

情報시스템의 評價와 分析(完)

金 英 男

<前 KORSTIC 企劃室>

6. 調査方法

調査領域의 決定에 所要되는 努力과 調査結果의 選別에 必要한 人力間에는 相關關係가 있다. 즉 調査領域을 緻密하게 設定하는 데 많은 勞力을 割愛한다면 事後編輯이 필요없는, 正確率이 높은 調査結果를 기대할 수 있을 것이다.

逆으로 調査領域을 概略적으로 設定한다면 調査結果에 대한 事後選別作業(Screening Operation)을 통해 調査結果의 正確率을 提高해야 할 것이다. 調査者가 調査結果를 選別하여 情報利用者의 要求를 滿足시킬 수 있다면, 後者의 方法이 許容된 正確率下에서 再現率을 提高시킬 수 있다는 점에서 더 有利한 方法이 될 것이다. 이것은 또한 費用-效果 面에서도 良好한 方法이다.

情報選別者의 意思와 最終情報 利用者의 意思(情報內容에 관한 意思)가 一致할 때는 情報選別作業이 經濟적으로 修行되고 있는가를 判斷해야 할 것이다. 또한 選別作業過程을 分析評價함으로써 不適合한 項目의 몇 %가 除去되고, 同時에 適合한 項目이 잘못 漏落된 것은 몇 %인가, 不適合한 項目의 除去에 所要되는 選別作業費用은 얼마인가 등을 判斷해야 할 것이다.

이러한 分析을 통해 平均調査正確率을 10%로 增加시키기 위한 事後選別作業에는 어느 정도의 費用이 所要되는가를 推定할 수 있다. 또한 正確率 增加에 따른 再現率의 減少도 豫測할 수 있다. 大規模 機械化 情報檢索시스템에서의 選別作業效率에 관한 最近의 研究에서 調査者가 選

別作業을 통해 調査正確率을 45%에서 75%로 增加시킬 수 있음이 나타났다. 이와 동시에 平均再現率은 10% 減少되었다.

費用-效果分析은 調査에 所要되는 費用을 構成要素別로 具體적으로 區分해야 하며 이를 통해 非能率的인 部分이 露出될 것이다.

앞서 언급한 情報시스템에서(Computer 情報檢索시스템)情報調査者의 情報調査件當 業務時間이 대략 아래와 같이 使用되었다.

情報要求者와의 面談	8分
調査範圍의 確定	20分
入力樣式構成	19分
調査結果選別	30分
計	77分

이 데이터를 통해 非能率的인 部分을 찾아냈다. 入力樣式을 作成하는 데 많은 時間이 消費되었다. 즉 調査項目의 組合(Combination)을 決定하는 데 많은 時間이 割愛되었고, 入力樣式에는 調査項目, 方法, 領域 등이 나타나며 記入에는 많은 勞力이 必要하다. 또한 情報調査者는 調査보다는 事務的인 일에 많은 時間을 消耗하고 있었다. 이런 경우 入力過程과 그 附隨過程을 簡素化한다면 調査過程의 經濟性을 期할 수 있을 것이다.

調査費用의 여러 要素를 分化함에 있어 데이터베이스의 크기에 대한 固定費用과 變動費用을 相互分離하는 것은 매우 重要한 일이다. 데이터베이스의 크기가 增加함에 따라 情報要求, 調査領域決定, 入力 등에 所要되는 時間이 同時에 增加한다고 말할 수는 없다. 데이터베이스의 크

기 증가는 調査에서 檢索되는 項目의 平均數만을 증가시키게 되어 情報選別費用만이 증가할 것이다.

索引作業에서와 같이 調査過程에 대한 費用-效果分析에서도 所要人力의 專門性 水準에 관한 評價가 並行되어야 한다. 專門性的 水準을 低下시킬 수 있다면 調査過程의 經濟化 達成은 容易할 것이다. 調査補助資料의 開發을 통해 調査者의 負擔을 輕減시키므로 經濟化가 이루어질 수 있다. 더 나아가 索引者와 調査者가 同一人이면 더 理想的이다. 2가지 能力을 兼備한 人力을 채용하여 一部時間을 索引에, 殘餘時間을 調査에 割愛한다면 가장 經濟的인 方法이 될 것이다. 그러나 索引作業이 調査作業보다 于先되는 경우가 많아 2가지 機能이 組合될 때 全体 生産性이 低下되는 경우가 있다.

