 評價試驗에서는 시스템의 平均再現率이 58%, 平均精度率이 50%이며, 探索件當 平均 175件的 文獻을 探索하는 것으로 나타났으며, 再現率을 85% 내지 95%로 올리기 위해서는 探索件當 平均 약 500내지 600개의 文獻을 檢索하여야 하며, 이와 같은 경우 精度率은 相對的으로 약 20%정도로 低下된다고 報告하고 있다. 또한 文獻探索을 個別的으로 調査해 보면 어느 경우에는 100%의 再現率과 100%의 精度率을 얻은 結果도 있고 어느 경우에는 再現率과 精度率이 모두 0%인 結果도 있다고 報告하고 있다.⁵¹⁾

이와 같이 特殊한 경우에는 再現率과 精度率과의 相反關係는 이루어질 수

50) 司空哲, *op. cit.* p.164.

51) Foskett, A. C. *op. cit.* o.17 and pp.453-454.

없으나 一般적으로 시스템의 平均值的 性能曲線은 이들 兩者間에 대체로 反比例關係가 이루어 진다고 볼 수 있다.

그러나 再現率이나 精度率의 어느 쪽도 그 自體로서 檢索效率을 完全하게 表示할 수는 없다. 100%의 再現率은 檢索方針, 檢索時間, 檢索費用에 關係없이 모든 蓄積情報를 하나씩 차례로 點檢한다면 언제나 可能하다고 볼 수도 있다. 그러나 이와같은 경우 一般적으로 精度率은 매우 낮게 되어 높은 再現率이 별로 중요한 意義를 갖지 못한다. 特히 情報의 生産量이 比較的 적었던 過去에는 再現率이 높은 檢索을 願하였을지 모르나 오늘날과 같은 情報의 洪水時代에는 情報利用者の 立場에서 볼 때 精度率이 높은 量의 情報가 더 바람직하다고 볼 수 있을 것이다. 왜냐하면 精度率은 利用者の 時間과 努力에 直接的인 影響을 주기 때문이다.

어떤 特定蓄積情報안에 利用者の 特定情報要求를 만족시켜 줄 수 있는 適合情報가 20개 들어 있다고 가정한다. 探索戰略등이 다른 A,B,C의 3方法으로 探索한 結果 모두가 20개의 適合情報중에서 各各 15개씩을 檢索하였는데 (따라서 A,B,C의 再現率은 各各 75%), A가 檢索한 全體情報는 30개, B는 60개, 그리고 C는 150개라고 한다면 精度率은 A가 50%, B가 25%, C가 10%로서, 만일 다른 條件들이 同一하다고 가정하면 A는 30개의, B는 60개의, C는 150개의 檢索情報중에서 15개씩의 適合情報를 區別해 내야 하기 때문에 결국 B는 A보다, 그리고 C는 A와 B보다 時間과 努力을 더 要求하게 될 것이다. 結果적으로는 A는 B와 C보다 그리고 B는 C보다 檢索效率이 높은 것이다.

한편, 앞에서 이미 指摘한 바와 같이 再現率과 精度率에 의한 檢索效率의 評價가 利用者指向의이라는 觀點에서 본다면 다음과 같은 測定基準도 關心事가 될 수 있을 것이다.

1) 相對再現率(Relative Recall Ratio)

情報檢索시스템에 대해 100%의 再現率을 要求하는 利用자가 있을 수도 있지만 앞에서 살펴본 바와 같이 대개의 경우 利用자는 適合情報를 區別하기 위해 時間과 努力이 더 要求되는 높은 再現率보다는 時間과 努力이 덜

要求되는 높은 精度率에 만족한다고 볼 수 있을 것이다. 따라서 어떤 利用者는 경우에 따라 시스템으로 부터 提供받고자 하는 情報의 量을 限定할 수도 있는 것이다.

