

情報檢索시스템의 評價에 관한 研究*

朴 寅 雄*

目 次

1. 序論
2. 情報檢索시스템의 分析
 - 2·1 情報檢索概念
 - 2·2 文獻檢索機能
 - 2·3 파일組織
3. 文獻檢索시스템의 評價方法
 - 3·1 檢索效率의 測定
 - 3·2 適合性
 - 3·3 文獻檢索시스템의 評價를 위한 原價計算方法
4. 文獻檢索시스템의 原價計算모델
 - 4·1 모델의 概念
 - 4·2 檢索活動
 - 4·3 檢索모델
5. 結論

1. 序 論

現代社會의 가장 두드러진 特徵중의 하나는 變化의 進行速度일 것이다. 그것은 멀리는 人間慾求에 기인하는 것이겠고 가까이는 科學技術의 發達로 인한 것이다. 人間慾求의 無限함을 생각할 때 科學技術도 無限히 發達될 것이 틀림없다.

이러한 發達과 變化를 이끄는 絶對的인 手段이 情報임은 또 情報에 대한 需要가 점점 增大할 것이라는 것은, 先進社會가 점점 情報社會

* 本 論文은 1981年度 淸州大學術研究助成費에 의하여 研究되었음.

*淸州大學校 圖書館學科

化하는 現狀에서 또 未來社會가 情報社會로 變化할 것이라는 科學者들의 理論에서 충분히 알 수 있는 일이다.

이와같이 個人이나 政府機關, 企業體 및 教育機關등으로 부터 더욱 많은 情報가 要求되고 提供됨으로써 더욱 效果的인 커뮤니케이션方法이 필요하게 되었다. 즉 점차 發展되는 社會에서 더 나은 情報流通方法이 필요하게 되었다.

情報流通을 活潑하게 하는 하나의 可能性은 컴퓨터의 이용이다. 컴퓨터 技術의 急速한 向上은 방대한 情報蓄積可能性을 가진 高速의 演算 및 記憶裝置의 開發을 초래하고 그러한 裝置는 情報의 蓄積, 檢索, 提供機能에 이용되게 되었다.

機械化된 檢索시스템의 發達로 內容을 認識하기 위한 多様な 文獻處理技術과 컴퓨터에 蓄積된 文獻을 檢索하는 여러가지 規則이 開發되었 다.

여기에서 關心을 두어야할 主要한 問題는 하나의 情報檢索技術이 다른 技術보다 나은 것인지 아닌지를 評價할 수 있어야 한다. 즉 情報檢索시스템의 業績을 測定하고 評價하지 않고서는 檢索시스템運用의 能率化를 기대할 수 없다.

그러나 실제로 情報檢索시스템의 業績을 測定하고 評價하는 일은 쉽지가 않다. 그것은 情報檢索의 業績은 情報檢索시스템의 目的, 文獻索引技術, 檢索方法 및 利用者의 知的 判斷등 諸要因에 의해 영향을 받기 때문이다.

이처럼 情報檢索시스템의 業績測定이 용이하지 않을지라도 效率인 情報檢索시스템의 運營을 위해서는 業績評價가 반드시 행해져야 한다.

本 研究에서는 그 評價方法을 찾는데 目的을 두어 原價를 중심으로 한 原價計算모형을 작성 제시하였다.

2. 情報檢索시스템의 分析

2.1 情報檢索概念

情報檢索機能을 수행하는 機械化된 시스템으로서는 質疑에 대한 解答을 제공하는 資料提供檢索시스템 (data providing retrieval system)

혹은 質疑解答시스템 (question-answering system)이 있고 다른 面으로는 質疑에 대하여 解答를 얻을 수 있는 參考文獻 혹은 文獻의 目錄을 제공하는 參考文獻檢索시스템 (reference retrieval system) 혹은 文獻檢索시스템 (Literature searching system)이 있다.

하나의 質疑가 質疑解答시스템에 주어지면 데이터파일에서 要求된 情報를 이끌어 내기 위하여 많은 文獻들을 시험한다. 청구된 情報가 주어지지 않을 때는 많은 관계文獻으로 부터 解答를 演繹해 내지 않으면 안된다.

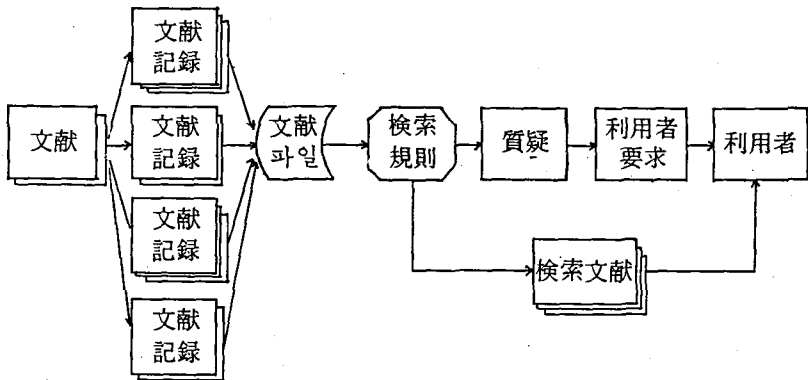
반면 文獻檢索시스템에서는 어느 質疑가 文獻記錄과 比較될 때, 文獻記錄속의 낱말이 質疑속의 낱말과 一致되면 그 文獻은 利用者の 要求를 만족시키는 것으로 판단한다.

