

情報檢索시스템의 評價試驗

Cyril Cleverdon 著*

金 殷 植 譯**

譯者註：이 번역문은 Journal of Documentation, Vol. 26 No. 1 pp. 55~67(1970)에 수록되어 있는 "Progress in Documentation: evaluation tests of information retrieval systems"를 옮긴 것이다.

지금부터 約 30年 以前의 도큐멘테이션分野에서 골짜기 문제가 되었던 것은 分類目錄과 알파벳順 主題名目錄의 長點에 관한 比較研究였는데 이에 대한 記述이나 討論은 아무런 論證이 없었기 때문에 識者들에 의해 많은 理論的 論爭이나 見解가 뒤따랐었다. 1950年代 初, Taube, Mooers 등의 學者들이 새로운 索引作成法을 提案했을 때 固陋한 學者들은 過去의 論爭을 再現했다. 1955年 American Documentation誌의 論說에서 Perry는 다음과 같이 그의 見解를 表明하였다.

「도큐멘테이션 시스템에 대한 試驗 結果의 評價만이 効率判定의 要素가 된다. 分類에 대한 因襲的인 觀念에서 惹起되는 陳腐한 論難이 未來의 機械化시스템에 있어서 는 거듭되지 않기를 바란다. 그러기 위해서는 도큐멘테이션 시스템의 効用성을 認定하고, 必要성과 經費에 있어서의 시스템의 利得에 대하여 論爭을 止揚한 理性에 의한 知的 判斷이 무엇보다 切實히 必要하다. 將次 到來할 機械化시스템은 우리가 그것을 기꺼이 받아들이고 또 우리들 自身이 딱딱한 頭腦訓練을 쌓는다면 機械의 束縛에서 解放될 수 있다.」

事實 美國과 英國에서는 이미 그 첫 試圖이 이루어졌다. 1953年 Mortimer Taube 博士가 設立한 Documentation Inc.(D.I.)에서는 概念整合索引(Coordinate Indexing)에 의한 유니텀시스템(Uniterm system)(當時 새로운 것이었음)과 Armed Services Technical Information Agency(ASTIA)가 開發한 主題名標目(Subject heading)에 의한 在來式 目錄의 性能을 比較研究했는데, 이 試驗에서는 D.I.에서 件當 平均 8個의 유니

텀으로 索引한 15,000件의 文獻을 使用하여 ASTIA가 利用者들로부터 의뢰받은 98種의 調査設問에 대하여 유니텀索引 및 ASTIA의 主題名標目別目錄(Subject heading catalogue)으로써 各各 探索하였다. 各 그룹에서는 探索된 output을 分析하고 어떤 文獻이 가장 適合한 것인가를 調査하였었는데 여기서 問題가 된 것은 適合性 決定(Relevance decisions)에 관한 것이었다. 兩 그룹의 深刻한 意見對立 때문에 全般的인 公式報告는 發表되지 않았으며 1956년에야 簡略한 計算結果⁽¹⁾가 發表되었는데 이것에 의하면, D.I. 및 ASTIA는 相互共通된 適合文獻 580件을 檢索하였다. 또 相對에서 適合하다고 認定하는 情報를 D.I.에서는 492件, ASTIA에서는 319件씩 各己 追加檢索하였다. 結局 ASTIA는 1089件의, D.I.는 488件의 不適合情報를 各各 檢索한 셈이 된다. 그러나 이와 같은 數字的인 統計는 한낱 歷史的인 關心事일뿐 其實 重要的 것은 그후 10年동안에 評價의 方法論에 미친 影響이라고 말할 수 있다. 만일 探索結果를 利用者가 받아들이고 利用者의 切半이 適合判定을 내렸다면, 그로부터 16年後에 있었던 MEDLARS의 評價까지 詳細한 分析을 위한 充分한 데이터가 나왔을 것이다. 適合性 決定에 대한 兩그룹간의 엇갈린 主張때문에 이 試驗이 情報檢索시스템(I.R. System)의 設計와 運營에 관한 先進된 基本事例로 될 수 없었다.

英國에서는 Thorne과 Cleverdon이 유니텀시스템에 관한 몇 가지 試驗⁽²⁾을 했는데, 이것이 歷史的 關心의 對象이 된 데에는 다음의 두 가지 理由가 있다. Mooers는 그 理由를 다음과 같이 說明하였다⁽³⁾. 即 그는 이미 以前에 그의 글에서 情報檢索(Information retrieval)이란 用語를 使用했었지만 이 말이 一般化된 것은 이 報告書에서 使用한 뒤부터라고 했다. 또 한 가지는 D.I.試驗에서 惹起되었던 適合性 決定에 있어서의 諸 問題에 관한 經驗과 더불어 이 太初의 試驗이 Cranfield I의 設計에 全적으로 影響을 주었다는 것이다.

이 分野에 있어서의 後續研究들에 대하여 考察하기 前에 우선 情報檢索시스템의 評價試驗의 定義의 限界에

* Cranfield Institute of Technology
** KORSTIC 情報處理部 物理主任.

관해서言及해야겠다. 그러기 위해서는 完全한 情報檢索시스템의 實態 또는 擬態環境下에서 이루어져야 하는데, 이 말은 반드시 完全한 시스템을 試驗해야 된다는 것은 아니고 [完全한 環境下에서 이루어지기만 한다면 가령 索引言語와 같은 特殊한 細部시스템(Sub-system)에 대하여 評價해도 좋다. 試驗環境 即 外的 要因中에 探索質問이나 適合性 決定등은 包含시키지 않았기 때문에 Kessler⁽⁴⁾의 研究나 索引者間 一致性(Inter-indexer Consistency)에 관한 試驗 등은 여기서 除外한다. 格外例를 들면, M.Henderson⁽⁶⁾의 "Evaluation of information systems: a selected bibliography"中的 論文들도 除外된다. 이와 같은 限界가 專斷的인 感이 없지는 않으나 이들 除外對象의 研究들(1948~1956年 사이에 發表된 Fairthorn의 論文은 例外)이 情報檢索시스템에 관한 우리의 知識에 대하여 크게 貢獻하지 못했다는 點에서 正當化될 수 있다.

