

圖書館學과 情報科學分野에 있어서 索引誌 6種의 檢索比較 (1)

E. Michael Keen著*
李孝淑譯**

圖書館學과 情報科學 分野의 5種의 抄錄誌와 1種의 索引誌가 主題檢索 作業의 效果面에서 비교되었다. 探索要求는 가장 정확한 主題요구를 하는 것에 대해 대리로 探索하는 것으로 계획된 실험 테스트의 방법을 통해 1年分の 각 색인지에 대해서 施行되었다. 그후 그 效果와 效率性에 대해서 7가지의 각기 다른 측정법이 만들어졌다(유네스코 보고서에는 附錄資料가 卷末에 添加되었다).

1. 테스트된 索引서비스

여기에 보고된 評價테스트는 T. Edwards의 연구(A comparative analysis of the major abstracting and indexing services for library and information science. December 1974)를 補完하고 그 索引 서비스의 평가에 있어서 남아 있는 問題點들을 해결하기 위해 계획되었다. 조사된 색인지로서 이 보고서에서 略語로 使用될 索引誌는 다음과 같다.

Library and information science abstracts 1972, 1973(LISA); Library Literature, 1972, 1973(LL); Information Science abstracts, 1972 (ISA); Bulletin signalétique—Science de l'information, documentation, 1972(BS); Referativnyi zhurnal—Informatics abstracts journal 1972, 1973(RZI); Computer and control abs-

tracts, 1972, 1973(CCA).

여기서는 각 索引誌의 1973年版이 조사될 계획이었으나, 실제로 調査期間인 1975년 1월까지 ISA의 主題索引은 나오지 않았으며, 또 다른 색인 방법으로 사용되었던 1972년 號들도 조사되지 않았다. RZI는 主題索引을 주지 않고 있으므로 그 색인지가 사용하는 主題排列로서 조사되었다. LISA, ISA와 BS는 抄錄番號를 参照하게 하는 年間 주제색인을 주고 있으며 CCA도 마찬가지로 半年間으로 1년에 2번 내고 있다. LL은 아무런 抄錄도 주지 않는 색인지인데 BS가 제공하는 초록은 佛語로 되었으나 英語로 된 主題索引을 주고 있다.

이러한 6가지의 다양한 主題接近 方法에 대해서 各各의 方式이 갖는 特性을 명확히 把握하기 위해서는 상당한 分析이 要求된다. "KWIC 색인"(주요어 색인)에 관한 文獻에 대하여 각 색인지가 제공하는 記入의 種類는 다음과 같다.

- (1) UDC : KWIC indexing 987 (LISA)
- (2) KWIC indexing (LL)
- (3) KWIC, 지학, 정보체제, 190 (ISA)
- (4) UDC, KWIC 용어색인, Central Information Department, United States 1110(BS)
- (5) IRS 전개와 조작(RZI)
- (6) 기계화와 KWIC 프로그램 3-6892에 적용 가능한 UDC 정보검색(CCA)

4가지의 색인 서비스(ISA, BS, RZI, CCA)는 특히 情報科學文獻을 많이 포함하고 있고, 다른 두가지(LISA, LL)는 圖書館學文獻을 많이

*웨일즈大學 圖書館學科

**KORSTIC 資料部

포함하고 있다.

情報科學主題에 대해서 각 索引別로 30가지의 탐색이 시도되었다. 그리고 LISA와 LL에 대해서는 36가지의 圖書館學主題 探索이 시도되었으며, ISA, BS, RI에서는 12가지의 圖書館學主題探索이 있었다. 이것은 앞으로 제시될 結果에 대한 標本이 될 것이다. 이와 같은 方法으로 도서관학과 정보과학의 結果는 구별될 수 있을 것이다.

2. 調査方法

記錄用紙와 調査時間을 測定하기 위한 stop-clock과 함께 探索要求書가 6명의 圖書館學科 學生들에게 配布되었다. 탐색된 方法은 書面으로 된 指示書에서 다음과 같이 拔萃한 點을 보면 잘 알 수 있다.