이와 같이 利用者들의 限定된 要求에 대한 시스템의 性能을 測定하기 위한 手段으로 相對再現率(Relative Recall Ratio 또는 Proportional Recall Ratio)을 使用하는데 그 測定公式은 다음과 같다.

$$\text{相對再現率} = \frac{\text{檢索된 適合情報}}{\text{利用자가 願하는 適合情報}} \times 100\%$$

예를들어 利用자가 願하는 適合情報가 5件이라고 하고 시스템이 5件의 適合情報를 檢索하였다면 檢索되지 않은 適合情報가 아무리 많이 있다 하여도 相對再現率은 100%이다. 그러나 시스템이 3件의 適合情報를 檢索하였다면 相對再現率은 60%가 되는 것이다.

또한 相對精度率도 생각할 수 있으나 그 測定公式은 相對再現率과 同一하게 될 것이므로 결국 相對再現率이 될 것이다.

2) 新情報率(Novelty Ratio)

檢索시스템이 利用者에게 提供한 檢索情報중에서, 또는 檢索된 適合情報중에서 利用자가 過去에 알지 못했던 새로운 適合情報가 얼마나 되는가를 나타내는 比率로서 아래와 같이 두가지 測定公式이 使用될 수 있다.

$$\frac{\text{檢索된 適合情報중 새로운 適合情報量}}{\text{檢索된 適合情報量}} \times 100\%$$

또는

$$\frac{\text{檢索된 適合情報중 새로운 適合情報量}}{\text{檢索된 情報量}} \times 100\%$$

이와 같은 新情報率은 最新情報周知서비스(Current Awareness Service)인 SDI서비스를 위한 情報檢索에 특별히 有效하다고 볼 수 있다.

3) 情報範圍率(Coverage Ratio)

利用자의 質問에 대해 探索코자 하는 特定主題에 관한 情報를 시스템의 特定데이터베이스가 얼마나 포함하고 있는가를 나타내는 比率로서 特定分野의 網羅的 索引誌나 리뷰誌를 통해서 測定될 수 있다.

어떤 데이터베이스의 特定主題에 관한 情報範圍率は 檢索시스템의 再現率에 直接的인 영향을 미친다. 이에 관해서는 檢索效率의 支配要因에서 說明 하겠다.

5. 檢索效率의 不確實性

情報의 適合性決定에 어떠한 方法이 適用된다 하여도 適合성에 의한 檢索效率의 測定은 不確實性(Uncertainty)을 排除할 수 없다고 본다.

不確實性的 要因을 살펴보면 첫째, 情報檢索시스템이 情報要求者에게 一次的으로 提供하는 書誌事項이나 抄錄 혹은 拔萃등은 原文獻의 代用物에 不適當한 것이다.

利用者는 이와 같은 代用物에 의하여 適合性 判斷을 第一次的으로 決定한 後 原文獻을 얻어 第二次的으로 最終的인 適合性判定을 하는 것이다. 따라서 代用物에 의한 適合性判斷과 原文獻에 의한 適合性判定은 반드시 一致한 다고는 볼 수 없을 것이다.

Belzer가 情報理論을 應用하여 여러가지 文獻代用物이 갖는 情報量을 測定한 試驗은 代用物에 의한 適合性判定에 參考할 價値가 있다고 본다.

이 試驗은 미국 Pittsburgh大學校 醫科大學에서 醫學分野研究者들로 구성된 利用者 그룹을 對象으로 다섯가지의 文獻代用物 (① 書誌事項, ② 抄錄, ③ 文獻의 첫째 파라그라프, ④ 文獻의 마지막 파라그라프, ⑤ 文獻의 첫째 및 마지막 파라그라프)이 原文獻이 갖고 있는 情報를 얼마나 傳達할 수 있는가를 測定한 試驗인데 그 結果는 原文獻이 갖는 情報量을 1 bit로 했을 경우 書誌事項은 平均 0.0953, 抄錄은 0.1233, 첫째 파라그라프는 0.1603, 마지막 파라그라프는 0.1659, 그리고 첫째 및 마지막 파라그라프는 0.313, bit의 情報를 傳達한다는 것이다.⁵²⁾

檢索시스템이 一次的으로 提供하는 書誌事項이나 抄錄같은 文獻代用物의 情報傳達量이 아주 적은 點으로 미루어 보아 이들로부터 一次的으로 判斷해

52) Belzer, J. "Information Theory as a Measure of Information Content," *Journal of the American Society for Information Science*, 24(1973), pp. 300-304.