Cranfield I⁽⁷⁾는 4種類의 索引言語의 性能比較를 위한 試圖로서 1,200件의 探索質問에 대한 18,000種의 文獻을 基礎로 했다. 이것은 後續研究들에서 認證된 바 많은 興味있는 結果를 주었는데, 中 中 가장 重要한 것은 評價試驗의 概念에 대한 發表이다. 이 試驗結果의 報告가 있는 後 많은 研究論文들이 發表되었는데 모두가 批判的인 것이었다⁽⁸⁾. 그러나 同時期의 Swanson⁽⁹⁾의 研究報告는 事實上 度外視되었지만 그는 그의 試驗에서 核物理分野의 100件의 記事를 對象으로 해서 매우 興味있는 調査를 하였다. 即 50件의 探索質問을 利用하여 從來의 主題名索引(Subject indexing)과 電子計算機를 利用한 完全文章索引(Computer searching of the complete text)의 效率를 比較檢討한 結果, 2시스템 모두 效率가 낮으나 主題索引보다 電子計算機에 의한 完全文章에 대한 索引이 效率가 높다는 事實이 發見되었다. 當時 이 驚異的인 事實에 대한 論理的 解明은 不可能했지만 그때까지 一般적으로 等閑視되었던 이 Swanson의 研究가 數年後 情報檢索시스템의 重要한 一局面을 認識케 해주었다. 1960년에 Schuller⁽¹⁰⁾에 의하여 荷蘭드國防省에서의 유니텀과 UDC에 대한 比較試驗報告가 있었는데 이때 Cranfield I에 의하면 이 2시스템이 全體的인 性能에 있어서 重要한 差異點이 없다는 것이 確認되었다.

그後 數年동안에 評價試驗에 관한 많은 研究報告가 있었는데, Cranfield Institute of Technology에서 創案한 多面索引(페시트索引, Facet index)과의 比較檢討를 內容으로 한 金屬工學關係文獻의 W.R.U. 索引의 試驗⁽¹¹⁾은 再現性(Recall)과 精確性(Precision)間의 反比例關係(Inverse relationship)를 論證한 첫 케이스가 되

었다. 特記할만한 것은 이 試驗에서 로울(Roles) 또는 相關要素(Relations)의 效率性에 있어서는 深刻한 疑問點이 있음이 確認되어 이것이 情報檢索의 諸 問題點에 대한 하나의 解決策으로 期待되기에 이르렀으며, Lancaster⁽¹²⁾, Sinnett⁽¹³⁾, Montague⁽¹⁴⁾ 및 Hyslop⁽¹⁵⁾ 등에 의한 各各 相異한 主題分野에 대한 試驗에서 로울에 관한 이와 같은 事實이 確認되었다.

이즈음 서로 다른 目的下에 많은 評價試驗들이 進行되었는데 中에는 特別한 應用에 대한 目的意識없이 情報檢索시스템에 관한 知識의 擴張만을 위하여 研究開發만을 目的으로 한 試驗이 많았다. 이것들은 研究를 위한 試驗(Research test)이라고 할 수 있겠는데 主로 새로운 시스템의 設計組織을 目標로 하여 特定環境下에서 이루어졌다. 中에는 造船技術圖書館局(Bureau of ships technical library)의 要請에 의한 것(12)도 있었는데, 이것은 새로운 시스템의 設計에 있어서 하나의 勸獎事例로 되었다. Montague⁽¹⁴⁾ 등에 의한 現場시스템(Operational system)과 直結되는 試驗들이 續出된 것은 그後의 일이었다. 評價試驗을 이와 같은 範疇속에서 考察하는 것은 앞으로의 論述에 있어서의 便利를 위한 것이다.

1964년에 와서 "SMART System"의 첫 試驗成果가 實用化되었는데, 이 作業은 처음에는 하버드大學에서, 中에는 코오넬大學에서 Salton의 指導아래 이루어졌으며, 오늘에 이르기까지 研究作業에 있어서 支配的인 位置를 차지하는 事例가 되고 있다. 그 內容은 "情報의 蓄積과 檢索"이란 題下의 15篇의 報告書中에 包含되어 있으며, Salton의 著書⁽¹⁷⁾의 바탕을 이루고 있다. SMART란 自然言語(Natural language)로 表現된 文獻 및 探索要求를 調整할 수 있는 自動主題處理시스템(Automatic text processing system)으로서 探索要求에 대한 最近接文獻의 順位表를 만들어준다. 이 시스템의 長點은 많은 相異한 制御分析過程에 의하여 文獻과 探索要求間의 鑑定者의 役割을 하는 것이다. Salton은 頭腦的 判斷과 自動化시스템과의 巧妙한 組合에 의하여 多量의 蒐集文獻을 對象으로 한 研究를 통하여 機械化시스템의 性能改善을 위한 가장 效率的인 方法을 開發할 수 있었다. SMART프로젝트에 대한 批判도 있었는데 그 內容은, 現場시스템에 應用될 수 있는 適格性을 論證할 수 없고, 또 電子計算機處理는 매우 複雜하기 때문에 現場規模의 文獻探索에 利用하기에는 經費가 너무 많이 든다는 것이다. 그러나 이것은 그리 重要한 理由가 못된다고 본다. 왜냐하면, 機械化 情報檢索시스템의 開發은, 將來의 더욱 強力한 電子計算機 開發에 의하여 現場에 應用될 수 있는 經濟的인 水準까지 經費를 低減시킬 수

있을 것이라는 假定위에 그 基礎를 두고 있기 때문이다. SMART프로젝트의 重要性은 情報檢索시스템의 性能에 대하여 考慮해야 할 限界點과 最適性能을 위하여 取하여야 할 方案을 提示한 데 있다.

