

# 온라인 檢索時代의 到來

韓 晟 熙  
(技術情報室長)

..... < 차 례 > .....

- I. 序 論
- II. 온라인時代의 到來
- III. 情報의 意義
- IV. 情報檢索
- V. 檢索效率
- VI. 檢索用語
- VII. 데이터베이스
- VIII. 온라인檢索의 準備
- IX. 質問式의 檢討
- X. 맺 음 말

## I. 序 論

情報라는 말이 오늘 날과 같은 뜻으로 사용된지도 오랜 시간이 지났다. 情報量의 增加에 따라 處理의 必要性도 그만큼 커져 이를 위해 電算機의 데이터處理機能도 多方面으로 開發되고 있으며, 이에 따른 周邊機器 역시 整備되어지고 있다. 이와 같은 諸機能은 處理指令에 따라 制御되게 되는데, 이 指令을 우리들은 “프로그램” 이라고 부른다. 現在 “情報處理” 란 機械를 利用하는 自動化란 意味와 같기도 하다.

이 自動化된 시스템과 시스템으로 얻어진 結果는 우리들 日常生活의 여러部門

에 파고들고 있는데 當然之事로 받아들이고 있다. 例컨대, 電氣, 水道, 電話, TV 등의 使用料金の 請求와 支拂處理, 預金の 引出, 俸給支給, 税金徵收, 高入, 大入 試驗의 處理 등이다. 그밖에도 우리들이 알지못하거나 느끼지 못하는 部門에 까지 파고들고 있다. 그런데 위의 例는 計算業務와 關係되는 것으로서 文獻檢索과는 다르다.

여기서는 우리들이 必要로하는 文獻情報를 電算機를 利用하여 檢索하는데 必要한 事項과 在來式 方法에 比하여 어떤 利點이 있고 缺點이 있는가에 대해 살펴보고 어떻게 對處할 것인가를 考察하여 보고자 한다.

## Ⅱ. 온라인時代의 到來

情報處理는 컴퓨터利用技術에 있어서 主要分野의 하나이다. 航空會社나 鐵道廳에서 座席을 豫約할 때 中央과 特定回線으로 連結된 端末裝置로 指定座席券의 豫約과 發賣의 機能 등이 컴퓨터에 의해 수행될 수 있다. 豫約은 온라인檢索에 該當되며 온라인出力機能은 發賣에 該當될 것이다. 즉, 利用者가 要求하는 航空機(列車)名, 利用月日, 乘下車驛名, 希望하는 座席 등을 情報파일(file)에서 探索한다. 이것은 端末裝置를 操作하는 擔當者가 對話를 하게되는데, 이 對話는 通信網을 통하여 販賣센터의 컴퓨터와 接續되어 該當하는 座席이 있으면 端末裝置에 發賣指令이 내려가 回答內容이 나오게 되고 料金計算이 되어 指定券이 나옴과 同時에 販賣센터에는 豫約되었다는 記錄이 된다. 이와 같은 處理를 하는데 要하는 時間은 不過 몇십秒에 지나지 않는다. 이런 作業을 手作業에 의한다면 處理能力이나 迅速性, 正確性은 不問可知이다.

이것이 中央의 데이터파일과 通信回線을 連結, 端末裝置를 利用하는 情報提供 서비스의 典型이라 할 수 있으며, 온라인情報檢索의 패턴이라 하겠다.

産業研究院은 1979년부터 온라인情報檢索을 開始, 1984년 5월에는 釜山支院에, 12월에는 大邱支院에 特定回線과 連結된 端末裝置를 設置, 支院에서도 情報를 즉시 檢索할 수 있으며, 그밖의 支院에도 設置하여 利用者가 直接 情報檢索이 可能하도록 計劃 檢討中이다.

이와 같은 端末裝置는 研究室內에도 設置할 수 있는데, 그 設置運營費와 本 研究院이나 支院에 가서 直接 檢索하는 費用과를 檢討하여 設置하여야 할 問題라고 생각된다.