야 하는 情報의 適合性決定은 그 不確實性이 排除될 수는 없을 것이다.

둘째, 再現率의 測定에 있어서 不確實性은 精度率보다 더욱 심각하다고 본다. 왜냐하면 精度率의 測定은 檢索된 情報(出力情報)에 대한 利用者의 適合性決定으로 可能하지만 再現率의 測定은 檢索되지 않은 情報 즉, 入力된 全體情報에서 檢索된 情報를 빼고 남은 情報의 適合性決定이 先行되자 않고는 不可能하기 때문이다.

圖表 1에서 全體情報(a+b+c+d)가 500,000文獻, 檢索된 情報(a+b)가 25文獻, 檢索情報중 適合情報(a)가 15文獻이라고 假定하면 우선 圖表 4와 같은 結果를 얻을 것이다.

<圖表 4>

	適合情報	非適合情報	全體情報
檢索된 情報	15	10	25
檢索되지 않은 情報			499,975
全體情報			500,000

따라서 精度率은 $\frac{10}{25} \times 100 = 60\%$ 가 된다. 그러나 再現率은 檢索되지 않은 適合情報(c)를 알지 못하고는 求할 수 없을 것이다. 그러나 c를 알기 위해 檢索되지 않은 499,975文獻을 하나 하나 檢索해 본다는 것은 매우 어려운 일이며 또 時間的으로나 經濟的으로나 非現實的일 것이다. 물론 全體情報가 50내지 100文獻 정도라면 可能할 수도 있지만 거의 모든 分野의 데이터베이스의 크기로 보나 또는 利用者의 情報要求의 特殊性등으로 보아 대부분의 경우 檢索情報나 適合情報는 全體情報에 비해 아주 적다고 본다. 그러므로 正確한 再現率의 測定은 全體의 適合情報(a+c)가 미리 把握되어 決定될 수 있는 實驗的 경우를 除外하고는 實際的인 情報檢索시스템에서는 거의 不可能할 것이다. 따라서 가장 近似值가 될 수 있는 再現率 즉, 概算再現率(近似再現率: Estimation of Recall Ratio)을 測定할 수 있는 方法이 模索될 수 밖에 없는 것이다.

概算再現率을 測定하기 위한 몇가지 方法이 文獻에 소개되었다.⁵³⁾ 여기서

53) Lancaster, F.W. *op. cit.* pp.129-132.

는 MEDLARS 評價試驗에 使用된 技法을 살펴보기로 한다.

A 라는 시스템의 概算再現率을 測定하기 위해 S 라는 主題에 관한 文獻探索을 실시한 結果 R 이라는 檢索結果가 出力되었다고 가정한다. 다시 S 라는 主題에 관한 文獻探索을 B 라는 檢索시스템에서 실시한 結果 14개의 文獻이 檢索되어 이들에 대한 適合性判定을 情報要求者에게 의뢰한 結果 12개가 適合文獻으로 判定되었다. 이 12개의 適合文獻이 A 라는 시스템이 探索한 全體情報(Database) 중에 몇개나 포함되어 있는지를 調査한 結果 10개가 포함되어 있었다고 가정한다. 이 10개의 文獻중 몇개가 A 라는 시스템이 檢索한 R 속에 포함되어 있는지 調査한 結果 7개가 있고 3개는 없었다면 A 라는 檢索시스템은 10개의 適合文獻 중에서 7개만 檢索하고 3개는 檢索하지 못하였으므로 이 시스템의 概算再現率을 $\frac{7}{10} \times 100 = 70\%$ 로 定하는 것이다.

圖表 4로 돌아가 이 시스템의 概算再現率이 위와 같은 方法으로 70%로 測定되었다면 이제는 圖表 5와 같은 結果를 얻을 수 있을 것이다.

<圖表 5>

	適合情報	非適合情報	全體情報
檢索된 情報	15	10	26
檢索되지 않은 情報	6*	499,966*	499,975
全體情報	21*	499,979*	500,000

(適合情報 15=70%이므로 100%의 適合情報은 약 21이 될 것이다. 따라서 별표(※)의 數值를 얻을 수 있을 것이다.