質疑解答시스템은 아직은 開發段階에 있어서 제한된 데이터 베이스를 사용하며 高速의 大量의 記憶이 요구되기 때문에 많은 어려움이 있다.

따라서 本 研究에서는 文獻檢索시스템 혹은 參考文獻檢索시스템을 대상으로 한다.

文獻檢索시스템의 모델은 「그림 1」과 같다¹⁾.

「그림. 1」 檢索過程



1) Cooper, M. D., *Evaluation of Information Retrieval Systems* (University of California, Berkeley, ph. D., 1971.) pp 8-9.

「그림 1」의 한 쪽끝은 自然語대로의 文献 이 시스템으로 入力되는 것을 나타내고 다른 쪽은 利用者의 情報要求를 표시한다.

機械화된 文献檢索시스템은 主題內容을 결정하기 위하여 文献을 認識 分析한다. 分析된 內容에는 그 內容을 가장 잘 表現할 수 있는 標識을 부여하고 그것이 文献記錄 즉 索引標目이 된다. 索引標目이 주어지면 文献과일속에 蓄積된다.

다음으로 利用者가 情報에 대한 要求가 생기면 그것을 檢索시스템이 解讀할 수 있는 質疑의 형태로 變換시켜야 한다.

이렇게 質疑가 주어지고 文献記錄이 蓄積되면, 檢索시스템은 質疑와 文献記錄을 檢索規則에 따라 比較하여 一致되는 것이 있으면 파일에 命令하여 文献이나 書誌 抄録이 檢出되도록 한다. 檢出結果에 따라 利用者의 要求가 충족되었는지 質疑를 再構成해서 다음 段階의 檢索을 실시하게 된다.

2.2 文献檢索機能.

現在까지는 人間과 文献檢索시스템間의 情報流通現狀에 관한 研究가 부족한 실정이나 效果的인 檢索시스템의 設計를 위한 實驗을 통하여 많은 成果를 이룩하고 있다.

이와같은 文献檢索시스템의 現행모델은 一般的으로 세部分으로 構成되어 있다.

첫째 構成要素는 收集된 文献의 內容을 確認하고 分析組織하는 機能을 수행하는 것이다.

두번째 構成要素는 質疑를 作成하는 것이고

세번째 機能은 質疑가 형성되고, 文献의 內容이 分析 組織되어 이차資料가 작성 蓄積되면, 質疑와 이차資料를 比較하는 것이다.

이러한 세가지의 機能을 아래에서 살펴본다.

2.2.1 文献內容의 分析과 組織

文献의 主題를 表現 敍述하는 데는 세段階의 作業 過程을 거친다.

첫째로 文献의 情報內容이나 意味를 알기위하여 文献을 읽고

둘째로 그 文献의 어느 部分이 그 시스템의 目的에 부합되어 記錄되어야 하느냐를 選擇하는 일이며

셋째로는 실제로 選擇된 文献의 內容을 그 시스템에서 쓰이는 言語

로 表現하는 作業이다.”

文獻內容의 分析은 文獻이 어떤 主題의 것인지를 確認하는 機能으로서 索引 抄錄 및 文獻分類過程의 기초가 되는 것이다.

收錄된 文獻은 情報管理者들이 內容을 파악하고 그 중에서 利用者集團이 요구하는 또는 필요로할 것이라 생각되는 것들 만을 빼내어 디소오리스(Thesaurus) 등 標準用語集으로부터 該當記事 主題를 가장 잘 나타내는 語句 즉 標識을 부여하는 活動이 分析業務이다.”

內容의 分析에는 分析者의 主觀的 判斷이 要求 되므로 報情管理者의 專門知識이 필요하다.

組織이란, 情報分析을 통하여 文獻속의 資料나 情報들이 認識分類되고, 적합한 索引標目들이 붙여지면 그것들을 利用者들이 쉽게 檢索할 수 있는 方式을 염두에 두면서 体系的으로 排列하여 目錄, 索引, 抄錄, 抜萃文등 소위 次文獻을 작성하는 活動이다.”

文獻記述을 위한 標識는 그 內容을 시스템이 받아들일 수 있는 公認되고 標準화된 人爲的인 言語이다.

이와같은 標識가 索引言語이며 특히 標準화된 索引言語를, 사용함으로써 情報管理者와 利用者등 兩者의 質疑表現을 標準化하여 檢索效率을 높일 수 있다.

2.2.2 質疑作成(Query formulation)

利用者と 檢索시스템과의 關係는 質疑의 作成으로부터 시작된다.

“利用者의 要求를 코드化하는 過程이 質疑作成過程이며, 코드화된 要求가 質疑이다.”⁴⁾ 라고한 「메도우」(Charles T. Meadow)의 설명처럼 利用者의 質問은 檢索시스템이 理解할 수 있는 標準化되고 制限된 言語로 變換되어야 한다.

- 2) Vickery, Brian C., *On Retrieval System theory* (London, Butterworths; 1965) 李淳子 訳. 情報科学과 컴퓨터, 서울, 韓國圖書館協會 1980. pp. 29에서 재인용
- 3) 崔成眞著, 情報學原論. 서울, 亜細亞文化社, 1976. pp. 19~20.
- 4) 上掲書, pp. 19~20.
- 5) Meadow, Charles T. *The Analysis of Information Systems*, 2nd Edition. (Los Angeles, California, Melville Publishing Co. 1973) pp. 121.