費用-效果面에 영향을 미치는 또 다른 要素는 調査者와 情報要求者간의 相互討議이다. 이와 같은 相互討議는 利用者의 情報要求를 보다 明確히 나타내기 위해 發生하는 것이다. 相互討議가 증가할수록 調査의 質은 再現率과 正確率面에서 더욱 良好해질 것이다.

調査前에 調査者와 情報要求者간의 討議에 投入된 勞力은 調査後 選別作業에서 消耗되는 勞力과 相關關係에 있다. 情報利用者と 調査者간의 相互討議는 여러 단계에서 자주 나타난다. 즉 情報要求時, 調査領域決定時, 調査結果選別作業 등에서 發生한다. Lesk와 Salton은 SMART 調査의 評價에서 情報調査後 相談이 調査前 相談보다 價値있음을 發見했다(1). MEDLARS 評價(2)에서는 情報利用者と 調査者간의 對面討議가 調査結果의 質을 向上시키는 데 별 도움이 못되고 오히려 情報의 曲解 원인이 되는 것으로 나타났다. 따라서 情報要求者는 우선 그의 要求를 書面으로 提出하여 明確히 한 뒤 情報專門家와 協議해야 할 것이다.

費用-效果面에서 利用者-시스템 相互作用(討議)은 情報시스템에서 가장 중요한 過程이 될 수 있다. 情報利用者の 情報要求가 不明確할 때는 索引語, 調査方法 등이 아무리 우수해도 조사는 거의 失敗로 끝날 것이다. 情報利用者가 情報시스템에 가장 效果的으로 情報를 要求할

수 있도록 入力樣式을 改善시킨다면 시스템의 效率는 全体的으로 크게 上昇될 것이다. 入力樣式을 細密하게 記入하는 것은 많은 노력이 필요하지만 그 結果 調査의 質은 改善되고 長期的인 費用節減을 꾀할 수 있다.

7. 시스템 要素間的 相互關係

前에 살펴본 내용을 통해 情報시스템이 수궁할 만한 結果를 生産하는 데는 여러 方法이 있음을 알 수 있다. 다시 말해 情報시스템 運用에는 여러 代案이 存在한다. 費用-效果分析은 이들 代案을 비교 檢討하고 一定水準의 性能獲得을 위해 各種 構成要素의 最上組合을 決定할 것이다. 시스템을 構成하고 있는 各種 要素間에는 밀접한 關係가 있다. 이들 關係中 가장 普編的인 關係는 入力과 出力간의 關係이다. 즉 入力過程의 經濟性 追究는 出力過程의 負擔을 증가시키고 出力費用도 증가시킨다. 反對로 入力過程의 세심한 配慮는 出力效率를 증가시키고 出力費用을 減少시킨다.

시스템의 效率를 極大化하기 위해서는 시스템의 基本構成要素를 파악한 뒤 基本要素에 대한 各種 代案을 蒐集 檢討하여 最上의 代案을 選擇하고 選擇된 각종 代案을 組合, 一元化시켜야 할 것이다. 아래의 몇가지 관계는 同一要素에 대해 相異한 代案을 對比시킨 것이다.

1) 緻密하게 構成되고 統制되는 索引言語 對 非統制 Keyword의 自由使用

統制된 言語를 構成하고 推持하는 데는 많은 勞力이 所要되며 이것을 索引에 適用시키는 데는 多額의 費用이 支出된다. 또한 統制된 언어로부터 어떤 項目을 選擇하는 데는 키-워드를 自由로 부여하는 것 보다 더 오랜 시간을 요한다. Keyword索引은 高度로 統制된 언어 사용시 보다 高級人力을 많이 필요로 하지 않는다. 그러나 自然語나 키-워드는 調査領域을 決定할 때마다 調査者에게 統制言語를 分類해 내야 하는 負擔을 준다. 예를 들어 Petrochemicals나 Textile Industry를 생각할 때 조사자는 키-워드나 自然語로 표현되는 모든 경우를 생각해야 할 것이다. 키-워드의 非統制使用은 平均調査

正確率을 減少시킨다. 따라서 調査結果에 대한 選別作業이 필요하고 그에 따른 費用도 증가하게 된다.

2) 改定作業을 통한 索引의 엄격한 品質管理 對 再檢討過程이 없는 索引作業

檢討作業도 索引費用을 증가시킬 것이다. 그러나 索引檢討作業을 통해 不適合한 項目을 除去시키는 調査結果 選別作業을 省略시키므로 結果적으로 出力費用을 減少시킨다. 入力의 檢討作業이 經濟적으로 正當化 될 수 있는지 여부는 發生된 索引誤謬의 數와 檢討作業에 의해 修正된 誤謬의 數를 評價함으로써 判定할 수 있다.