온라인檢索法의 特徵은 中央의 데이터베이스와 檢索者와의 對話方式이다. 그 對話를 어떻게 잘하는가는 情報部門의 專門家가 아니라 情報를 必要로하고 要求하는 研究者 自身이 直接 操作하는 것 以外는 없다. 그러나 研究者의 條件으로서 情報檢索法의 習得을 義務化시킨 研究機關은 별로 없는 것으로 안다. 情報檢索의 必要性을 認識하고 있다하더라도 실제로 研究檢討하는 者는 매우 적을 것이다.

通信手段이 發達되고 多様な 데이터베이스가 製作되어진 現在, 過去와 같이 圖書館에서 文獻을 뒤적거리는 時代는 이미 지나고 必要한 者가 스스로 端末機에 앉아 探索하는 온라인時代가 이미 닥쳐왔다고 본다.

### Ⅲ. 情報의 意義

情報(Information)의 定義를 論議할 必要는 없겠지만 簡單히나마 살펴보지 않을 수 없다.

TV, 라디오, 新聞, 週間誌 등과 같은 매스미디어가 生産되어 傳達되는, 社會의 事情을 알리는 것은 매우 많다. 그런데 이들을 情報라고는 부르지 않는다. 이는 自己가 遮斷 혹은 無視하면 存在하지 않음과 같기 때문이다. 그러나 어떤 意志를 가지고 이들에게 接하였을 때 비로소 情報로서의 存在가 認識되고 때로는 사라졌던 過去의 것도 情報로서 되살아난다. 文獻檢索에 있어서 遡及檢索이 이것이다.

情報는 發信(生)源이 있는 同時에 受信者가 있음으로써 비로소 兩者間에 存在(成立)하는 것이다. 사람이 意志를 가졌을 때에 受信者가 되는 것은 그 사람이 어떤 價値體系를 바탕으로 判斷하였기 때문에 情報로서 認識되어지는 것이다. 따라서, 情報에는 價値가 附加되어져 있다. 換言하면, 價値를 상실하였다고 判斷되면 이미 情報는 아니다. 그런데 價値觀은 사람에 따라, 狀況에 따라 差異가

있고, 時間과 더불어 變動하기 때문에 情報의 評價는 사람에 따라 다르며 時間과 더불어 變化하는 것이 普通이다.

#### IV. 情報檢索

情報檢索은 Information Retrieval 또는 Information Storage and Retrieval의 譯語로서 略하여 IR이라 하기도 한다. 즉, 情報의 蓄積과 蓄積된 情報에서 必要로하는 情報를 探索하는 2가지 機能을 뜻한다.

情報를 蓄積하려면 먼저 情報蒐集의 問題가 따른다. 왜냐하면 情報의 發生源이 分散되어 있기 때문이다. 우리들은 이 問題에 대해서 自己分野의 密度 높은 雜誌(Key Journal)를 選定하여 直接 購讀한다. 反面 密度가 낮은 雜誌는 抄錄誌(二次資料)에서 檢索하는 方法을 취한다.

必要로하는 情報의 50%를 커버하는 情報源을 Key Journal이라고 하면, 10餘種의 雜誌면 될 것이다. 이를 購讀함은 可能하겠지만 研究對象分野가 여럿에 걸쳐있다면 不可能에 가깝다. 한 分野에 대한 情報의 檢索範圍(coverage)를 10% 올리는데 2배의 雜誌가 必要하다고 한다. 만약 10種에서 50%의 情報를 얻었을 경우 60%로 올리려면 20種, 70%라면 40種이 必要하므로 時間과 勞力의 效果는 너무나 적다고 본다. 만약 Key Journal의 選定이 잘못되었다면 必要情報의 50%를 얻지 못할 것이다.