이것은 情報管理者가 文献의 内容을 記錄하면서 標準化된 用語를 사용하여, 파일에 入力되는 單語의 數를 制限하여 經濟의 利益을 얻고, 情報管理者와 利用者 양쪽 主題表現의 標準化를 꾀하고 또한 파일안에 들어 있는 利用者の 要求에 부합되는 모든 文献을 찾아내는 確率을 높이고자 하는 活動과 상응되는 것이다.

그리고 質疑의 作成을 위하여서는 利用자가 自己가 必要로 하는 主題内容과 要求하는 情報의 範圍나 깊이를 精確하게 파악할 需要가 있다. 그것은 어느 主題의 情報를 어느 정도의 再現率과 正確率로 檢索할 것인가 즉 만족스런 檢索 結果를 얻는 일의 前提가 되기 때문이다.

2.2.3 檢索方法

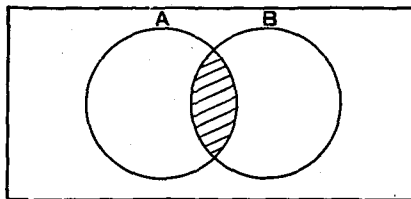
檢索의 目的은 檢索시스템에 들어 있는 모든 文献중에서 特定한 要求를 가진 質疑에 맞는 文献의 集合만을 가려내는 것이다. 後組合시스템에서는 質疑가 索引時에 사용된 索引語와는 다른 單語의 集合으로 表現될 수가 있다. 즉 부울演算子 AND, OR, NOT 으로 원래 索引時에 쓰인 索引語의 集合에서 여러가지 組合이 이루어져서 質疑가 작성된다. 이때 檢索活動은 質疑에 기초를 둔 論理關係에 의해서 수행된다.

여기에는 AND, OR, NOT 의 세가지 論理가 적용된다.

① 論理積 (AND 論理)

檢索過程에서 어떤 文献의 索引標目속에 그 檢索標目を 구성하는 키워드(Keyword)들이 모두 들어있는 文献記錄이 이 論理關係를 만족시키는 文献이다.

「그림 2」 論理積

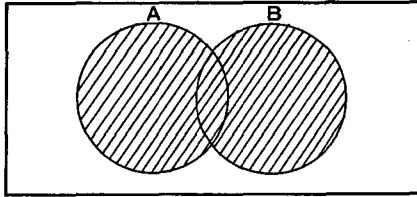


「그림 2」에서 A와 B에 共通으로 포함되는 즉 $A + B$ 인 斜線部分이 要求를 충족시키는 部分이다.

② 論理和 (OR 論理)

檢索過程에서 어떤 文献의 索引標目속에 그 檢索標目を 구성하는 키워드들이 들어 있는 總文献記錄이 論理和의 關係를 만족시키는 文献이다.

「그림 3」

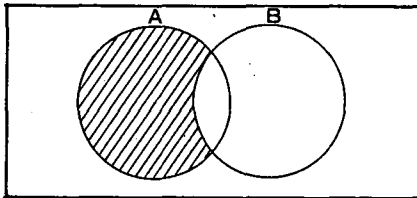


「그림 3」에서 A와 B의 모든 원들로 구성하는 즉 A OR B인 斜線部分이 要求를 충족시키는 文献이다.

③ 論理否定 (NOT 論理)

檢索過程에서 어떤 文献의 索引標目속에 그 檢索標目を 구성하는 키워드중 A 키워드가 들어 있는 文献은 檢索하되 B 키워드가 들어 있는 文献은 索檢하지 말라는 것이다.

「그림 4」



「그림 4」에서 A에 속하는 원 가운데 B에 속하지 않는 원들로 구성되는 즉 A OR B인 斜線部分이 要求를 충족시키는 文献이다.

2·2·4 檢索規則

檢索規則은 文献의 內容을 認識分析하여 얻어진 文献記錄과 파일과 質疑를 비교하는데 이용되는 규칙이다.

「쿠우퍼」(Cooper, M. D.)는 機械化된 檢索시스템에 적용될 檢索規則으로 比較規則(Matching rules), 共同規則(Associative searching) 및 集團化(Clustering) 등 세가지를 제시하였다.*

먼저 比較規則이란 가장 단순한 것으로 質疑속의 낱말을 조사 하고 그 낱말을 담고 있는 文獻記錄을 확인하는데 이용된다. 質疑에 맞는 記錄은 要求에 적합한 것으로 判斷하고 맞지않는 記錄은 檢索의 対象에서 제외된다.

이러한 過程으로 要求에 적합한 文獻과 그렇지 않은 文獻을 구분한다.

다음 段階의 檢索規則으로서는 質疑와 文獻記錄과의 一致의 程度를 나타내기 위하여 關聯文獻을 程度에 따라 分類하기도 한다. 一致의 程度는 檢索된 各 文獻에서 關聯用語의 數를 기초로한 順序로서 또는 質疑와 文獻間의 거리(Distance)로서 정한다.

다음으로 共同檢索過程은 檢索節次를 一般化하는 것으로 이 規則에 의하면, 質疑속의 用語는 본래의 質疑用語와 결합하게 되는데 이 결합된 質疑를 文獻파일과 비교하는 것이다.