(1) 탐색자는 탐색요구가 한 索引誌當 1年間 發刊分에 대해 조사되어야 할 것을 가정해야 한다. 질문들은 물음(question)의 형태이거나, 한 주제에 대한 說明을 요구하는 형태일 것이다. 또 이 두가지 경우는 모두 願하는 要求主題에 대해 論한 文獻이다. 따라서 探索者는 索引誌의 探索에서 찾아냈던 引用(citation)에 따라 圖書館書架로 갈 것을 豫想해야 한다. 探索者는 探索質問이 자신의 것이거나 혹은 다른 利用者의 것을 대신하여 탐색하고 있다고 생각해도 좋다.

(2) 탐색자는 가능한한 빨리, 또는 15분이내라는 制限된 時間內에 될수록 많은 引用目錄을 만들어야 한다.

(3) 탐색자는 그가 할 수 있는 한, 가장 적합한 인용이 무엇인지 판단해 내야 한다. 하나의 인용이 適合한 것인가 아닌가를 알 수 없을 때는 그것을 목록에 올려야 한다. 왜냐하면 抄錄들을 상세히 조사하는 것이 아니라 필요한 경우 그 초록들을 대강 훑어보는 것으로 그치기 때문이다.

(4) 探索記錄用紙에 記入해야 할 3가지 사항은 다음과 같다.

① 探索者가 適合하다고 認定하여 찾아낸 引用

② 각각의 적합한 引用들을 찾아낸 시각

③ 調査는 했지만 不適合하다고 決定된 引用
(5) 探索記錄用紙에는 탐색이 끝난 最終時刻을 적어야 한다. 또한 탐색을 끝내고 그에 대해서 다음의 4가지 중 하나에 표를 해야 한다(적합한 경우에는 한가지이상 표를 해도 좋다).

① 제한된 時間에 到達했다.

② 완전히 適合한 參考資料를 찾았다.

③ 더 적합한 資料는 없을 것 같다.

④ 더 적합한 探索用語가 없었다.

(6) 한 가지 探索質問을 處理한 후에 탐색자는 그 質問을 다시 調査하지 않을 것이다. 그리고 탐색자는 같은 색인지에서나 다른 索引誌에서, 다른 質問을 探索하려고 할 때 그 이전의 탐색에 대해서는 잊어버리도록 해야 한다.

(7) 우리는 탐색자의 技術이나 速度를 테스트하고 있는 것이 아니라 단지 한 색인지에 대한 것과 또 다른 색인지에 대해서 각 探索者를 比較하는 것이다. 그래서 重要的 것은 探索能力이 높으나 낮으나가 문제가 되는 것이 아니라 각기 다른 색인에 대해서 가능한한 一貫性있게 探索하는가가 중요한 것이다.

探索者는 주어진 探索要求를 위해서 각각의 색인지가 唯一하고 유용한 索引이라고 가정하고 그것에서부터 가장 최상의 것을 얻도록 해야 한다. 어떤 색인지에 대해서는 好感이 가고 어떤 색인지에 대해서 好感이 가지 않는 것은 당연한 것이다. 우리는 이 實驗探索이 끝난 후 이러한 探索者의 感情을 알 수 있다.

이 實驗計劃은 각 探索者들이 각각의 索引誌에 대해서 같은 數의 調査를 하지만 直接 學習效果(direct learning effect)를 論하기 위해서 한 번 이상은 같은 탐색요구를 탐색하지 않는 Latin squares에 기초하고 있다.

이러한 테스트를 하기 이전에 제시되는 指示와 探索은 그 탐색과 또다른 테스트가 계속되기 3½일(3½ days) 前에 먼저 시행되었다.

탐색은 탐색자가 기록하는 對象, 페이지를 찾는 데 걸린 時間, 페이지를 넘기는 일 등이 다른 사람이 관찰하는 가운데서 탐색되었다. 또다른 테스트에서 탐색자들은 세가지 點에 근거해서 그들이 찾아냈던 인용들의 適合性을 判斷했으며, 索引이 提供하는 記入의 적합성 豫測을

결정하는 방법으로서 그들 스스로 完全文獻을 調査하고 그 적합성을 再判斷했다. 6명의 탐색자들은 이 實驗探索이 끝났을 때 각색인들에 대한 그들의 見解를 묻는 設問紙에 답했다. 그리고 탐색자들에게 탐색요구를 하는 이유는 알리지 않았고 또한 어떤 색인지에서도 탐색요구에 대해 적어도 하나의 가장 적합한 문헌이 있다는 것도 탐색자들은 알지 못했다.