機械에 의한 檢索을 可能하게 하려면 情報의 蓄積을 完全하게 하지 않으면 안 된다. 그래야만 비로소 機械檢索시스템에 대한 信賴性이 부여되기 때문이다.

그런데 이와 같은 機械檢索에 있어서 뒤따르게 되는 問題點으로 利用者의 要求에 充分히 對應하여 나아가기 위한 充分한 情報를 準備해 놓지 않으면 안되는 일이다. 또한 充分한 情報가 準備되어 있다 할지라도 利用者의 擴大를 도모하지 않게 되면 이 機能의 最大值를 발휘하게 되지 못하는 問題가 뒤따르게 된다.

## V. 檢 索 效 率

檢索結果에 대한 效率을 측정하는 척도로서,

$$\text{再現率}(\alpha) = \frac{\text{檢索된 適合情報의 量 (H)}}{\text{蓄積된 情報中 適合情報의 量(A)}}$$

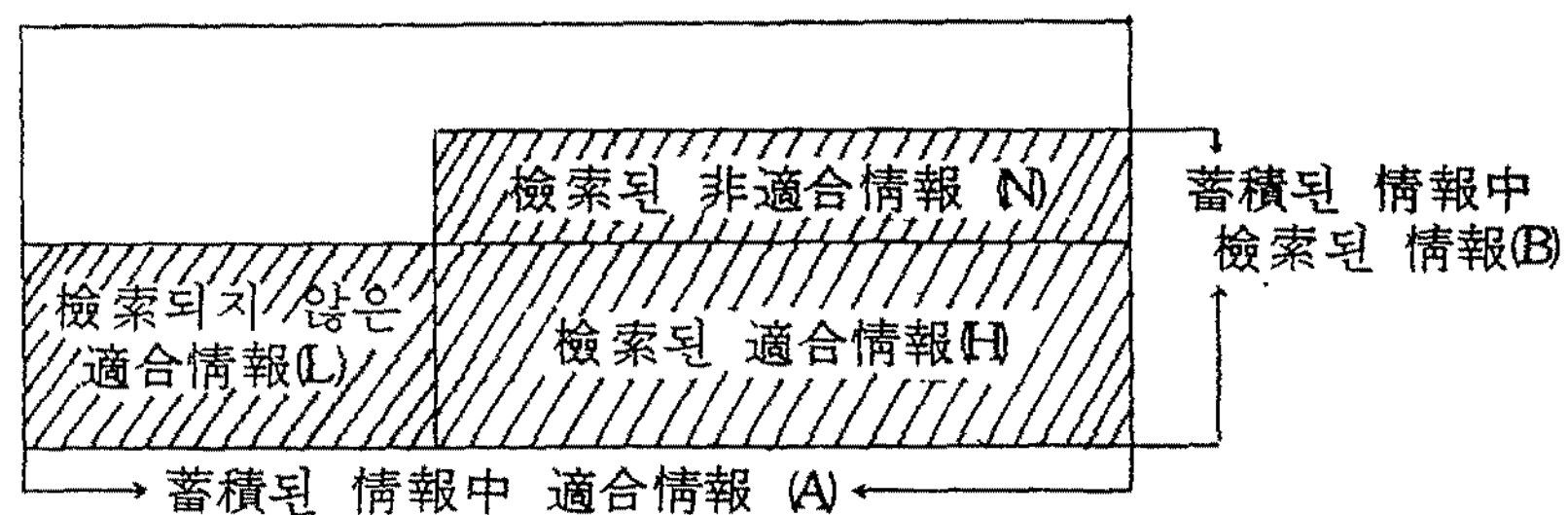
$$\text{適合率}(\beta) = \frac{\text{檢索된 適合情報의 量(H)}}{\text{蓄積된 情報中 檢索된 情報의 量(B)}}$$