이와 같이 適合性에 의한 再現率測定은 絕對再現率(Absolute or True Recall Ratio)이 아닌 概算再現率의 測定일 수 밖에 없는 것이며, 이와 같은 見地에서 볼 때 不確實性은 더욱 排除될 수 없다.

셋째, 시스템이 提供하는 文獻代用物이든 原文獻이든 利用者는 어떤 順序로서 文獻의 適合性을 決定할 것이다. 따라서 제일 먼저 査定한 文獻과 마지막에 査定한 文獻의 適合性判定은 달라질 수 있을 것이다.

文獻의 개별적 適合性判定은 먼저 調査된 文獻의 適合性與否에 따라 영향을 받을 수 있을 것이다. 利用者의 情報要求에 틀림없이 適合한 情報인데

만일 뒤에 査定되었다면 앞에서 調査되어 얻어진 適合情報 때문에 별로 必要없는 情報가 될 수 있어 適合性判定에 있어서 非適合情報로 決定될 수 있을 것이다. 이와 같은 時差때문에 適合性判定의 可變性이나 一時性은 排除될 수 없을 것이다.

6. 檢索効率의 支配要因

情報檢索效率의 價値는 窮極的으로 利用者에 의해서 決定될 것이다. 아무리 훌륭한 시스템이 提供하는 情報라도 知識이나 經驗이나 情報要求의 動機(Motivation)등이 다른 利用者에 따라 情報의 價値는 달라질 것이다. 따라서 再現率이나 精度率의 價値도 多様な 情報利用者の 多様な 情報要求에 따라 달라질 것이다.

冊을 쓴다든지, 또는 批評記事(Review Article)를 準備한다든지, 또는 長期研究 프로젝트를 着手하는 사람이라면 처음에는 흔히 綜合的인 文獻探索을 必要로 할 것이므로 關聯分野의 網羅의 檢索 즉, 再現率의 높은 檢索을 원할 것이다. 反面에 製品生産이나 品質改善을 위한 特殊問題解決, 論文記事를 위한 特殊情報, 또는 短期研究프로젝트 등을 수행하는 사람에게서는 問題解決을 위한 特殊主題에 관한 正確한 少數의 最新情報 즉, 精度率의 높은 檢索이 더 重要할 것이다. 또한 SDI서비스를 위해서는 대체로 再現率의 높은 檢索이 重要할 것이며 RS를 위해서는 대체로 精度率의 높은 檢索이 重要할 것이라고 생각한다.

한편, 再現率의 價値는 모든 適合情報가 대략 同等한 價値를 갖는다는 前提條件 아래에서 成立될 수 있다. 그러나 앞에서 살펴본 바와 같이 情報의 適合性에는 差異가 있게 마련이다.

利用者の 特定情報要求에 대해 시스템의 探索結果가 다섯개의 適合정보를 檢索하고 10개의 適合정보를 漏落하였는데 (再現率은 33%) 다섯개의 檢索된 適合정보가 10개의 漏落된 適合정보보다 훨씬 價値가 많은 情報라면 漏落된 10개의 適合정보는 剩餘정보가 되어 결국 그 價値는 低下될 것이다. 이와 같은 경우 再現率에 의한 檢索效率은 실질적으로는 33%보다 더 높게 評價

되어야 할 것이다.

또한 精度率은 이미 살펴본 바와 같이 실질적으로는 利用者의 時間과 努力에 많은 영향을 주기 때문에 精度率은 一面 利用者의 時間과 努力의 間接的測定이라고 볼 수 있을 것이다.

이와 같은 精度率은 어디까지나 利用者가 情報檢索을 시스템에 의뢰하는 委託探索(Delegated Search)의 性能測定에 보다 有効한 것이다.

精度率이 매우 낮은 檢索結果(대부분의 경우 書誌事項이나 抄錄 또는 拔萃 등 소위 代用物的 情報만 포함하고 있기 때문에)로부터 第一次的 適合性 判斷을 決定한 후 第二次的으로 原文獻등을 入手하여 最終的인 適合情報를 얻기 위한 利用者의 時間과 努力과 經費등은 상당할 것이다.

그러나 委託探索이 아니고 利用者 스스로가 수행하는 非委託探索(Nondelegated Search)의 경우에는 이와 같은 時間이나 努力이나 經費등이 委託探索의 경우와는 상당히 다른 것이다.