이것은 소위 原文方法(source document method), 즉 요구를 蒐集해서 適合文獻如否를 判斷하는 방법이 사용되었기 때문이다. 즉, 실험탐색을 하기 전에 6가지의 색인지에서 찾아낼 文獻들을 著者索引 방법에 의해 찾아내야 하고 그 실험을 위해서 적합한 1972, 1973년도 색인지에도 그 문헌들이 포함되어 있어야 하는 방법이다. 이 문헌들에 대한 초록을 사용할 때 탐색요구는 사실성있고(realistic), 원문의 제목만을 再叙述하는 것이 아닌 것으로 만들어져야 한다.

약 30가지의 탐색요구는 정보과학 주제분야를 위해서 이 방법을 택했었다. 또한 圖書館學分野를 위해서는 18가지의 탐색요구가 LISA와 LL에 대해서 같은 원문(source document)방법을 사용했다. 그리고 각 6개의 探索要求로 形成된 그룹들은 ISA, LL, RZI에서 적합한 원문을 갖고 있다. 情報科學 탐색요구는 각 색인에서 각각 한번씩 탐색되었고, 도서관학분야를 위한 요구는 적합한 색인지에서 두번씩 탐색되었으며, 그 결과는 다음의 데이터에 나타나 있다.

3. 테스트 結果

研究된 7가지 探索基準과 사용된 重要 測定方法은 다음과 같다.

- (1) 적합한 原文을 찾는 能力(再現率)
- (2) 探索에 의해 適合하다고 判斷된 文獻을 찾는 능력
- (3) 不適合하다고 판단된 文獻들을 삭제하는 能力(正確한 選擇率)
- (4) 탐색하는데 걸린 時間(探索時間)
- (5) 肉體的인 探索作業(페이지 넘기기)
- (6) 索引記入의 明瞭함(適合性 豫測能力)
- (7) 主觀的 趣向(設問紙 結果, 全体的 趣向,

價值順位)

3.1 完全探索 結果

5가지 實驗基準에 의한 결과는 表 1에서와 같으며 두 주제로 나뉘어 있다. 그리고 나열된 索引順序에는 아무 의미도 없다.

LISA, LL, ISA는 情報科學과 圖書館學 分野에서 모두 再現率과 適合文獻 選擇方法에서 BS, RZI, CCA보다 우수하다. 정확히 선택하는 비율은(selected precision ratio)抄録이 부족한 LL을 제외하고는 모두 비슷하다.

페이지 넘기기(page turns)같은 노력의 測定方法은 情報科學探索을 위해서 만들어진 것으로 LL이 가장 效果的이고 ISA는 效果的인 2단계 색인이다. CCA는 페이지를 넘기는 수는 비교적 적으나 再現率과 適合率이 매우 낮아서 실제로 이것은 페이지를 넘기는 데 보다 오히려 페이지를 調査(consultation)하는 데에 시간이 많이 걸려서 非能率的이다. 그리고 LL의 1972년 蓄積版은 1973년의 6種의 非蓄積版보다 探索力이 낮다. 이에 대한 分析은 이후의 報告書에서 발표될 것이다.

거의 모든 방법에서 도서관학 探索은 정보과학탐색보다 약간 더 좋은 결과를 나타내고 있는

表 1. 완전탐색결과

서비스	재현율	선택된 적합 문헌	정확히 선택한 비율	탐색 시간	페이지를 넘긴 수
정보과학					
LISA	67	3.1	44	11.4	26.9
LL	53	3.1	60	10.8	14.7
ISA	63	2.9	48	11.2	20.3
BS	33	2.2	44	12.4	29.7
RZI	30	2.0	43	14.9	39.4
CCA	30	1.4	50	12.8	19.7
도서관학					
LISA	75	3.7	52	10.9	
LL	78	3.9	84	12.0	
ISA	67	3.3	57	11.1	
BS	42	1.8	51	12.2	
RZI	33	1.4	35	13.5	