理想的인 檢索의 경우 ABH가 완전히 일치하여  $\alpha, \beta$  모두 1이 된다.  
誤差의 觀點에서

$$\text{除外率} = \frac{\text{檢索되지 않은 適合情報(L)}}{\text{蓄積된 情報中 檢索된 情報의 量(B)}}$$

$$\text{雜音率} = \frac{\text{檢索된 非適合情報(N)}}{\text{蓄積된 情報中 檢索된 情報의 量(B)}}$$

〈蓄積情報 파일〉



要求者가 必要로 하는 情報를 빠짐없이 모두 檢索하여 不適合한 情報가 하나도 나오지 않는다면 가장 理想的인 檢索結果라 하겠다. 그러나 이와 같은 結果를 얻는다는 것은 어려운 일이기도 하다. 그 理由로는 檢索用資料의 蓄積過程中 索引作成이 不充分하였다든가, 누락 또는 資料의 收錄範圍 등에 問題가 있으며, 또한 要求主題에 대한 主題分析의 不充分, 要求者와의 意思交換不足 등 많은 要因이 發生하기 때문이다.

이와 같은 問題點들의 改善을 위해서는 많은 蓄積情報파일의 活用과 經驗이

必要하며, 檢索方式의 檢討 등으로 改善하여 나아갈 수 있는 것이다.

## VI. 檢 索 用 語

情報檢索을 컴퓨터側에서 보았을 때의 問題點은 自然語의 處理이다. 自然語를 使用하여 情報의 傳達을 하고 있지만 機械檢索時에는 自然語가 適合하지 못하다. 自然語는 柔軟性이 많으며 人間의 思考에 適合토록 되어있어서 複雜하고 애매함이 많기 때문이다. 例컨대 컴퓨터와 電算機, 電子計算機는 同義語라고 人間은 理解하지만 機械는 別個의 言語로 判斷한다. 때문에 情報檢索의 效率을 높이려면 言語의 管理가 必要하다. 自然語의 表現概念을 하나의 言語로 統一하여 索引語라 하였고 索引語에 對應하여 檢索에 使用하는 말을 檢索語라 하였는데, Keyword는 이에 該當한다. 위의 自然語, 索引語, 檢索語間의 關係를 規定하여 文獻檢索을 하는 體系 혹은 辭典의 形體를 나타내고 있는 것을 소위 Thesaurus라 일컬고 있다.

그리고, 檢索語와 檢索語의 上下關係를 表示함으로써 檢索을 效率化시킬 수 있다. <表 1>은 그 關係를 나타낸 것이다. 이외에 關聯語(R.T.=Related Term)가 設定되어 있으며, 또 同義語中 使用하지 않는 것과 使用하는 것을 UF 또는 USE, use For (...을 보라) 등의 表示에 의해 使用된다. 그리고 同義異語의 區別은 ( )로 表示한다.

例: bonds (chemical)

bonds (adhesive)

우리말에도 많은 問題點이 뒤따르는 바가 있는데, Thesaurus 를 만들때 同義語 (例: 미캐니즘, 機作, 機轉, 機序, 機構)나 同音異義語 (例, 容水量, 用水量)가

<表 1>

略 語		譯 語
T T	Top Term	最上位語
B T	Broader Term	上位語
N T	Narrower Term	下位語

問題가 될 것이다.

Thesaurus에 記錄된 索引語를 Descriptor라 하는데 Keyword와 대개 같은 意味로 생각하면 된다. 前述한 컴퓨터는 Descriptor이고, 同義語인 電算機나 電子計算機는 非디스크립터이다. Thesaurus에서는 利用頻도가 높은 것이 디스크립터로 되어있다.

## VII. 데이터베이스

*Chemical Abstracts* 誌나 *Biological Abstracts* 誌와 같은 抄錄誌(二次資料)는 그대로의 形態로는 컴퓨터로 檢索할 수 없다. 이것을 磁氣테이프에 넣어 機械可讀型으로 만든 것을 Data Base라 한다. 즉 어떤 目的으로 蒐集, 編輯된 情報의 集합을 말한다. 데이터베이스를 定義하면,

- ① 大量의 데이터가 組織化된 形態로 整理되어 있다.
- ② 1내지 여러개의 電子計算機로 읽을 수 있는 形態로 파일에 들어 있다.
- ③ 데이터의 蓄積, 檢索, 維持管理를 위해 시스템 및 프로그램이 完備되어 있어야 한다.