索引用語를 印刷한 索引集이나 혹은 온-라인 시스템에서 對照해 가며 探索을 수행하는 경우에는 利用者 스스로가 適合性判斷을 同時에 決定할 수 있기 때문에 精度率은 委託探索의 경우와는 다르게 될 것이다. 이 경우의 精度率은 利用者가 適合하다고 判定한 文獻數를 利用者가 실제로 對照한 文獻數(따라서 檢索된 全體文獻數가 아님)로 나눈 것이기 때문에 委託探索의 경우와는 그 값이 달라지게 된다.

따라서 精度率은 委託探索의 性能測定에는 妥當性이 있다고 보겠으나 非委託探索에는 별로 意義가 없다고 보겠다. 그러나 情報檢索은 대부분의 경우 시스템에 의뢰하는 委託探索으로 수행된다.

7. 檢索効物の 支配要因

情報檢索시스템의 檢索効率을 支配하는 主要要因들을 살펴보면 다음과 같다.

- 探索情報の 範圍
- 索引시스템

- 利用者の 情報要求
- 探索戰略

A. 探索情報の 範圍

利用者の 情報要求에 대해 探索하고자 하는 蓄積情報の 範圍는 再現率을 支配하는 重要な 要因이 된다. 즉, 探索하고자 하는 主題에 關係 시스템의 파일속에 포함되어 있는 情報の 範圍(文獻의 數)는 곧 再現率의 높고 낮음에 영향을 미칠 것이다.

利用者로 부터 再現率 65%에 精度率 25%의 要求가 있다고 하자. 시스템이 要求된 特定主題에 關係한 總情報量의 95%를 축적하고 있다면 68%의 再現率만으로도 要求條件인 65%의 再現率을 만족시킬 수 있지만 蓄積情報量이 75%라면 同一한 結果를 얻기 위해서는 87%의 再現率을 必要로 한다.

이와 같이 要求된 特定主題에 關係한 蓄積情報量이 많으면 높은 再現率을 얻을 수 있고 反對로 적으면 낮은 再現率을 얻게 될 것이다.

시스템이 探索할 特定데이터베이스內에 入力되어 있는 特定主題의 情報量을 測定하기란 容易한 일이 아니다. Martyn과 Slater에 의해 研究되어⁵⁴⁾ 一般적으로 使用되고 있는 하나의 測定技法은 데이터베이스에 수록된 主題分野의 專門書誌資料들을 利用하여 算出하는 方法이다. 正確한 情報量의 算出은 매우 어려우며 概略적인 算出方法에 不過하다.

主題分野의 綜合索引誌(例: Index Medicus)나 抄錄誌(例: Chemical Abstracts) 혹은 리뷰誌(例: Annual Review of Information Science and Technology)나 리뷰記事등은 特定主題의 情報量을 調査하는 좋은 書誌資料가 된다.

B. 索引시스템

再現率과 精度率을 支配하는 가장 重要な 要因으로 索引시스템의 網羅性(Exhaustivity)과 特定性(Specificity)이 있는데 網羅性은 再現率을, 그리고 特定性은 精度率을 支配하는 要因이 된다.

54) Martyn, J. and Slater, M. "Tests on Abstracts Journals." *Journal of Documentation*, 20(1964), pp. 212-235.

網羅性은 索引作成時 文獻當 索引用語(Index Term) 혹은 索引語彙(Index Vocabulary)의 數를 나타내는 것이고 特定性은 主題를 正確하게 記述하는 索引言語(Index Language)의 能力을 말하는 것이다.⁵⁵⁾

따라서 網羅性은 索引作成時 索引語彙를 얼마나 확장해서 使用할 것이냐 하는 시스템의 索引作成의 政策決定이지만 特定性은 索引言語 自體의 本質의 質을 나타내는 시스템의 機能으로서 特定索引言語(分類表, 主題名標目表, 디소오러스 등)를 채택할 때 自動的으로 索引言語의 特定性은 決定되는 것이다.