表 2. 정보과학 주제탐색을 위해 혼합된 측정방법을 사용한 완전탐색결과

서비스	적합문헌을 찾는 데 걸린 시간	적합문헌을 찾은 최종 시각	적합문헌 에서 페이지 를 넘길 수	10분간 페이지를 넘길 수
LISA	3.64	2.98	7.2	22.4
LL	3.46	2.76	4.4	13.4
ISA	3.81	2.88	6.6	17.7
BS	5.65	4.43	9.9	23.0
RZI	7.47	6.15	14.3	27.7
CCA	9.14	7.52	11.2	14.8

데, 이것은 아마도 탐색자들이 도서관학을 훨씬 더 잘 알고 있음을 반영하는 것이다.

表 2는 情報科學探索을 위해 이제까지와 같은 자료에 基礎하여 2가지 평가기준을 반영시켜 4개의 追加된 測定方法을 나타낸 것이다. 두번째의 "適合文獻을 찾는 데 걸린 시간" (time per relevant)은 최종적으로 適合한 文獻이 발견된 시간을 기초로 한 것인데 이것은 각각의 탐색에서 여러 索引誌를 多樣하게 使用해서 한 것이 아니라 한 가지 방법으로 계속 찾아서 그 文獻을 발견하기까지에 所要된 시간을 말한다.

適合한 방법마다 所要된 時間과 適合文獻을 찾기 위해 페이지를 넘긴 數에 있어서는 LISA, LL, ISA가 다른 것들보다 훨씬 효율적이다. 그리고 10분간 페이지를 넘긴 수에 있어서는 LL, CCA, ISA가 가장 探索成果가 좋다.

適合한 방법마다 걸린 시간에서도 일반적으로 같은 결과를 내고 있으며, LL이 가장 성과가 좋다. 적용범위를 자유롭게 하는 방법(coverage-free measure)에서는 LISA보다 ISA가 두번째로 효과적이고, 그리고 ISA는 탐색이 끝났을 때 적합한 문헌을 많이 提供하지 못하는 데 이것은 아마도 適合文獻을 비교적 적게 包含하고 있음을 나타낸다.

3.2 낮은 再現目標 結果

表 3에서는 낮은 再現目標 結果를 위해서 利用者가 단지 하나의 적합한 項目만을 원하는 것으로 假定하여, 탐색자는 하나의 적합한 文

表 3. 하나의 적합문헌을 찾는 것을 목표로 한 낮은 재현목표에서의 탐색결과

서비스	탐색된 재현율 (%)	정확히 선택한 비율 (%)	탐색시간 (minutes)	페이지를 넘긴 수
정보과학				
LISA	93	36	4.3	12.8
LL	90	75	4.0	6.3
ISA	100	50	3.7	6.4
BS	80	43	5.9	11.0
RZI	77	47	9.5	20.3
CCA	63	40	9.0	13.4
도서관학				
LISA	94	39	4.5	
LL	94	74	4.3	
ISA	100	55	4.3	
BS	75	43	8.5	
RZI	67	40	10.0	

獻을 찾아내는 즉시 탐색을 끝내고, 그 결과를 기록해 놓은 것이다. 탐색된 再現率(the sought recall ratio)은 탐색된 것중에 얼마정도가 이러한 하나의 적합문헌을 찾아내는 목표를 달성하는가를 말하는 것이다.

ISA는 거의 모든 방법에서 다른 색인지보다 우세하다. 그리고 또 LISA, LL, ISA는 재현, 정확성, 시간에 있어서 다른 것보다 훨씬 효과적이다. 정확히 주제선택을 하는 방법(selected precision measure)은 그것이 완전탐색결과에서 나타난 것보다 차이가 좀더 많이 난다. 그리고 表 1과 비교해 보면 LL은 측정방법의 차이를 잘 나타내고 있다. LISA, LL, ISA간의 탐색 시간에서의 차이는 미약한 것이고 LISA, ISA는 각각 CCA, RZI보다 0.2% 혹은 1%정도 탐색성도가 좋고 BS는 그 두 그룹의 중간이다. BS는 ISA보다 10%정도 못한 반면 CCA나 RZI보다 1%정도 좋다.