마지막으로 集團化는 類似한 特性을 가진 文獻을 한 그룹으로 조직하면 그 그룹내의 文獻은 그룹외의 다른 文獻들 보다는 서로 더욱 類似할 것이라는 假定下에 論理的으로 關係있는 文獻은 미리 모아서, 檢索해야 할 文獻數를 최소로 하기 위한 것이다.

2.3 파일組織

索引作業이 끝나면 各 文獻에 대하여 索引레코오드(Record)가 작성되는데 그 레코오드에는 文獻內容, 文獻의 位置등 文獻에 대한 설명이 포함된다. 이러한 레코오드는 檢索을 용이하게할 수 있는 일정한 規則에 따라 組織하게 되는데 이렇게 組織된 것을 파일(file)이라 한다.

파일은 많은 레코오드로 구성되고, 各 레코오드는 하나 또는 둘이상의 필드(field)로 이루어 진다. 따라서 파일의 構造는 레코오드의 順序 또는 排列이다. 파일은 레코오드에 들어 있는 필드의 順序에 따라 排列된다.

6) Cooper. M. D., op. cit., pp. 18~23.

그리고 필드의 順序는 파일의 利用에 가장 적합하도록 정하여 진다. 즉 辭典體카아드目録에서 著者, 書名, 主題名등 세가지의 다른 필드가 하나의 알파벳 順序로 排列되어 있는 것과 같다.

파일안에서 一次的으로 順序를 정해주는 필드를 一次分類키이 (Primary sort key), 一次分類키이 안에서 다시 順序를 정해주는 필드를 二次 分類키이 (subordinate sort key)라 한다.⁷⁾

가장 단순한 파일構造는 連續組織 (Sequential Organization) 이다. 파일이 論理的인 順序에 따라 排列된 組織을 말한다. 連續組織에서는 아주 간단하다는 長點은 있으나 파일속의 특정한 레코오드를 찾기 위해서는 그 레코오드가 發見될 때까지 차례로 레코오드를 조사해야 하는 不便이 있다. 새로운 레코오드의 追加時에는 카아드目録인 경우에는 쉽지만 印刷된 藏書目録인 경우에는 組版을 다시 해야 하는 어려움이 있다.

다른 파일構造로는 無作為組織 (Random Organization) 혹은 直接組織 (Direct Organization)이 있다. 파일의 어느 곳이나 直接接近할 수 있기 때문에 이 시스템에서는 어느 레코오드나 찾는 時間이 같다.

直接接近方法을 理解하는 데는 키이概念과 레코오드의 蓄積位置概念이 중요하다. 無作為 혹은 直接파일組織은 키이와 蓄積位置사이의 서로간의 變化에 의하여 레코오드를 組織하고 位置를 정하게 되는 技術이다.

또 다른 파일構造로는 逆組織 (Inverted Organization) 이 있다.

대부분의 機械화된 檢索시스템은 逆組織이된 索引파일을 사용한다.

이것은 한 主題名이 표시된 한장의 카아드에 그 主題를 포함하는 모든 文獻들의 住所를 記入하는 索引파일이다.⁸⁾

마지막으로 체이닝 (Chaining)도 파일組織의 하나의 方法이다.

主題, 著者, 書名사이의 論理的 關係를 보유하고 있는 書誌記錄의 蓄積에서 특정 主題로부터 그 主題에 관해 著述한 파일속의 모든 著者에게로 유도시키려할 때 또 著者로부터 그 著者が 著述한 主題의 各

7) 李淳子 譯, 情報科学과 컴퓨터, 서울, 韓國圖書館協會, 1980 pp 54-55

8) Cooper, M. D., op. cit., pp. 30-31

9) 崔成眞著, 前掲書, pp. 225

文献을 지시하고자 할 때 체이닝技法을 사용한다. 재래식의 「보라」 「도 보라」参照가 이런 성질의 것이다.

機械화된 시스템에서 체이닝技法을 쓰는 것은 관계있는 레코오드를 바로 잇달아 계속해서 넣을 수 있는 자리가 蓄積媒体에 없거나 또는 앞으로 들어갈 새로운 레코오드의 자리를 미리 잡아놓는 費用을 절감하기 위해서 이다.

3. 文献檢索시스템의 評價方法

檢索시스템의 評價에 이용될 수 있는 方法으로서는 檢索效率의 測定과 時間方法 및 原価方法등이 있고 그 중에서 效率測定方法이 重要하게 다루어져 왔다. 그러나 本研究에서는 時間方法과 原価方法을 살펴 보 고자 한다.

3.1 檢索效率의 測定

檢索시스템의 評價와 관계되는 많은 研究가 檢索시스템의 成果를 반영하기 위한 效率의 測定에 集中되어 왔다. 理想的인 情報檢索能力은 不必要한 情報를 포함하지 않고 要求된 情報모두를 찾아 내는 것이다. 이러한 能力을 數量的으로 測定하기란 매우 어려운 것으로서 主로 利用者 자신들이 判斷하여 決定할 問題인 것이다.

지금까지 檢索시스템의 效率測定方法으로서는 再現率과 正確率이 가장 널리 사용되고 있다.

「表1」 檢索效率統計表

	관련 문헌수	관련없는문헌수	
검색문헌수	a	b	a + b
검색않된문헌수	c	d	c + d
	a + c	b + d	a + b + c + d

「表1」에서 “a”는 檢索된 關聯文献數, “b”는 關聯되지 않는 檢索된 文献數, “c”는 關聯되나 檢索되지 않는 文献數, “d”는 關聯 되지도 않고 檢索되지도 않는 文献數를 나타낼 때, 再現率(R)은 $R = a/a + c$ 로.