1962~1968年 사이에 SMART프로젝트는 Cranfield II 프로젝트와 서로 密接하게 併行하여 이루어졌다. Cranfield I에서는 4種의 索引言語가 試驗의 結果 根本적으로 매우 類似하다는 事實이 밝혀졌다. 알파벳順 主題名 索引, 國際十進分類 및 概念整合索引시스템間的 差異點은 實際보다는 더 明確했는데 이것은 각 索引言語가 數個의 別途디바이스를 가지고 있기 때문이다. 모든 索引言語디바이스는 다음 2가지 中의 한 가지 目的을 가지고 있다. 即 檢索된 適合文獻의 數를 增加시켜 再現率(Recall ratio)을 높이든가 또는 相對적으로 非適合文獻의 數를 減少시켜 精確率(Precision ratio)을 높인다. Cranfield II 프로젝트의 目的은 이 디바이스(Devices)들을 各各 獨立적으로 또는 相互結合하여 試驗하는 것이다.⁽¹⁸⁾ 이 프로젝트에서는 29個의 索引言語에 대하여 檢討했는데 主로 航空力學分野의 221件의 探索質問에 대한 1,400種의 文獻을 使用하였다. 이 프로젝트에서 얻은 成果는 驚異적인 것으로, 自然言語(即 文章中의 낱말)를 使用하는 單一用語를 바탕으로 한 索引言語가 가장 效率的임이 밝혀졌다.

1963年, 케이스 웨스턴 리저브大學에, 各種 實驗用 情報檢索시스템의 特性比較를 目的으로 하는 比較시스템 研究所(Comparative system laboratory)가 設立되었다. 初期에는 主로 試驗의 方法論에 勞力을 集中하였으며, 最終報告書⁽¹⁹⁾도 既知의 事實에 대해서는 거의 言及하지 않고 方法論에 대해서만 論述하였다. 實際試驗에 있어서는 試驗資料로서 熱帶病에 관한 600種의 文獻을 使用, 25人의 利用者로부터 接受한 124件의 質問을 이미 試驗한 5種의 索引言語로 索引化하였다. 이 프로젝트에는 50人에 達하는 人員이 隨時로 動員되어 進行되었으나 情報檢索시스템에 관한 새로운 事實은 얻지 못했다. 그러나 過去의 試驗에서 얻은 여러가지 事實들 例를 들면 網羅性의 效果, 再現성과 精確性間的 反比例關係 및 檢索性能向上을 위한 로울導入의 失敗 등을 確認 하는데는 크게 도움이 되었으며, 結局 情報檢索시스템의 效率는 期待以下임이 確實視되었다.

같은 期間에, Farradane은 相關索引(Relational indexing)의 效率에 관한 一連의 試驗을 했는데⁽²⁰⁾, 여기서 그는 Sugar Industry Abstracts에서 1,085件의 文獻을 選定하여 索引化하고 127件의 質問을 利用하였다. 이 試驗의 結果는 試驗의 範圍制限에 의하여 限定된 感이 없지 않으나 무엇보다 重要한 것은 로울이 質問의 77%

에 대한 Output에 대해서 아무런 效果도 없었다는 事實로서, 따라서 이와 같은 種類의 디바이스의 低價値性이 再確認되었다.

Cranfield II프로젝트에서 얻은 資料의 一部를 使用하여 Karen Sparck Jones博士는 劍橋大學에서 重要한 一連의 實驗에 着手했는데, 이것은 劍橋大學의 語學 研究所(Cambridge Language Research Unit)⁽²¹⁾에서 遂行한 研究의 繼續으로 大型電子計算機와 檢索프로그램의 成果로 말미암아 完全히 새로운 局面으로 突入했다⁽²²⁾. Sparck Jones博士는 探索프로그램과 索引의 網羅性 및 索引言語의 精確性間的 微妙한 關係를 論證할 수 있게 되었다. 再現성과 精確性間的 反比例關係 때문에 일어나는 諸 問題의 解決을 위해서는 既存索引言語디바이스로서는 不可能하다는 事實이 證明되었으며, 따라서 既存시스템에 있어서는 檢索된 總適合文獻數의 增加가 檢索된 非適合文獻數의 不均衡的인 增加를 가져와서 檢索成績은 理想值에 훨씬 未達된다는 事實이 全評價試驗을 通하여 確認되었다. 分類技術이나 디소오러스만으로는 이 問題解決의 幾微가 보이지 않으나 Sparck Jones博士의 研究가 이 問題解決의 方法을 提示해 줄 것이라는 疑心이 가지 않는다.

Keen은 各種 索引言語間的 比較를 內容의 一部로 하는 一連의 試驗을 展開했는데 從前의 試驗에 있어서 多少 看過된 바 있는 探索戰略(Search strategy)의 多變化에 特別히 注意를 集中시켰다.

시스템의 開發을 위한 試驗에 대하여 考察해보면, INSPEC프로그램에 있어서 Aitchison은 情報選擇提供(SDI) 및 溯及調査(Retrospective searching)의 電子計算機化의 最適成就를 위한 研究에 앞장섰다. 지금도 進行되고 있는 SDI프로젝트⁽²³⁾에는 500人 以上の 利用者가 對象으로 되어 있는데, 이들은 每週 定期的으로 電子工學分野에 있어서 그들이 願하는 引用參考文獻(Basis citations)을 받고 있으며, 이 서비스를 위하여 每週 300種의 文獻을 索引한다. Science Abstracts에 包含된 主題의 全領域을 커버할 수 있는 如他시스템을 위한 모델로도 利用이 可能한 實現性있는 SDI시스템의 開發과는 別途로 利用者가 1年間 Output을 서비스받은 後 情報蒐集節次를 變更했는지를 評價한다. 또한 範圍를 좁혀서, 電子工學分野의 1,000種의 文獻과 100件의 質問을 綜合하여 INSPEC을 위한 여러가지 可能한 索引用語에 대하여도 研究를 했다. 또 索引者가 選定한 制御用語(Controlled term) 및 Science Abstracts中의 主題記入(Subject entries)뿐만 아니라 題目, 抄錄 및 索引者가 選擇한 用語에 使用한 自然言語와 比較檢討하였다. 그 結果⁽²⁴⁾ Cranfield II프로젝트에서 밝혀진 網羅的 索引의

重要性和自然言語의 期待以上の 效果가 立證되었다.