예를 들어 “Index Medicus”에서 文獻當 平均 6개의 索引用語로 索引되고 “MEDLARS”에서는 文獻當 平均 10개의 索引用語로 索引된다고 假定하면 이 두 시스템의 索引作成의 網羅性은 3:5가 되는 것이다. 만일 많은 數의 索引語彙를 갖고 있는 索引言語가 主題를 正確하게 記述할 수 있는 主題表現의 能力이 높다고 假定한다면 “MEDLARS” 시스템의 索引言語가 “Index Medicus”시스템의 索引言語보다 索引言語의 特定性이 높은 것이다.

먼저 索引作成의 網羅性을 살펴본다. 4개의 主題(A, B, C, D)가 포함된 文獻이 있다고 가정한다. 索引作成時 4개의 主題가 모두 認識(把握)되어 索引이 되었다면 (網羅性이 높은 索引) 4개의 主題중 어떠한 主題로 探索한다 하여도 이 文獻은 索引될 것이다. 즉, 檢索確率이 많아 이 文獻의 再現率은 높을 것이다. 그러나 索引作成時 D主題가 認識되지 않아 索引되지 않았다면 D主題에 관한 情報要求는 이 文獻을 探索하지 못할 것이다. 따라서 이 文獻의 再現率은 그만큼 낮아지는 것이다. 이와 같이 索引作成이 網羅性은 再現率을 상승시키는 要因이 된다.

그러나 이와는 反對로 索引作成의 網羅性이 높으면 精度率은 低下된다. 가령, A主題를 探索하고자 할 때 檢索된 이 文獻은 A와는 별로 관계없는 B, C, D라는 主題도 포함하고 있기 때문에 결국 이 文獻속에는 A主題 以外의 雜音이 많아 精度率은 低下되는 것이다.

55) Guha, B. *Documentation and Information*. Calcutta, The World Press Private Ltd., 1978. p.205.

그러나 前述한 WRU의 索引實驗에서 나타난 結果를 보면 索引作成의 網羅性에도 어떤 最適度가 있다는 것이다. 즉, 網羅性이 낮은 경우에는 높은 再現率을 期待할 수 없지만 網羅性이 어느 程度以上 높아지고 나면 그것을 더 높인다 해도 再現率이 별로 높아지는 것이 아니고 오히려 精度率만 低下시키는 結果가 된다고 Cleverdon 등은 報告하고 있다.⁵⁶⁾

다음으로 索引言語의 特定性を 살펴본다. 시스템이 채택한 索引言語가 주어진 特定主題를 詳述할 수 있는 索引用語가 없다면 그 主題는 보다 一般的인 用語로 記述될 것이다. 가령 主題를 正確히 記述할 수 있는 用語는 “진도개”인데 이 用語가 시스템이 채택한 索引言語에서는 使用되지 않았다면 이 主題는 보다 包括的인 用語인 “개”로 索引되었을 것이며 結果적으로 “진도개”에 관한 情報要求가 있다고 해도 “개”에 관한 文獻을 探索할 수 밖에 없는 것이다. 따라서 이 경우 檢索된 文獻은 “진도개”만을 전적으로 취급한 文獻, “진도개”가 部分的으로 취급된 文獻과 “진도개”는 전연 취급되지 않은 文獻등이 모두 檢索될 것이므로 檢索效率의 精度率은 “진도개”라는 索引用語가 채택된 索引言語보다 분명히 낮을 것이다. 이와 같이 索引言語의 特定성은 精度率을 높여주고 再現率을 減少시키는 支配要因이 되는 것이다.

오늘날 여러 分野의 디소오러스가 作成되어 活用되고 있는 것은 이와 같은 索引言語의 效率성을 높여 檢索效率을 增大시키고자 하는데에 그 目的이 있다.

C. 利用者の 情報要求

人間の 言語表現은 各樣各色이어서 表現方法에 따라 利用者の 情報要求에 대한 把握은 달라질 수 있는 것이다.

正確하게 表現된 情報要求는 檢索效率을 높일 수 있는 要因이 된다. “探索者가 利用者가 무엇을 要求하고 있는지를 正確히 把握하면, 그 探索의 半은 이미 成功한 것”이라고 Collison은 指摘하고 있다.⁵⁷⁾

56) Cleverdon, C. W., Lancaster, F. W. and Mills, J. “Uncovering Some Facts of Life in Information Retrieval.” *Special Libraries*, 55(1964), pp. 86-91.