正確率(p)은 $P = a/a + b$ 로 표시된다. 再現率は 關聯된 文献總數 $a + c$ 에 대한 檢索된 關聯文献數 a 의 百分率을, 正確率は 檢索된 文献總數 $a + b$ 에 대한 檢索된 關聯文献數 b 의 百分率을 말한다.

이상의 두가지 比率에서 檢索시스템의 目的은 많은 數의 質疑에 대하여 再現率과 正確率을 最大로 되게 하는 것이다.

이러한 再現率과 正確率에 關連하여 檢索效率測定方法에 關한 研究 몇가지를 살펴본다.

먼저 「뮤어즈 (Mooers, E. F.)가 非關聯文献比率이 $\frac{b}{b+d}$ 임을 밝혔다.¹⁰⁾

그리고 「펠리」(J. W. Perry.) 「켄트」(Allen Kent) 「벨리」(Berry, M. M.)가 아래와 같은 比率을 얻었다.¹¹⁾

$$\text{不動率(Resolution factor)} = \frac{m}{n}$$

$$\text{排除率(Elimination factor)} = \frac{n-m}{n}$$

$$\text{正確率(Pertinency factor)} = \frac{W}{m}$$

$$\text{雜音率(Noise factor)} = \frac{m-W}{m}$$

$$\text{再現率(Recall factory)} = \frac{W}{x}$$

$$\text{漏落率(Omission factor)} = \frac{x-W}{x}$$

위 式에서 “n”은 시스템에 蓄積된 文献數, “m”은 檢索된 文献數, “W”은 檢索된 文献中 關聯文献數, “x”은 蓄積된 文献中 關聯文献數를 나타낸다.

물론 이러한 比率은 檢索規則, 索引의 깊이와 一貫性 및 파일속의 文献의 質疑에 影響을 받는 것이며, 또 이와같은 간단한 比率로서는 시스템의 部分的인 評價는 可能할지언정 全體的인 評價를 완전히 할수 없다는데 우리의 不滿이 있다.

10) Mooers, E. F. "Evaluation of the Performance of an Information Retrieval System by Modified Mooers Plan" *American Documentation*, 14 : 1 (January 1963) pp. 28~34

11) Kent, Allen, *Information Analysis and Retrieval* (New York, Awiley-Becker-Hayes Pub. 1966) pp. 315~316.

3·2 適合性

앞서 檢索效率의 測定方法을 살펴보았고 이젠 시스템에 의하여 檢索된 文獻이 利用者의 要求에 適合한지의 여부를 알아 보는 問題를 考察해 본다.

「랑카스터(Wilfrid F.Lancaster)의 지적대로 適合性判斷은 利用者의 情報要求에 관련된 것이어야 하기 때문에 判斷은 情報要求者에 의해서만이 이루어 진다.¹²⁾ 따라서, 適合性이란 利用者의 情報要求에 관련된 文獻에 대한 利用者 스스로의 滿足性判斷에 근거를 둔다.

適合性判斷에 관한 한 研究는 判斷에 影響을 미칠 수 있는 變數를 文獻自体, 文獻의 形態, 質疑, 判斷立場, 表現方法 및 人間의 能力 등을 지적하고 있다.¹³⁾

이상의 變數들 중 가장 基本的인 것은 人間의 能力이며, 主題에 관한 專門의 知識이 그 能力을 좌우하는 것이다.

나아가 特定한 文獻에 대한 利用者의 滿足度를 예상할 수 있으면 情報檢索過程의 效率를 상당히 向上시킬 수 있을 것이다. 利用者의 滿足度를 예상하는 데는 시스템이 文獻을 제공하기 전에 利用者의 知識水準을 파악하는 것이 또한 중요한 것이다.

利用者가 어느 한 시스템에서 다른 시스템에서 보다 더욱 滿足하는 가를 評價하기 위하여 各 構成要素의 單位原価를 계산하고 그 原価를 기준으로 시스템을 比較하여야 할 것이다.

適合性 概念을 기초로한 檢索效率의 測定方法에 의한 시스템評價에서는 評價의 範圍가 極 制限될 수 있음을 알 수 있다.

다음 項目에서 논의되는 原価計算方法이 檢索시스템의 包括的인 評價를 위한 더 넓은 展望을 제시하게 될 것이다.

3·3 文獻檢索시스템評價를 위한 原価計算方法

12) Lancaster, F. Wilfred, *Information Retrieval Systems, Characteristics, Testing and Evaluation* (New York, John Wiley & Sons, Inc. 1968) pp. 121.

13) 李淳子 譯. 前掲書 pp. 65-66

시스템分析은 特定시스템의 모든 部門의 評價過程으로서¹⁴⁾ 檢索시스템評價를 위해서는 그것의 分析이 선행되어야 하고, 分析에서 먼저 고려되어야 할 事項은, 시스템이란 全体目的을 성취하기 위해 作用하는 여러가지 要素로 構成되어 있다는 점이다. 시스템分析은 먼저 시스템의 定義와 目的 그리고 시스템作用의 制限条件등을 살피고 그것들에 따라서 分析者는 시스템의 各 構成要素들을 시험하게 된다.

檢索시스템의 構成要素로서는 内容分析과 索引作成, 蓄積 및 比較檢索등이 있다¹⁵⁾

各 構成要素들을 分析함으로서 시스템全体로서의 機能遂行을 目的成就方向으로 이끌 수 있을 것인가를 評價할 수 있게 된다.