노팅검大學에서는 化學會情報管理研究會(Cheical Society Research Unit in Information Documentation and Retrieval)에서 各種 形態의 化學抄錄을 對象으로 하여 一連의 評價試驗에 들어 갔는데 SDI서비스의 確立과 評價에 集中勞力했을 뿐만 아니라 索引代用物(Indexing surrogates)^(25, 26)의 使用效果와 같은 附次的인 問題에 대해서도 研究했다.

워싱턴에 있는 國立醫藥博物館(National Library of Medicine)⁽²⁷⁾에서 實施한 바 있는 MEDLARS의 現場 시스템에 대한 評價試驗에 있어서는 302件의 實際의 探索質問을 對象으로 했는데 그 結果 精確率이 50%로 比較的 높은데 比하여 再現率은 不過 58%라고 밝혀졌다. 또 잘못 檢索된 非適合文獻 또는 檢索되지 않은 適合文獻 등 約 4,000項目에 達하는 失敗에 대하여 分析한 結果, 誤謬의 가장 核心的인 部分은 利用者의 實際要求에 대하여 滿足할만한 探索戰略을 導入하지 못한 點이며, 其他는 Cranfield I에서 實施했던 類似한 分析에서 指摘된 것과 같았다.

迦及調查시스템에 대한 試驗이 進行되는 동안 一部에서는 信賴度가 매우 높은 데이터를 內包한 詳細한 報告書를 내놓았는데 이 研究에서는 SDI시스템을 強調하고 있다. 情報蓄積檢索의 機械化시스템에 관한 第2次國際會議에서, 쉘研究所의 Dammers⁽²⁸⁾, 유네버研究所의 Rowland⁽²⁹⁾ 및 앨버타大學의 Heaps⁽³⁰⁾ 등은 現場SDI 시스템의 評價에 관한 論文을 發表했었다. 그러나 現在 進行中인 가장 重要한 試驗中의 하나는 옥스포드大學에서 P. Leggate博士가 推進하고 있는 것으로 여겨지는데, 이 試驗에서는 太初의 D.I.—ASTIA試驗 以來 처음으로 現場시스템間의 性能比較를 試圖하였다. 約 60個그룹의 利用者들이 노팅검에 있는 Chemical Titles Services, 필라델피아의 Institute of Scientific Information Service 및 Scientific Documentation Center로부터 同時에 Output를 配布받고 있는데, 初期의 安定期間이 지난 후 情報의 適合性, 利用하게 된 日字 및 其他 因子들에 관한 이 3시스템間의 效率比較를 위하여 各 利用者그룹에 傳達된 引用文獻에 대한 詳細한 記錄이 相互協助下에 數個月間 保存된다⁽³¹⁾.

論 評

本稿는 過去 20年동안에 研究報告된 評價試驗中에서 重要한 것들에 대해서 概觀한 것이다. 여기에는 美國特許局의 Bryant 諸氏⁽²³⁾ 및 해리다이어먼드研究試驗所⁽²⁴⁾의 Altmann 등의 研究內容도 包含되었다. 事實 評價試驗의 定義에 관한 過去의 文獻만해도 하나의 解說記事

속에 網羅시키기에는 너무나 老大하다. 이것은 綜合的인 目錄을 必要로 하는 讀者들이 반드시 參考해야 하는 Annual Review of Documentation誌⁽³⁵⁾를 構成하는 標準章의 主題로 選擇되곤 하는데, 著者들에 의한 論評을 보면 그 定義가 一般的으로 매우 曖昧하다. 評價試驗의 方法論이 完璧하다고 評하는 것은 잘못된 生覺일런지 모른다. 그러나 애당초부터 試驗設計 自體에 缺陷이 있다는 根據없는 斷定을 하고 試驗結果의 妥當性을 否認하는 批評家들이 적지는 않았으나, 더러는 이들 結果가 後續研究에 의하여 確證되는 등, 方法論은 그 幼年期로부터 計劃된 시스템의 開發試驗은 勿論 MEDLARS와 같은 大規模의 現場시스템에 대한 評價試驗을 設計할 수 있는 實質的인 段階에까지 發展하였다. 이같은 事實은 情報檢索시스템에 관한 知的 水準의 向上에 一部起因한다고도 볼 수 있으며, 하나의 시스템의 可能的인 實行과 性能을 理解하고 展望할 수 있기 때문에 全體시스템 또는 細部시스템의 評價試驗을 쉽게 設計하고 實行할 수 있게 되었다. 한편 研究試驗은 漸次 어렵게 되어 가고 있다. 即 研究의 主題로 되는 索引 및 索引言語에 있어서는 作業이 매우 複雜하며, 將次 電子計算機의 도움이나 SMART 또는 Needham에서 쓰고 있는 프로그램 및 分析에 있어서의 相當한 勞力 등이 없이는 重要한 結果를 豫測하기 어려운 問題點을 內包하고 있다. 索引과 索引言語에 대하여 研究를 集中하는 理由는 그것이 試圖해 보기에 比較的 單純하기 때문이다. 研究作業에서 거의 考慮되지 않았던 어려운 點은 시스템의 施行結果와 關係있는 것으로서, 實行과 그에 대한 利用者의 反應을 커버하는 것이라고 말할 수 있다. 이分野에 있어서의 研究프로젝트는 實際時間에 있어서의 利用者와의 相互作用의 必要性 때문에 매우 困難하다.