따라서 데이터베이스는 文獻情報 뿐만 아니라 數值, 人名 등도 해당된다. 즉, 情報가 一定한 基準下에 蒐集, 編輯되고 새로운 情報가 追加, 更新된 情報集團이 데이터베이스인 것이다.

데이터베이스의 特徵은 共有化에 있다. 相異한 시스템으로도 access되게끔 設計되어 利用者의 多樣한 要求에 對應되게끔 管理運營된다. 이 점이 對象範圍를 限定한 電算機의 파일과 다른 點이다.

現在 世界各國에서 市販되고 있는 데이터베이스는 約1,300種 가량이 된다하며, 그밖에 大學이나 企業에서 獨自적으로 開發한 데이터베이스도 많다. 그러나 이와 같은 데이터베이스들은 公開되어 지지 않은 것이 대부분이다.

우리들은 市販되고 있는 데이터베이스의 모두를 直接利用할 수 없으므로 購入, 整備하고 있는 機關을 通하여 access할 수 있는데 産業研究院에서는 그 業務도 수행한다.

데이터베이스중 CAS (Chemical Abstracts Service) file에 대해 略述하면, 이는 生化學, 有機化學, 物理化學, 應用化學, 化學工學 등의 情報가 收錄되어 있고, 情報源은 125 個國 以上の 15,000 餘種에 해당하는 學術誌의 情報와 27 個國의 特許 등이다. 연간 收錄되는 情報量은 45 萬件이 넘으며, 每年 그 量은 追加된다. 自然語를 Keyword로 使用할 수 있어 온라인檢索에 接하지 못하였던 사람도 타이프라이터와 같은 要領으로 簡單히 檢索할 수 있다. Subject Index로 檢索하는 從來의 方法에 比하여 簡便성과 迅速성은 特記할 만하다.

## VIII. 온라인檢索의 準備

온라인檢索을 시작하기에 앞서 準備하여야 할 事項을 알아본다. 왜냐하면 檢索操作을 하여 놓고 머뭇거리는 사이에 時間과 經費가 計算되므로 事前에 充分한 準備를 갖추어 놓는 것이 무엇보다 重要하다.

### 1) 目標을 明確히 할 것.

온라인檢索法에서 檢索結果는 該當文獻數가 몇件있다고 回答하므로 質問內容을 機械가 틀리지않게 理解하게끔 充分히 檢討한다. 關聯語, 上位語를 잘 使用하면 폭넓게 檢索할 수 있다. 막연한 테마의 調査는 機械檢索이 안되므로 檢索目標을 具體적으로 設定함이 重要하다.

### 2) 데이터베이스의 選擇

取扱機關(데이터뱅크)의 데이터베이스中 自己가 檢索하고자하는 情報가 어느 파일에 있나를 찾아야한다. 물론 각각의 파일에 관한 收錄內容, 範圍, 性格 등은 充分히 알고 있어야 하며, 그에 대한 매뉴얼도 갖추어야 한다.

데이터베이스의 選擇은 檢索의 누락, 適合性, 時間, 經費 등 모든 것을 좌우하게 되므로 신중을 기하고 사전에 專門人 또는 經驗者와의 相談은 檢索의 適合率을 높여 줄 수 있으므로 유의해야 한다.