이러한 分析을 통한 評價에 原價計算方法이 效果적이다.

原價計算方法을 이용한 圖書館業務의 評價研究는 상당한 水準에 이르고 있다. 한동안 圖書館에서 收書와 編目業務의 原價와 동시에 參考質疑回答業務의 原價算定에 많은 흥미를 가졌었고,^{16) 17) 18) 19) 20) 21)} 圖書館

-
- 14) Brightman, Richard W., *Information System for modern management*. (New York, The Macmillan Company, 1972) pp.28
 - 15) Meadow, Charles T., op. cit., pp 121~131
 - 16) Landau, Herbert B. "The Cost Analysis of Document Surrogation : A Literature Review," *American Documentation*, 20: 4 (October, 1969) pp.302-309.
 - 17) Leimkuhler, Ferdinand F and Michael D. Cooper "Cost Accounting and Analysis for Univesrsity Libraries", *College & Research Libraries*, 32 (November, 1971) pp.449-464.
 - 18) Brutcher, Constance, Glend Gessford and Emrmet Rixford, "Gost Accounting for the Library," *Library Resources and Technical Services*, 8:4 (Fall, 1964) pp.413~429.
 - 19) Wynar, Don and others, "Cost Analysis in a Technical Services Division," *Library Resoures and Technical Services*, 7 (Fall, 1963) pp.312~326.
 - 20) Fasana, Paul J., "Processing Cost for Science Monographs in the Columbia University Libraries," *Library Resources and Techical Services*, (winter 1967)pp.97-107.
 - 21) 朴寅雄, 종합대학도서관업무의 원가계산에 관한 연구. 成均館大學校大學院 碩士學位論文 1976.

業務의 分析을 위한 原価모델 역시 研究開發되었다.”²²⁾

그러나 文獻檢索시스템의 分析을 위한 分析의 原価모델에 관한 研究는 부족한 실정이다. 다음 章에서 이 問題를 살펴 보기로 한다.

4. 文獻檢索시스템의 原価算計모델

4.1 모델의 概觀

「켄트」(Allen kent)는 利用者側에서본 檢索시스템의 成果를 判斷하는 基準이 되는 要素를 다음과 같이 보았다.²⁴⁾

- ① 再現率
- ② 正確率
- ③ 勞力: 關聯情報를 얻는데 소비된 時間과 에너지
- ④ 應答時間(Response time): 質疑作成으로 부터 만족스런 解答을 얻을 때까지의 時間

⑤ 情報提供과 出力의 形應: 適合性判斷能力

⑥ 文獻入力政策: 蓄積된 文獻의 出版時期, 信憑性 및 言語

이상의 要素중에서 原価算計과 관계되는 것은 勞力과 應答時間이다. 本研究에서는 原価모델에 포함될 要素를 다음과 같이 選定하였다.

① 시스템側 檢索原価: 확정된 價格으로 利用者에게 奉仕한다.

② 利用者側 檢索原価: 시스템의 이용으로 소비된 時間과 勞力에 대한 價值이고 이것은 시스템이용으로 지불되어야 할 費用에 더하여진다.

③ 利用者が 特定시스템利用으로 얻을 수 있는 利益.

이상의 要素를 고려한 原価모델에서 利用者は 質疑를, 시스템은 文獻檢索에 도움될 檢索規則을 작성한다.

4.2 檢索活動

22) Jain, Aridaman K. A statistical Study of Book Use, ph D. Dissertation, Purdue University, Lafayette, Indiana, 1967.

23) Morse, Philip M. *Library Effectieeness: A system Approach*, Mit, Press, Cambridge, Massachusettes, 1968.

24) Kent, Allen, op. cit., pp. 314~315,

自動化된 檢索시스템에 의한 情報檢索活動은 利用者와 컴퓨터間的 相互作用으로 이룩된다. 人間과 機械間的 相互作用에서 利用者の 要求가 많고 專門的일 수록 目的을 이루기 위한 시스템의 活動이 더욱 복잡해질 것이다.

아래에서 利用者와 시스템間的 相互作用을 分析해 본다.

利用者と 시스템間的 相互作用을 檢索活動의 進行상황에 따라 段階로 구분하고 各 段階別로 利用者와 시스템活動間的 關係를 살펴보면 다음과 같다.²⁵⁾

첫째단계 : 檢索前 活動段階

① 利用者活動

- 1) 利用者が 시스템에 要求할 事項을 결정한다.
- 2) 要求를 시스템言語로 變換시킨다.
- 3) 質疑作成

② 시스템活動

- 1) 質疑의 構文을 체크한다.
- 2) 디소오러스(Thesaurus)를 調査한다. 디소오러스는 質疑속에 있는 낱말의 一般性이나 特殊性을 利用자가 알도록 하고 質疑의 범위를 넓게 또는 좁게 하도록 한다.

둘째단계 : 檢索活動段階로서 利用자가 待期 하는 동안 시스템은 比較方式 즉 檢索規則을 선택한다. 記錄된 利用者要求가 蓄積된 文献의 記錄 즉 2次資料와 比較되는 段階이다. 檢索過程에서 시스템의 活動이 가장 큰 段階이기도하다.

마지막단계 : 檢索後 活動段階이다.