性能의 測定(Performance measures)은 언제나 論爭의 對象이 되어 왔다. 再現率과 精確率은 가장 널리 使用되어온 測定手段으로서 1957년에 이미 Perry와 Kent에 의하여 提案된 것인데,⁽³⁶⁾ 一般的으로 "Cranfield Measures"라고 알려져 있다. 最近에는 Swets⁽³⁷⁾, Brooks⁽³⁸⁾ 및 Robertson⁽³⁹⁾ 등이 내놓은 測定方法에 대하여 數學的인 接近을 試圖한 論文들이 많이 나오고 있다. 그러나 여기서 가장 重要한 것은 研究試驗과 現場試驗間의 區別이다. 시스템의 實行成果, 敏感度 또는 標準化測定 등은 研究試驗에 있어서의 特定目的에는 適合할런지 모르나 現場시스템의 試驗에는 쓸모가 없다. 現場시스템의 試驗에 있어서는 精確率이 測定對象要素中의 하나가 되어야 하는데, 多幸이 이것은 比較的 單純한 測定이어서 利用자가 받은 文獻의 適合性을 分析함으로써 利用者 自身이 決定할 수 있다. 그러나 現場시스템에 있어

서 再現率을 確認하기는 좀 어렵다. 그렇지만 MEDLARS試驗 以前에 이를 위한 合理的인 近似解法이 이미 開發되었다. 아무래도 單純한 再現率은 現場의 狀況에 있어서 그 價値性에 限界가 있는데 이와 같은 事實은 비록 그 眞面貌는 아니지만 시스템自體가 얼마나 不完全한가를 말해준다. 遡及調査와 SDI시스템에 있어서 100%의 再現率을 要求하는 利用者가 있을 수도 있지만 大概의 경우 그 以下로도 滿足한다. 또 遡及調査에서 利用者A는 20件의 適合文獻이면 足하다고 미리 말해두는 수가 있을 것이고, SDI의 경우 利用者는 適當 平均 5件의 適合引用文獻을 要求하는 수도 있다. 이와 같이 利用者들에 의하여 指定된 要求에 대한 시스템의 性能을 評價함에 있어서 相對再現率(Relative recall ratio)이란 말을 使用한다. 即 어떤 시스템이 利用者A의 要求에 대하여 20件의 適合文獻을 檢索했다면 檢索되지 않은 適合文獻이 많더라도 相對再現率은 100%이다. 現場試驗에 要求되는 再現率의 其他變換에 대해서는 參考文獻40에 잘 記述되어 있다.

評價試驗을 통하여 얻은 經驗과 知識에 의하여, 情報檢索시스템의 設計와 實行을 위한 가장 좋은 單一方法은 없다는 것이 漸次 明白해 짐을 알 수 있다. 주어진 特定의 性能要求에 대해서는 여러가지 選擇的 解決方案이 있다, 예를 들어, 再現率 65%에 精確率 25%의 要求가 있다고 假想해 보자. 첫째 適用範圍와 再現性 사이에는 隨意性이 存在한다. 即 주어진 主題에 관한 全世界의 文獻의 95%를 커버할 수 있다면 68%의 再現率만으로도 要求條件인 65%의 再現率을 만족할 수 있다. 그러나 만일 適用範圍가 不過 75%라면 同一한 結果를 얻기 위해서는 87%의 再現率이 必要해진다. 要求精確率을 얻기 위해서도 赤是 索引言語, 探索戰略 또는 Output의 整理 등에 있어서 隨意性이 存在한다.