57) Collison, R. L. *Library Assistance to Readers*, 5th ed. London, Crosby Lockwood, 1965. p. 62.

그러나 利用者가 自己의 情報要求事項을 처음부터 正確히 記述하여 시스템에 提出하는 경우는 매우 드물다고 본다.

利用者の 情報要求를 위한 質問形成의 過程은 보통 心理的인 것에서 論理的인 것으로, 모호한 것에서 明確한 것으로, 복잡한 것에서 단순한 것으로 變化하기 때문에⁵⁸⁾ 委託探索의 경우 대개 利用者와 시스템(探索者)間에 여러가지 方法(書信, 電話, 面談등)에 의한 커뮤니케이션으로 利用者の 情報要求는 보다 자세하고 精確하게 把握되는 것이다.

探索者가 利用者の 情報要求를 잘 알면 알수록 높은 수준의 適合性決定을 하게 되어 결국 精度率이 높은 檢索을 할 수 있으나 反對로 情報要求를 잘 모르면 모를수록 낮은 수준의 適合性決定으로 인하여 많은 文獻을 適合하다고 判斷하게 되어 결국 精度率이 낮은 檢索을 하게 될 것이다. 이는 利用者の 特定情報要求에 대한 探索者の 專門的 知識의 程度가 適合性判斷能力을 좌우하게 된다는 事實로서 結果的으로 利用者の 情報要求에 대한 探索者の 專門知識程度와 檢索된 適合文獻數와는 反比例關係가 成立된다고 볼 수 있다.

D. 探索戰略

再現率과 精度率의 性能曲線은 探索戰略에 따라 상당히 調節될 수 있다. 探索의 範圍를 넓게하여 包括的인 探索을 하면 다소라도 關聯이 있는 많은 文獻이 檢索되어 再現率이 높게 될 것이고, 反對로 探索의 範圍를 좁게하여 세밀한 探索을 하면 文獻量은 적더라도 主題와 關聯이 많은 文獻만 檢索되어 결국 精度率은 높은 結果를 얻을 수 있는 것이다.

利用者の 情報要求를 精確하게 把握하여 “부울”演算子(Boolean Operators)인 AND, OR, NOT 등의 集合理論을 適用하여 利用者の 要求條件에 부응하도록 再現率과 精度率을 調整하는 것은 檢索效率을 向上시키는 探索戰略인 것이다.

그러나 이와 같은 探索戰略에 의한 調整은 어느 適正線까지가 重要하다고 본다. 이미 MEDLARS 評價研究에서 指摘된 것처럼 再現率이나 精度率의

58) 司空哲, op. cit. p. 149

어느 한쪽을 위한 다른 한쪽의 지나친 희생은 投入된 時間이나 努力이나 經費면에서 볼 때 特殊한 경우를 除外하고는 별로 큰 成果를 期待할 수 없는 것이다.

Shaw의 “Milkweed”에 관한 文獻調査는 이 點을 뒷받침해 주는 좋은 實例라고 본다. 이 調査에 의하면 “總 4천개의 文獻을 調査해 본 結果, 얻을 수 있는 모든 情報는 96개의 文獻속에 모두 포함되어 있었다는 것이다.”⁵⁹⁾ 그렇다면 3,904개의 文獻속에 들어있는 Milkweed에 관한 情報는 결국 96개의 文獻속에 포함되어 있는 情報의 重複이라는 結論이다.

따라서 利用者의 情報要求의 類型에 따른 再現率과 精度率의 調整도 어느 수준까지의 適正線 維持가 바람직하다고 본다.

8. 結 論

情報檢索시스템評價의 第一의 基準이 되는 檢索效率의 評價基準 및 測定方法, 情報의 適合性 및 效用性, 檢索效率의 價値性 및 支配要因등에 關係考察해 보았다. 그 要點을 간추려 보면 다음과 같다.

1) 檢索效率에 의한 情報檢索시스템評價는 情報의 適合性이라는 評價尺度가 多분히 相對的이며 主觀的이기 때문에 많은 研究와 試驗의 對象이 되고 있다.