① 利用者活動

- 1) 利用者は 出力된 文献을 읽고
- 2) 關聯文献을 確認하고
- 3) 關聯文献을 利用한다.

② 시스템活動

- 1) 시스템은 利用者の 要求를 만족시킬 수 있는 文献을 제공한다.
- 2) 檢索成果를 測定한다.

25) Cooper, M. D., op. cit., pp. 59~61

3) 피이드백機能으로 利用者の 満足度에 비추어 檢索節次上的 오류를 수정한다.

이상과 같은 利用者和 시스템間的 相互作用에 영향을 미치는 가장 중요한 要因으로서는 利用者の 精確한 情報要求의 表現이며, 그러한 表現에는 다음과 같은 事頂들이 고려되어야 한다.”

- ① 利用者の 要求事頂에 대한 正確한 說明.
- ② 檢索의 目的認識
- ③ 利用者가 요구에 관련된다고 믿는 文献의 論題把握.

4.3 檢索모델

4.3.1 檢索시스템의 原価

質疑에 대한 檢索시스템運用의 總原価는 시스템原価와 利用者原価를 合한 것이다. 즉

$$\text{檢索시스템總原価} = \text{시스템원가} + \text{이용자원가} \dots\dots\dots (1)$$

시스템原価는 시스템의 單位時間當 原価에 檢索에 소요된 시스템의 總時間을 곱한 것이고, 利用者原価는 利用者の 單位時間當 原価에 檢索에 소요된 利用者の 總時間을 곱한 것이다. 따라서 위의 式은 다음과 같이 표시된다.

$$\begin{aligned} \text{檢索시스템總原価} &= (\text{시스템총시간} \times \text{시스템단위시간당원가}) \\ &+ (\text{이용자총시간} \times \text{이용자단위시간당원가}) \dots\dots\dots (2) \end{aligned}$$

4.3.2 시스템原価

시스템의 活動은 檢索前活動, 檢索活動 및 檢索后活動으로 구분되고 시스템原価는 다음과 같이 표시된다. 즉

$$\begin{aligned} \text{시스템原価} &= (\text{검색전활동시간당원가} \times \text{시간}) \\ &+ (\text{검색활동시간당원가} \times \text{시간}) \\ &+ (\text{검색후활동시간당원가} \times \text{시간} \dots\dots\dots (3) \end{aligned}$$

各活動의 單位時間當原価는 다음과 같이 計算된다.

26) Lancaster, F. Willfrid, op. cit., pp. 181~189.

檢索前活動의 單位時間當原價

= 채널시간당원가 + 중앙처리장치시간당원가 + 정보기록매체시간당원가 (4)

檢索活動의 單位時間當原價 = 비교원가 + 문헌기록원가 (5)

檢索后活動의 單位時間當原價 = 채널시간당원가 + 중앙처리장치시간당원가 + 정보기록매체시간당원가 (6)

「式 5」에서 檢索시스템의 比較原價는 檢索方法에 따라 다르기 때문에 일정하지가 않다. 즉 比較原價는 質疑에 포함된 用語의 數, 論理記号의 數, 파일속의 文獻記錄의 數 및 各記錄속의 낱말의 數와 관계되고 그 밖에 컴퓨터의 中央處理裝置, 情報記錄媒体 및 채널과도 關係된 것으로 一定한 單一한 公式으로 計算될 수 없다.

다음으로 文獻記錄에 대한 總原價는 分析組織原價, 蓄積原價 및 파일管理原價등으로 나누어 진다.

먼저 情報센터에 文獻이 收集되면 內容을 파악하고 필요하다고 생각되는 資料나 情報를 분리하여 후에 檢索하기 편리하도록 배열해 두기 위하여 標識를 부여하고 目錄 索引 抄錄 拔華文등 索引標目으로서의 2次文獻을 작성한다. 이와같은 機能遂行에 소요된 原價가 分析組織原價이다.

그러나 지금까지 2次資料 作成原價를 計算하는 一定한 方法은 볼 수 없다.

두번째 構成原價인 蓄積原價에 영향을 미치는 여러가지 要素들이 있다. 컴퓨터 蓄積裝置의 賃借費用, 蓄積裝置의 統制費用, 이런 裝置의 利用能力 및 記錄속의 言語의 數등이 그것이다.

마지막 構成原價로서는 파일管理原價가 있다. 檢索시스템에서는 두 종류의 管理機能이 수행된다. 즉 파일로 부터 레코오드를 檢索하는 일과 그 파일을 작성하고 유지하는 일이다.

그 밖에 管理原價로서 채널原價, 中央處理裝置原價 및 蓄積媒体原價가 있다.

파일管理原價式은 기본적으로 (4)式과 같다. 즉
파일管理原價 = 채널원가 + 중앙처리장치원가 + 정보기록매체원가 (7)

결국 文獻記錄의 原価는 다음과 같은 式으로 표시된다. 즉

$$\text{文獻記錄原価} = \text{분석조직원가} + \text{축적원가} + \text{파일관리원가} \dots\dots\dots(8)$$

4 · 3 · 3 利用者原価

檢索過程에서 고려될 두번째 主要原価要素는 利用者이다.

利用者の 總原価는 利用者の 單位時間當原価에 質疑作成에서 부터 關聯文獻의 提供時까지 利用자가 소비한 時間을 곱한 것이다. 즉

$$\text{利用者原価} = \text{이용자 활동 시간당원가} \times \text{이용자총시간} \dots\dots\dots(9)$$

利用者原価에도 檢索前活動原価, 檢索活動原価 및 檢索后活動原価 등 세가지 原価要素로 구분된다.