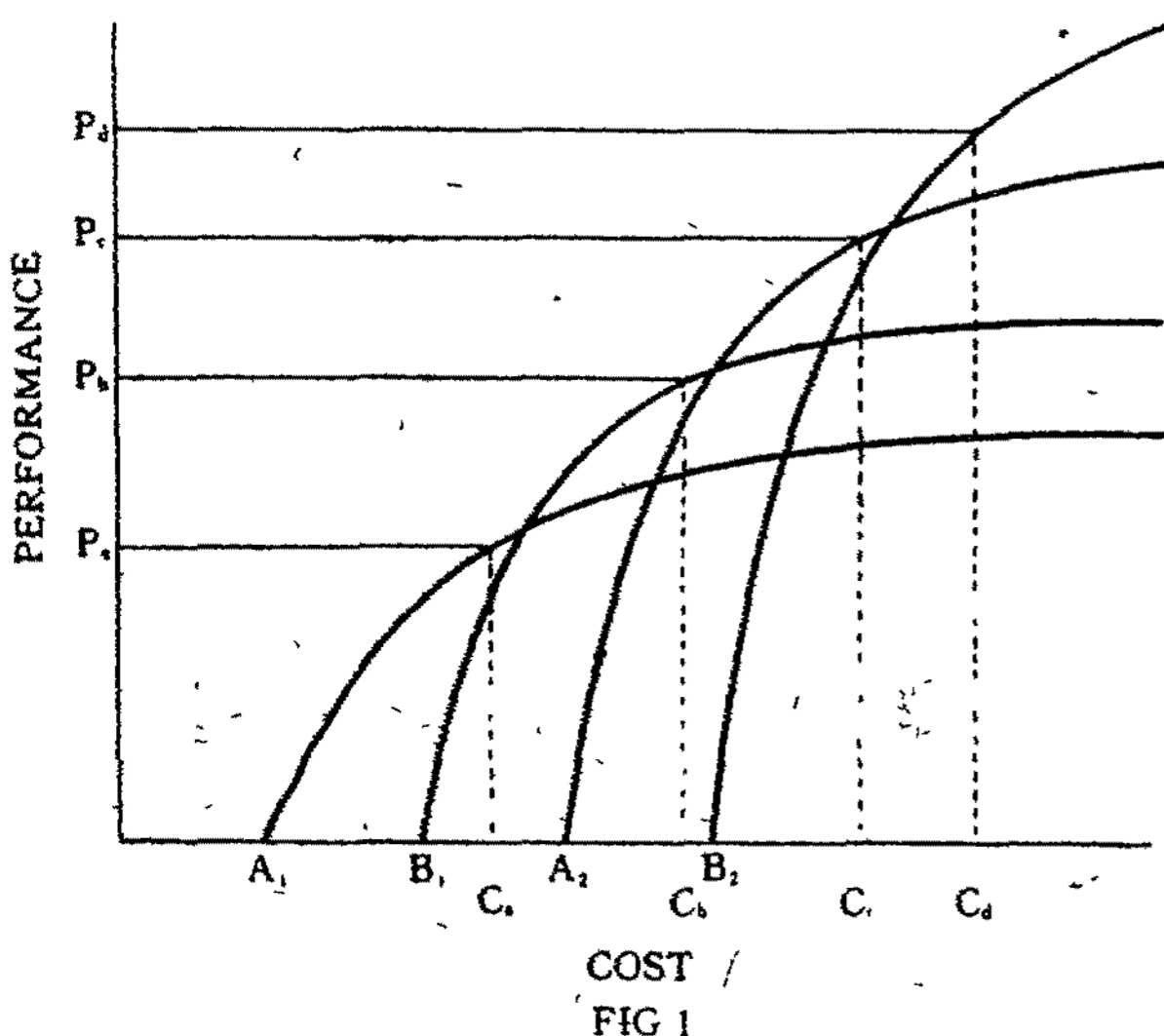


FIG 1

주어진 狀態에 있어서 同一한 目標에 到達할 수 있는 여러가지 方案의 選擇에 直面했을 때는, 經費效率(Cost effectiveness)의 바탕위에서만 論理的인 決定을 내릴 수 있다. 情報檢索시스템에 있어서 經費效率의 實例를 說明하기 위해서 Fig. 1과 같은 單純한 境遇를 生覺해 보자. A1은 100種의 雜誌를 適用範圍로 하는 시스템이고 B1은 200種의 雜誌를 適用範圍로 하는 同一한 시스템으로서 全然 索引化하지 않고 input는 雜誌記事의 題目中의 用語를 使用했다. A2와 B2는 雜誌類는 A1, B1과 同一하나 記事는 網羅的 索引을 하였다. 點 A1, B1, A2, B2는 각 시스템의 資料蒐集 및 input의 基本經費를 나타낸다. 여기에 各種 探索經費(Search cost)를 追加하여 性能-經費曲線(Performance cost curve)이 作成된다. 이 圖表에 의하면 Pa의 性能이 要求될 때는 A1시스템을 利用해야 된다는 것을 알 수 있다. 따라서 Pb에 대해서는 B1시스템이 가장 效率的이며 Pc에 대해서는 A2시스템, Pd에 대해서는 B2시스템이 適合하다는 것을 알 수 있다. 다시 말하면 經費를 Ca로 固定시킬 때는 A1시스템을 適用해야 된다.

위의 例는 하나의 假定이지만, 情報檢索시스템의 評價에 있어서 先行되어야 할 接近方法을 提示해 준다. 어느 限度內에서는 여러가지 方法에 의하여 必要한 成果를 얻는 것이 理論적으로 可能하나 最終 決定은 經費에 의하여 左右된다. 여기서 必要해지는 것은 새로운 測定方法의 導入인데 MEDLARS試驗(40)의 結果, 한가지 方法이 開發되었다. 이 方法은 適合文獻의 檢索經費와 關係가 있으며 다음과 같이 數式으로 表示할 수 있다.

$$Cr = \frac{C + F \left\{ \left(D \times \frac{R}{100} \right) \left(\frac{100}{P} \right) - \left(D \times \frac{R}{100} \right) \right\}}{D \times \frac{R}{100}} = \frac{100C}{DR} + \frac{100F}{F} - F$$