### 3) 質 問 式

컴퓨터와의 對話는 質問에 대한 것만 機械는 대답하며 質問樣式이 檢索結果의



成敗를 決定한다. 質問은 Keyword 로하게 되는데, 예컨대 Information 이라고 하는 主題語로서 質問을 위한 打字를 하면 Information 을 갖는 該當文獻數(Documents)가 回答된다. 1 Keyword 로도 檢索이 되지만 正確히 檢索하려면 論理演算法으로 Keyword 를 組合시킨다. 이것을 質問式이라하며, 이 質問式의 作成이 온라인檢索에서 가장 重要한 作業이다. 그런데 利用者는 컴퓨터의 利用이 熟達치 못하여 利用者가 作成한 프로파일은 利用者의 要望을 滿足시키지 못하는 結果를 가져오는 일이 많으므로 經驗있는 檢索專門家의 助言은 檢索效率을 높이는 데 도움이 될 것이다. 그렇다고 專門家에게 一任하면 利用者의 專門知識을 살릴 수 없기 때문에 兩者의 協同作業에 의해서만 完全한 質問式이 成立된다고 보아야 한다.

## IX. 質問式의 檢討

質問에 使用된 檢索方式이 適當한가 否適한가를 確認하기 위하여 먼저 몇 件 程度를 出力(Browse)시켜 適合情報인가를 標題에서 檢討하여 보는 것이 바람직하다.

그리고, 檢索情報數가 豫想보다 많았을 때와, 적거나 없을 경우의 對策도 세워야 할 필요가 있다. 이 結果는 質問式이 잘못 作成되었던가 정말로 없던가 중 하나이기 때문이다.

檢索專門家は 가령  $(A+B+C+D) \times (E+F) \times D$ 와 같은 檢索式을 만든다 해도 專門家가 아닌 利用者는  $A \times B$ 이거나  $A \times B \times C$ 와 같은 單純한 形態를 취할 수 있다. 이러한 경우 文獻數가 極端으로 많이 나오든가, 그렇지 않으면 없을 경우가 있는데, 이 때에는 A와 B의 Keyword 를 再檢討하여야 할 것이다.

예를 들어 說明하면 INSPEC (Information Services for the Physics and Engineering Communities) 데이터베이스에서 Condenser \$에 관한 主題를 檢索할 때('76년~) 1041 Documents 에 2344 Occurrences 라는 회답이 나온다. 그런데 이와 같은 말로서 Capacitor \$라는 말이 있다. 이것을 檢索하면 17485 Occurrences 에 6608 Documents 라는 회답이 나온다. 만약 前者도 많다고 다음

質問을 하였을 때는 1041 Documents 에서만 찾아지므로 檢索率이 떨어진다. 그러므로 Condenser \$ or Capacitor \$ 로 하면 19829 Occurrences 에 7556 Documents 로 나온다. 이후 콘덴서에 대한 다음 질문을 하게 된다.

이와 같이 質問을 거듭하면 利用者가 願하는 情報를 찾아낼 수 있게 되지만 檢索專門家는 이를 짧은 時間에 누락없이 檢索하여 낼 수 있으므로 숙달될 때까지는 프로파일 작성연습을 하여 둘 必要가 있다.

## X. 맺 음 말

現代를 情報化社會라고 한다. 情報化社會에서 살고있는 우리들은 情報에 뒤떨어지지 않으려고 모든 勞力을 다한다. 하루가 다르게 情報産業이 發達해가고 있음은 情報의 利用이 그만큼 높아가고 있음을 의미하는 것이다.

從來와 같이 情報의 生産者와 情報의 檢索者가 따로따로 있음이 아니라 生産者가 곧 檢索者가 되고 必要情報도 研究室內에서 찾아볼 수 있는 온라인時代에 접어 들었음을 실감하고 이에 順응하여야 할 것이다.

다행히 産業研究院에서는 過去부터 온라인 情報檢索 時代를 대비하여 敎育과 實習을 통하여 普及하여 왔으며, 이번에 이를 더욱 強化하여 實行하고 있으므로 詳細한 技法과 實習으로 스스로가 必要情報를 檢索하는 온라인時代를 맞이하기 바란다.