3) 高度로 專門化된 統制언어 對 비교적 擴範圍하게 統制된 언어

一般的으로 前者를 推持하고 適用시키는 데 많은 費用이 所要된다. 索引語가 專門化 될수록 索引의 一貫性을 推持하기가 어렵고 이것을 利用키 위해서는 高級入力が 필요하다. 高度의 專門的 言語는 높은 調査正確率을 나타내며, 調査結果의 選別時間을 節約시킨다.

시스템의 觀點을 入力過程, 出力過程 중 어디에 두느냐에 따라 여러가지 相異한 作用素가 意思決定에 동원된다. 이들 作用素中 주요한 몇가지를 열거하면 아래와 같다.

1) 量(Volume)-索引處理된 文獻의 量과 每年處理된 情報要求量을 말한다. 極斷적으로 많은 文獻이 索引되었으나 極少量의 情報要求만이 處理되었을 경우 入力費用을 最少로 하고 많은 勞力을 出力 쪽으로 轉換시키는 것이 合理的인 것이다. 反對의 경우도 물론 存在할 수 있다.

2) 迅速한 入力-어떤 경우는 入力を 迅速히 해야 할 때가 있다. 例를 들어 文獻速報와 같은 경우는 다른 어떤 作業보다 入力を 중시하여야 할 것이다. 이런 경우는 入力보다 索引에 대해 經濟性을 追究해야 할 것이다.

3) 出力의 迅速性-正確하고 迅速한 出力이 要求될 때가 있다. 이런 경우 入力過程에 대해 經濟性을 追究했다고 하면 이것은 出力이 正確하고 迅速할 때에 한해서만 正當化 될 수 있는 것이다.

4) 副產物-어떤 경우 情報調査에 이용할 수 있는 데이터베이스를 低廉한 價格으로 構入할

수 있다. 例를 들어 自然語로 된 機械讀取可能한 데이터베이스를 冊發刊의 副產物로 얻을 수 있다. 비록 入力樣式과 그 內容이 費用-效果面에서 우리의 要求를 充足치 못한다 해도 데이터베이스의 價格이 普通價格이라면 이것을 利用하는 것이 바람직하다. 가능하면 데이터베이스를 약간 변조시켜 사용할 수도 있을 것이다. 곁들여 調査作業에 더 많은 노력을 投入하는 것은 더욱 바람직한 일이다.

前에 우리는 情報시스템의 下部시스템과 關係 있는 各種 要素를 費用-效果面에서 살펴 본 적이 있다. 사실 費用-效果分析에서도 效率評價分析에서와 같이 下部시스템을 個別的으로 評價하는 것은 非現實的이며 매우 위험한 일이다. 왜냐하면 이들 시스템은 相互 密接한 關係를 推持하고 있어 一部시스템에서의 變化는 시스템 全体에 영향을 미치기 때문이다. 이것은 매우 重要한 概念이다.

費用-效果分析에서는 直接的이고 卽刻的인 波及效果 뿐 아니라 間接的이고 長期的인 效果까지도 念頭に 두어야 할 것이다. 例를 들어 緻密하게 統制된 高度水準의 索引語로부터 간단한 索引語로 變更 使用기로 決定하였을 때 다음의 몇가지 直接的인 效果를 기대할 수 있을 것이다.

1) 索引語의 統制 및 推持費用 節減

2) 索引時間의 節約

3) 全体 索引時間의 效率性 提高

이와 더불어 長期的이고 間接的인 效果도 있다.

1) 調査領域決定에 필요한 時間의 증가와 이에 따른 調査費用의 증가

2) 調査正確率의 減少와 調査結果의 選別作業 필요

3) 選別作業의 필요에 따라 시스템이 所有한 各種 文獻代用物의 質的 向上 필요

이와 類似한 現象이 索引의 網羅性을 증가시킬 때도 나타난다. 直接的인 效果로는 索引時間과 費用의 증가, 檢索對象文獻의 增加, 再現率의 증가, 平均正確率의 減少 등이 나타난다. 長期的인 效果로는 正確率을 一定水準에 推持시키기 위한 調査結果의 選別作業 필요성, 選別作業의 效率的 運用을 위한 補助資料의 質的 改善 등

이다. 情報시스템은 매우 複雜한 組織이다. 따라서 어떤 變化에 대해 部分的인 效果만을 기대해서는 안될 것이다.