여기서 C=單一探索의 시스템經費

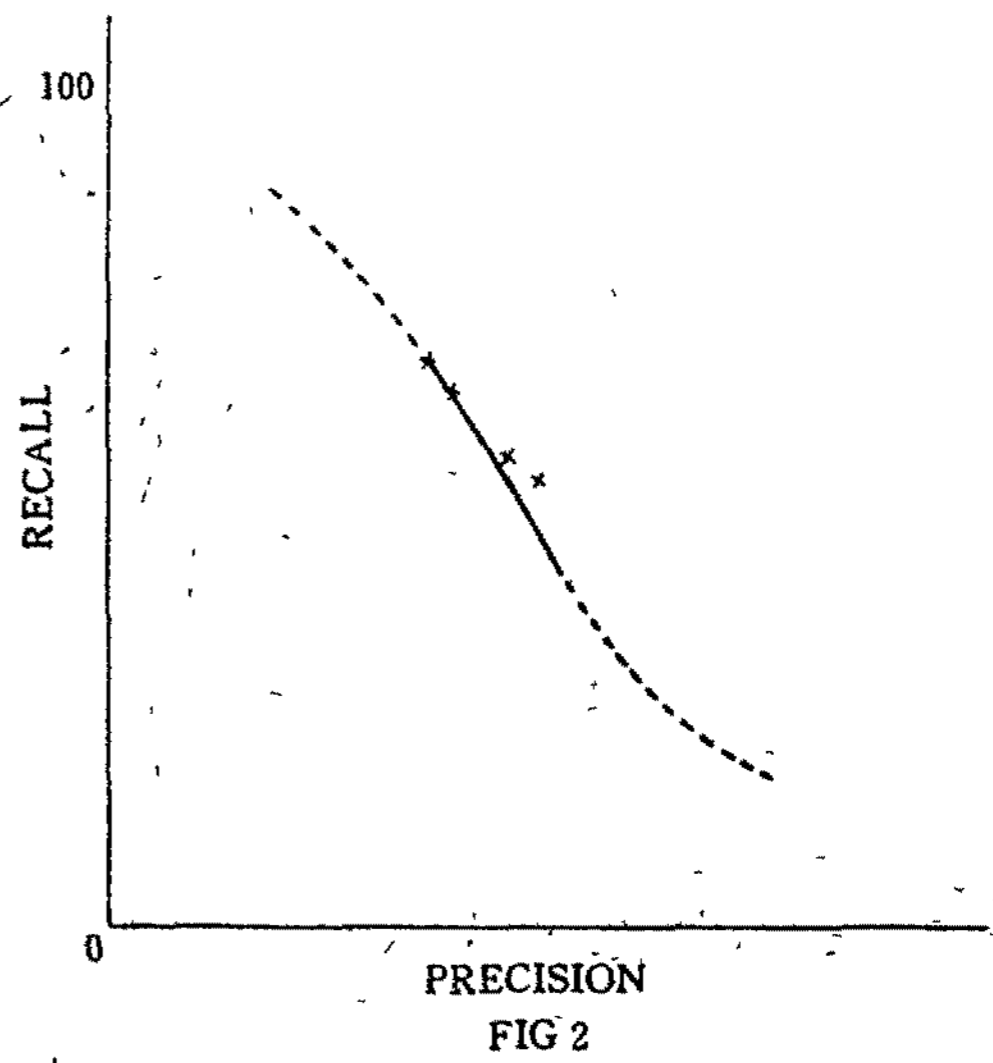


FIG 2

D=適合文獻 豫想件數

R=再現率

P=精確率

F=非適合文獻에 대한 費用

이 式의 재미있는 應用例는 MEDLARS의 評價에서 찾아볼 수 있다. 即 5個所의 MEDLARS Center에서 各各 相異한 探索方案을 採擇했는데 結果는 다음과 같았다.

	再現率(%)	精確率(%)
Center A	69.2	40.7
" B	64.6	43.2
" C	57.9	50.9
" D	55.5	55.6
" E	43.3	57.2

即, 再現率이 低下함에 따라서 精確率은 增加한다는 것을 알 수 있다. 그리고 Fig.2에 의하면 各點은 아주 定常性能曲線을 形成함을 알 수 있다. 따라서 Center A와 Center B間의 性能上의 優劣은 말할 수 없다. 그러나 시스템의 運營費를 安後 經費効率測定法을 適用해 보면 Center A가 가장 效率的이며 適合文獻當 檢索費用이 最低이다.

	Cr(\$)
Center A	1.34
" B	1.39
" C	1.49
" D	1.57
" E	1.99

비록 現場시스템의 試驗이 實行不可能할 程度로 複雜하지는 않다해도, 性能値를 測定하고 缺陷을 分析하기 위해서 시스템間을 直接 比較하는 것은 아직은 不可能하다. 例를 들어, 主題分野가 化學인 하나의 시스템이 80% 再現率과 40% 精確度의 性能을 가졌다고 해서, 60% 再現率과 40% 精確度를 가진 主題分野가 物理인 하나의 시스템보다 優秀하다고 볼 수는 없다. 아직은, 主題分野, 利用者그룹의 種類에 따른 相異한 探索要求 및 操業環境 등의 變化要因에 대한 許容限度를 決定할 수 없을 뿐만 아니라 經費效率의 比較도 不可能하다. 또 現在로는 情報檢索시스템의 經費算出에 대한 아무런 標準도 없기 때문에 그것은 더욱 不可能하다.

向後 10年內에는 開發될 것으로 믿어지는 3가지의 路線이 있는데, 其1은 遡及調査에 대한 連續品質管理技術의 開發이다. SDI시스템에 있어서는 이와 같은 品質管理가 可能하여 探索操作의 最適化를 期할 수 있다. 遡及調査에 있어서는 이것이 不可能하기 때문에, 利用者의 要求를 相當히 滿足할 만한 性能을 保障할 수 있는

比較的 簡單한 管理節次가 반드시 開發되어야 한다. 其2는 이미 記述한 바 있는 시스템間의 比較에 관한 것이다. Leggate의 研究는 이를 위한 첫 試圖이다. 그러나 그의 方法은, 年間注文量으로써 經費를 算出할 수 있는, 同一主題分野를 커버하는 SDI시스템을 基礎로 해서 그것을 單純化한 것에 不過하였다.

其3은 가장 重要하고 또 어려운 問題이다. 모든 情報檢索시스템은 한 個人이나 機關에 대한 所望스러운 行爲라고 보는 것이 아마 옳을 것이다. 空白狀態의 情報檢索시스템은 存在하지 않는다. 可能한 方法에 의하여 情報의 探索은 繼續되고 시스템은 間斷없이 設置되며 새로운 서어비스方案이 樹立되고 있다. 大部分의 시스템은 試驗을 거치지 않고 必要以下の 效率로써 運營되고 있다. 經費效率에 있어서도 마찬가지다. 왜냐하면, 大概의 경우 經費效率에 대해서는 크게 關心을 기울이지 않는다. 많은 提案된 試驗方案의 未備함과 不適當함을 勘案할 때 情報關聯人(Information people)들이 그들 自身の 文獻을 吟味하지 않는다는 것은 明白하나 評價에 관해서만은 變化의 希望的인 조짐이 보인다. 이러한 局面에 處해서 收益性的 바탕 위에서 情報檢索서어비스를 正當化시킬 수는 없다. 2重研究의 經費가 不安의 要素가 되지만 無意味한 것으로 치더라도 우리는 情報라는 것이 "必要不可缺(Good thing)"한 것이라는 事實만이라도 依存하지 않을 수 없다. 한 個人이 他 情報源으로부터 完全に 孤立되었다고 假定하고, 個人 自身の 利用을 目的으로 索引시스템을 樹立했을 때 對經費收益을 圖模하기란 그다지 어려운 問題가 아니다. 그러나 이것을 많은 사람들이 利用할 수 있는 專門情報서어비스시스템 樹立의 收益性 決定의 段階로까지 發展시키기 위해서는 많은 研究가 뒤따라야 하며 情報서어비스에 이바지해야 할 勞力에 대한 成果를 보기 위해서도 달리 方法이 없을 것 같다.