따라서 利用者原価는 아래와 같은 式으로 표시된다. 즉

$$\text{利用者原価} = (\text{검색전 활동시간당원가} \times \text{소비시간}) + (\text{검색활동시간당원가} \times \text{소비시간}) + (\text{검색후활동시간당원가} \times \text{소비시간}) \dots\dots\dots(10)$$

여기에서 利用자가 시스템을 이용하지 않을 경우에는 그 시스템은 다른 利用者나 또는 다른 일을 위하여 奉仕할 수 있을 것이다. 그 때 利用者は 利用하지 않는 時間에 대한 費用은 지불하지 않을 것이며 반면 시스템이 다른 業務를 수행하고 있다면 利用者は 시스템과의 對話를 위하여 기다려야할 것이다.

이러한 事實은 (10)의 式을 다음과 같이 변경되도록 한다.

$$\begin{aligned} \text{利用者原価} = & \{ \text{검색전활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \} \\ & + \{ \text{검색활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \} \\ & + \{ \text{검색후활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \} \\ & \dots\dots\dots(11) \end{aligned}$$

위의 式에서 待期時間이란 質疑가 접수되고 檢索活動이 시작될때 까지의 時間을 말한다.

利用자가 시스템을 이용하는 時間은 利用者の 資質과 知識뿐 아니고 시스템에 대한 經驗과도 관계가 있다. 그러나 이러한 時間의 測定은 매우 어려운 것이다.

5. 結 論

前章에서 밝혀진 바와 같이 本 研究에서는 檢索시스템의 原価中心

점을 시스템과 利用者로 선정하고 먼저 이들의 原価를 計算하여 둘을 합하면 檢索시스템의 總原価가 계산된다.

그런데 시스템活動과 利用者活動은 각각 檢索前活動, 檢索活動, 檢索後 活動으로 구분된다. 시스템活動중 檢索前活動과 檢索後活動은 채널이용活動 中央處理裝置利用活動 및 情報記錄 媒体에 관한 活動으로, 檢索活動은 比較活動과 文獻記錄活動으로 구성된다. 이들 細区分된 活動의 原価를 계산함으로써 檢索시스템의 總原価가 계산된다.

이렇게 計算된 原価로서 利用者は 더 效率的인 시스템을 選擇할 수 있게 된다.

本研究에서 밝혀진 原価計算모델은 다음과 같다.

① 文獻檢索시스템의 總原価計算모델

$$\text{文獻檢索시스템의 總原価} = (\text{이용자총시간} \times \text{이용자활동시간당원가}) + (\text{시스템총시간} \times \text{시스템활동시간당원가})$$

② 시스템原価計算모델

$$\text{시스템原価} = (\text{검색전활동시간당원가} \times \text{시간}) + (\text{검색활동시간당원가} \times \text{시간}) + (\text{검색후활동시간당원가} \times \text{시간})$$

1) 檢索前活動의 單位時間當原価 = 채널시간당원가 + 중앙처리 장치 시간당원가 + 정보기록매체시간당원가

2) 檢索活動의 單位時間當原価 = 비교원가 + 문헌기록원가
文獻記錄原価 = 분석조직원가 + 축적원가 + 파일관리원가

3) 檢索後活動의 單位時間當原価 = 채널시간당원가 + 중앙처리 장치 시간당원가 + 정보기록매체시간당원가

③ 利用者原価計算모델

$$\text{利用者原価} = \{ \text{검색전활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \} + \{ \text{검색활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \} + \{ \text{검색후활동시간당원가} \times (\text{소비시간} + \text{대기시간}) \}$$

A study on evaluation of information retrieval system

In Ung Park
Dept of Library Science
Cheongju University

Information is an essential factor leading the rapid progress which is one of the distinguished characteristics in modern society. As more information is required and as more is supplied by individuals, governmental units, businesses, and educational institutions, the greater will be the requirement for efficient methods of communication.

One possibility for improving the information dissemination process is to use computers. The capabilities of such machine are beginning to be used in the process of Information storage, retrieval and dissemination. An important problems, that must be carefully examined is whether one technique for information retrieval is better for worse than another.

This paper examines problem of how to evaluate an information retrieval system. One specific approach is a cost accounting model for use in studying how to minimize the cost of operating a mechanized retrieval system. Through the use of cost analysis, the model provides a method for comparative evaluation between systems.

The general cost accounting model of the literature retrieval system being designed by this study are given below.

1. The total cost accounting model of the literature retrieval system.
The total cost of the literature retrieval system =

(the cost per unit of user time X the amount of user time) + (the cost per unit of system time X the amount of system time)

2. System cost accounting model

system cost =

(the pre-search system cost per unit of time X time)
+ (the search system cost per unit of time X time) + (the post-search system cost per unit of time X time)

1) Pre-search system cost per unit of time

= cost of channel per unit time + cost of central processing unit per unit time + cost of storage per unit time

2) Search system cost per unit of time

= comparison cost + document representation cost.

3) Post-search system cost per unit of time.

= cost of channel per unit time + cost of central processing unit per unit time + cost of storage per unit time

3. User cost accounting model

Total user cost =

[pre-search user cost per unit of time X (time + additional time)] + [search user cost per unit of time X (time + additional time)] + [post-search user cost per unit of time X (time + additional time)]