8. 費用分析要素

費用-效果, 費用-利潤分析 등을 情報시스템에 適用시키는데 가장 큰 障害는 情報서비스나 情報 그 自体에 대해 費用을 부과시키는 절차 (Costing Procedure)가 결여되어 있다는 사실이다. 적어도 費用에 관한 데이터가 여러가지 情報서비스에서 導出된다 해도 公式적으로 確認되는 경우는 극히 드물다. National Library of Medicine은 각종 費用데이터를 MEDLARS 運用に 適用시켜 보았다(3). Landau(4)는 索引과 抄錄에 관한 有用한 論文을 多數 發表하였다. 어떠한 情報시스템에서 認知possible한 媒体(Dollar 혹은 Cent)로 經濟性 向上을 測定할 수 있다면 시스템運용에도 費用概念을 導入해야 할 것이다. 우리는 어떤 節約이든(索引의 網羅性 減縮을 통한 節約이든, 어떤 要素의 제거를 통한 節減이든) 測定해 낼 수 있다. 그러나 예를 들어 시스템性能이 60%의 再現率과 50%의 正確率로부터 75%의 再現率과 50%의 正確率로 向上되었을 때 이 結果를 費用效果面에서 어떻게 測定할 것인가?

情報시스템에서의 費用賦課(Costing) 문제는 一般企業에서 使用하는 一般會計節次와는 相異하다. 文獻과 蒐集情報의 減價償却은 機械나 장비의 그것과 同一하지 않다. 더우기 장서수집에 지출되는 費用은 情報서비스의 提供量과 無關하다. 全体費用을 각 情報製品別로 分割(抄錄誌, 文獻速報, 遡及調査, SDI)시킨다는 것은 매우 어려운 作業이다. 어떻게 入力費用을 모든 情報製品과 서비스에 分配할 것인가? 우리가 入力費用을 實際로 分割할 수만 있다면 各種 書誌的 製品과 서비스에 대한 單位費用을 算出해 낼 수 있을 것이다. 典型的인 單位費用으로는 遡及調査單位費用, 檢索項目單位費用, 項目當 印刷單位費用, 提供된 項目當 單位費用 등이 있다. 그러나 이러한 單位費用은 量의 變化에만 매우 敏感할 뿐 性能의 變化에는 感應하지 못한

다.

예를 들어 National Library of Medicine에 의해 遡及調査의 費用計算에서 單位로 利用된 檢索項目當 單位費用을 생각해 보자. 이 費用은 量의 變化에는 매우 敏感하나 性能의 變化는 조금도 反映치 못했다. 예를 들어 檢索項目當 費用이 75¢라 하자. 이 單位費用은 出力의 量에 따라 變化한다. 우리가 調査領域을 變更하여 평소의 2배되는 項目을 檢索한다고 하자. 그러면 單位비용은 대략 37¢로 半感될 것이다. 사실 調査領域의 變更은 시스템性能의 급격한 低下原因이 될 수도 있다. 즉 調査領域의 擴大는 不適合한 項目의 증가원인이 되고 이에 따라 平均調査 正確率은 50%에서 25%로 떨어진다.

費用-效果分析을 위해서는 시스템의 效率變化에 敏感한 單位費用이 필요하다. 이러한 單位費用中的 하나가 檢索된 項目중 有用한 項目만을 對象으로 算出한 單位費用 C_R (Unit Cost per Relevant Citation Retrieved)이다.

시스템이 50%의 正確率로 運用되고 있다 하자. 이 시스템의 C_R 은 1.48달러이다. 시스템의 再現率과 正確率을 증가시킬 때 C_R 은 減少될 것이다. 따라서 C_R 은 情報시스템의 費用-效果分析에서 어떤 向上을 表現하는 데 有用한 單位費用이 될 수 있다. 이 單位는 同一 시스템의 多様な 運用型態를 相互 比較하거나, 根本적으로 相異한 시스템들의 費用效果를 測定하는 데 사용할 수도 있다. 이 測定置를 索引過程, 索引言語調査領域, 利用者와 시스템간의 相互作用型態 등에 影響을 미치는 經濟性 效果를 測定하는 데 使用할 수도 있다.

예를 들어 平均索引網羅性的의 增加는 再現率의 증가로 나타나고 索引費用의 증가로 나타난다. 또한 이것은 C_R 의 減少로도 나타난다.