<引用 文 獻>

1. Gull, C.D. "Seven years of work on the organization of materials in the special library". American Documentation, vol.7, 1956, pp.320~329.
2. Cleverdon, C.W., Thorne, R.G. "A brief experiment with the uniterm system of coordinate indexing for the cataloguing of structural data". R.A.E. Library Memo 7, 1954.
3. Mooers, C.N. "Comments on the paper by Bar-Hillel." American Documentation, vol.8, 1957, pp. 114~116.
4. Kessler, M.M. "Comparison of the results of bibliographic coupling and analytic subject indexing. American Documentation, vol. 16, 1965, pp. 223~233.
5. Jacoby, J., Slamecka, V. "Indexer Consistency under Minimal Conditions." Documentation Inc. November, 1962
6. Henderson, M.M. "Evaluation of information systems: a selected bibliography with informative abstracts. N.B.S.

- Tech. Note 297, 1967.
7. Cleverdon, C.W. "Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems". 2 vols. Cranfield, 1962.
 8. Swanson, D.R. "The evidence underlying the Cranfield results". *The Library Quarterly*, vol. 35, 1965, pp. 1~20.
 9. Swanson, D.R. "An experiment in automatic text searching". Final Report. Thompson Ramo Wooldridge Inc. 1960.
 10. Schuller, J.A. "Experience with indexing and retrieving by U.D.C. and Uniterms". *Aslib proceedings*, vol. 12, 1960, pp. 372~389.
 11. Aitchison, J., Cleverdon, C.W. "Report on a test of the index of metallurgical literature of Western Reserve University Cranfield, 1963.
 12. Lancaster, F.W., Johanningsmeter, W.F. "Project SH ARP: evaluation of indexing procedures and retrieval effectiveness." NAVSHIPS Report 250-210-3, 1964.
 13. Sinnet, J.D. "An evaluation of links and roles used in information retrieval". Tech. Report ML-TDR-64-152.
 14. Montague, B.E. "A analysis of the designing, installation, and operation of a coordinate indexing system using links and roles for the Plastics Dept. of the DuPont Company". *Journal of Chemical Documentation*, vol. 4, 1964, pp. 210~215.
 15. Hyslop, M.R. "Role indicators and their use in information searching; relationship of ASM and EJC systems". *Proc. Amer. Doc. Inst.*, vol. 1, 1964. pp. 99~107.
 16. Salton, G., et al. "Information Storage and Retrieval". I.S.R. Reports I-15, 1962~1969.
 17. Salton, G. "Automatic information organization and retrieval. New York, McGraw Hill, 1968.
 18. Cleverdon, C.W., Mills, J., Keen, E.M. "Factors determining the performance of indexing systems. 2 vols, Cranfield, 1966.
 19. Comparative Systems Laboratory. "An inquiry into testing of information retrieval systems". 3vols. Case Western Reserve University, 1968.
 20. Paradane, J. "Report on research on information retrieval by relational indexing, 1968 (To be published)
 21. Needham, R.M., Sparck Jones, K. "Keywords and clumps: recent work on information retrieval at the Cambridge Language Research Unit". *Journal of Documentation*, vol. 20, no. 1, 1964, pp. 5~15.
 22. Sparck Jones, K., Jackson, D.M. "The use of automatically obtained classifications for information retrieval". *Information Storage and Retrieval*, 1970.
 23. Clague, P. "SDI investigation in electronics". in: *Computer based information retrieval systems*. Bingley, 1969.
 24. Aitchison, T.M. "INSPEC Comparative evaluation of index languages. (To be published)
 25. Kent, A.K. "Problems of remote processing of magnetic tape information files". *Aslib proceedings*, vol. 20, 1968 pp. 502~508.
 26. Veal, D.C. "Report on evaluation of an experimental computer based Current awareness service for chemists". (To be published)
 27. Lancaster, F.W. "Evaluation of the MEDLAS demand search service". Washington, National Library of Medicine, 1968.
 28. Dammers, H.F. "Computer use in information and data handling: an appraisal of its economic aspects". (To be published in *Information Storage and Retrieval*)
 29. Rowlands, D.G. "The Unilever Research SDI system". (To be Published in *Information Storage and Retrieval*)
 30. Heaps, H.S., Thiel, L.S. "Optimum procedures for economic information retrieval". (To be published in *Information Storage and Retrieval*)
 31. Leggate, P. "Evaluation of current awareness services." Experimental Information Unit, Oxford. (Unpublished working paper)
 32. Bryant, E.C., et al. "Analysis of an indexing and retrieval experiment for the organometallic file of the U.S. Patent Office." Westat Research Analysts Inc. Report WAR-PD-10, 1963.
 33. Altmann, B. "A multiple testing of the ABC method". 2 vols. Harry Daimond Labs. Reports TR-1295 and TR-1296.
 34. Vaswani, P., Cameron, J. "The National Physical Laboratory experiment on mechanized indexing and retrieval. (To be Published in *Information Storage and Retrieval*.)
 35. Cuadra, C.A. (ed.) *Annual Review of Information Science and Technology*. vols. 1-4, 1966-69. New York, Wiley and Sons.
 36. Derry, J.W., Kent, A. *Documentation and information retrieval*. Western Reserve University, 1957.
 37. Swets, J.A. "Effectiveness of information retrieval methods". Report AFC RL-67-0412, 1967.
 38. Brookes, B.C. "The measures of information retrieval effectiveness proposed by Swets". *Journal of Documentation*, vol. 24, no. 1, 1968. pp. 41~54.
 39. Robertson, S.E. "The parametric description of retrieval tests". *Journal of Documentation*, vol. 25, no. 1, 1969, pp. 1~27.
 40. Cleverdon, C.W. "A cost-effectiveness measure for information retrieval systems". (To be published in *Information Storage and Retrieval*).