2) 檢索效率測定の 여러가지 基準이 研究되었으나 利用者指向의인 再現率과 精度率에 의한 檢索效率測定技法이 가장 널리 適用되고 있다. 그러나 必要에 따라서는 시스템 指向의인 再現率과 非適合率에 의한 檢索效率評價도 併用되어야 한다.

3) 情報의 適合性에 의한 檢索效率評價가 支配的이나 情報의 適合성과 效用性의 概念은 區分되어야 하며 眞正한 意味의 檢索效率評價를 위해서는 情報의 適合성과 效用성에 의한 두가지 評價尺度가 共히 適用되어야 한다.

4) 絶對再現率(특히 現場시스템의 絶對再現率)은 蓄積情報量의 크기로 보

59) Shaw, R. Private Communication, quoted by Cleverdon in *Computer Journal*, 30(1974), p. 174.

아 그 測定이 거의 不可能하므로 一般的으로 概算再現率의 測定으로 評價된다.

5) 再現率과 精度率의 性能曲線은 特殊한 경우를 除外하고는 대체로 逆關係가 成立한다. 시스템의 檢索效率은 平均再現率과 精度率의 性能曲線으로 比較·評價된다.

6) 再現率과 精度率은 委託探索과 非委託探索의 경우 상당히 달라질 수 있다.

7) 委託探索의 경우 利用者가 시스템으로 부터 提供받고자하는 情報量을 限定하는 경우에는 相對再現率 또는 相對精度率이 測定되어야 한다.

8) 檢索效率의 主要支配要因으로는 探索情報의 範圍, 索引作成의 網羅性, 索引言語의 特定性, 利用者의 情報要求의 表現, 情報要求의 把握, 要求主題에 대한 探索者의 專門知識水準, 시스템의 探索戰略등이 있다.

한마디로 檢索效率의 評價는 情報의 適合性이라는 評價尺度가 애매하여 客觀的이고 數量的인 測定이 매우 어렵다. 많은 研究와 試驗이 계속되었으나 評價基準이나 評價方法에 뚜렷한 意見의 一致는 보지 못하였다. 그러나 情報檢索의 두가지 法則은 정립되었다고 생각한다.

第一의 法則은 '再現率과 精度率은 反比例關係'이며

第二의 法則, '完全한 檢索은 不可能하다는 事實'이다.

A Study on the Effectiveness of Information Retrieval

Koo-ho Yoon*

Retrieval effectiveness is the principal criterion for measuring the performance of an information retrieval system.

The effectiveness of a retrieval system depends primarily on the extent to which it can retrieve wanted documents without retrieving unwanted ones. So, ultimately, effectiveness is a function of the relevant and nonrelevant documents retrieved.

Consequently, "relevance" of information to the user's request has become one of the most fundamental concept encountered in the theory of information retrieval. Although there is at present no consensus as to how this notion should be defined, relevance has been widely used as a meaningful quantity and an adequate criterion for measures of the evaluation of retrieval effectiveness.

The recall and precision among various parameters based on the "two-by-two" table(or, contingency table) were major considerations in this paper, because it is assumed that recall and precision are sufficient for the measurement of effectiveness.

Accordingly, different concepts of "relevance" and "pertinence" of documents to user requests and their proper usages were investigated even though the two terms have unfortunately been used rather loosely

* Assistant Professor, Keimyung University.

in the literature. In addition, a number of variables affecting the recall and precision values were discussed.

Some conclusions derived from this study are as follows :

Any notion of retrieval effectiveness is based on "relevance" which itself is extremely difficult to define.

Recall and precision are valuable concepts in the study of any information retrieval system. They are, however, not the only criteria by which a system may be judged.

The recall-precision curve represents the average performance of any given system, and this may vary quite considerably in particular situations. Therefore, it is possible to some extent to vary the indexing policy, the indexing language, or the search methodology to improve the performance of the system in terms of recall and precision.

The "inverse relationship" between average recall and precision could be accepted as the "fundamental law of retrieval", and it should certainly be used as an aid to evaluation.

Finally, there is a limit to the performance (in terms of effectiveness) achievable by an information retrieval system. That is : "Perfect retrieval is impossible."