Westat Research Inc는 U. S. Patent Office에서 여섯 개의 實驗用 組合索引시스템의 費用效果面을 測定하기 위해 檢索된 文獻中 有用 文獻만을 對象으로 算出된 單位費用을 測定單位로 삼았다. 또한 特許分類파일에서의 이들 시스템의 檢索效率를 受動式 調査方法의 效率과 比較하는 데도 이 單位費用을 利用하였다. 여기에서의 單位費用은 金錢이 아닌 時間이었다. 즉 이

곳에서 사용된 단위비용의 구성요소는 特許異議에 대항하기에 充分할 정도의 特許를 찾아내는 데 所要된 特許調査者의 平均所要時間이다. 有用項目에 대한 費用(C_R)은 出版된 索引이나 抄錄誌의 效率測定에도 利用된다. 이런 경우의 C_R 은 有用項目을 찾는 데 消費되는 所要時間이다.

印刷된 索引 2가지와 索引利用者 2人 A와 B가 있다고 가정하자. 또한 年間 索引中 25페이지가 A에게 有用하고 30페이지가 B에게 有用하다고 하자. A는 그에게 유용한 20page를 2시간에 걸쳐 찾아내고(再現率 80%), B는 28페이지를 1시간에 찾아냈다(再現率 98%).

누구도 100%의 再現率達成은 어렵다. 왜냐하면 利用者가 필요로 하는 項目中 一部 項目은 調査者가 認知치 못한 題目에 屬해 있거나 利用者 自身이 題目이나 抄錄으로부터 正確한 項目을 發見치 못했기 때문이다.

有用한 項目에 대한 單位費用(時間)은 利用者 A에게는 페이지당 6分(2시간×60分/20項目), B는 2分(1시간×60分/28項目)이다. A와 B가 正常的으로 調査한다면 索引은 A보다 B에 더욱 効果的이다. 왜냐하면 索引語가 B의 요구에 더욱 적합하게 構成되어 있기 때문이다.

利用者は 索引을 이용하여 有用한 項目을 찾을 때 그가 부딪치는 難易度에 따라 索引을 評價한다. 즉 有用한 項目에 대한 單位費用에 의거 판단한다. 利用者에 대한 索引의 效率는 索引利用者의 平均調査時間을 減少시킴으로써 改善될 수 있을 것이다. 이것은 多様な 方法으로 施行될 수 있다. 즉 索引의 網羅性 증가, 索引語의 專門性 증가, 참조체제의 구성 등을 통해 이룩될 수 있다. 이런 型態의 變化를 유발시키는 費用은 測定 可能하다. 즉 平均調査效率이라는 면에서 그 結果를 測定할 수 있다. 이들 여러 方法의 費用效果面은 情報製品生産費와 調査效率을 檢索된 有用項目當 平均調査時間을 媒体로 비교하므로 評價할 수 있다.

結 論

지금까지 3회에 걸쳐 몇가지 基本概念을 中心

으로 情報시스템의 分析評價方法을 紹介하였다. 특히 시스템의 分析評價方法으로 費用-效果分析技法을 導入, 그 基本要素를 概括적으로 살펴 보았다. 費用-效果分析은 어떤 特定경우 보다 一般的이고 보편적인 경향을 따라 遂行되어야 할 것이다.

예를 들어 마이크로피시도 費用-效果面에서 매우 우수한 情報傳達媒体에 속한다. 그러나 대부분의 情報利用者가 이 媒体의 受容을 거부한다면 費用-效果分析은 아무런 意味도 지니지 못한다. 위에서 언급한 바와 같이 人的 요소는 모든 他 要素에 우선한다. 따라서 모든 分析에 앞서 人的 變數에 관한 연구가 先行되어야 할 것이다.

参 考 文 獻

- 1) Lesk, M., G. Salton, Interactive Search and Retrieval Methods Using Automatic Information Displays, Proceedings of the Spring Joint Computer Conference, 435~446(1969)
- 2) Lancaster, F. W., Evaluation of the MEDLARS Demand Search Service. National Library of Medicine, Bethesda, 1968
- 3) Cummings M. M. Needs in the Health Science, Electronic Handling of Information, Thompson Book Co., Washington, D. C. 1967
- 4) Landau, H. B. The Cost Analysis of Document Surrogation a Literature Review, American Documentation 20(N. 4) 1969
- 5) Donald W. King, Edward C. Bryant, The Evaluation of Information Services and Products, Information Resources Press Washington, D.C. 1971
- 6) Carolos A. Cuadra, Annual Review of Information Science and Technology, American Society for Information Science. Vol.6, Encyclopedia Britanica, Inc, Chicago. 1971
- 7) F. W. Lancaster, The cost-effectiveness Analysis of Information Retrieval and Dissemination Systems, Journal of American Society for Information Science(February), American Society for Information Science p. 12~p. 26, 1971