2단계 색인 (ISA)은 1단계색인 LL이 놀라운 결과를 내는 것만큼 페이지를 넘기는 수고가 없어야 한다. RZI는 主題索引이 부족해서 앞에서 설명했듯이 페이지를 넘긴 수, 探索時間, 再現率에서 가장 非效果的이다.

圖書館學探索은 情報科學探索보다 첫번째 적

합한 문헌을 찾아내기까지 조금 더 시간이 걸리지만 다른 모든 방법에 있어서는 이 두 主題가 모두 비슷하다.

3.3 適合性 豫測

適合性 豫測이란 探索者가 索引誌에서 그가 찾는 것이 얼마나 正確한 것인가를 아는 것이다. 다시 말해서 探索의 結果는 탐색자가 適合할 것이라고 豫測한 인용에 따라 文獻을 充分히 調査했을 때 그것이 정말 적합했다고 증명되는 경우 探索者의 豫測 正確성에 의해 그 결과가 판단된다. 적합성예측이 높다는 것은 적합하다고 예측된 하나의 인용이 적합한 것으로 증명되는 것이라기보다 그러한 경우가 종종 많이 일어나는 것을 말한다. 그런데 적합성 판단은 단지 索引이 提供하는 것 즉, 索引用語, 文獻題目, 提供된 文獻抄錄에 의해서만 이루어진다. 왜냐하면, 이러한 실험탐색자들은 색인지에서 適合性を 判斷할 때 너무 오랜 시간을 보내지 않도록 指示를 받았고, 단지 그 抄錄들을 대강 훑어보도록 했기 때문이다.

탐색자들에 대한 指示文에서 다음과 같이 발췌한 것들이 이러한 實驗方法을 설명해 주고 있다.

우리는 몇가지 탐색을 하고자 한다. 그런데 그 탐색에서 탐색자는 실제로 CLW 도서관의 서가에 가서 탐색에서 찾아낸 문헌들을 보아야 한다. 탐색자가 해야 할 첫째 작업은 그가 먼저 했던 탐색을 回想해서 탐색기록용지(retraced search record)에 기록을 해야 한다. 탐색자는 전에 목록에 올렸던 인용을 다시 찾아 봄으로써 각 인용에 대해서 다음과 같이 分類하면서 판단을 내릴 것이다: H(質問에 매우 適合한 것); L(質問에 대해 適合성이 낮은 것); CB(質問에 適合할 수 있는 것)

탐색자는 이것을 끝내자마자 圖書館 書架로 가서 그 문헌들을 찾아보라. 그리고 질문에 대한 그 文獻의 適合성을 다음과 같이 判斷해야 한다: H-높다; L-낮다; N-적합치 않다.

만일 그 完全文獻을 보았을 때 索引誌를 보고 내렸던 判斷과 아주 다른 판단을 내리게 된다면 그것은 문제가 되지 않는다. 탐색자는 그 문

表 4. 적합성 예측의 결과

(각 색인지간의 구별은 되지 않았음)

	색인지 적합성 원문 적합성 (%)	가능한 색인지 적합성 원문 적합성(%)	문헌의 수
두 주제분야의 모든 문헌	91	55	150
" 원문	100	100	55
" 비원문	84	41	95
정보과학의 비원문	86	19	56
도서관학의 "	82	67	39

헌을 완전히 읽지는 말아야 한다. 즉, 탐색자는 그 문헌을 좀더 參考하기 위해서 圖書館에서 빌려올 것인지를 決定하기 위해서 각 文獻을 약 5분정도 훑어보아야 한다.

이 테스트에서 36가지 情報科學探索과 30가지의 圖書館學探索이 이루어졌다. 適合資料判斷은 209種의 文獻에 대해서 하려고 했으나 150가지로 줄었다. 왜냐하면 몇가지 文獻은 그 시각에 그 도서관에 없었으며, 또 몇가지는 英語로 되어 있지 않았으므로 정해진 探索時間內에 적합성 판단을 할 수가 없었다.

우리가 알고 있는 한, 색인지에 대한 適合性 豫測이 測定된 것은 이번이 처음이다. Legate의 *The development and evaluation of a mechanized SDI service for biologists*는 機械化한 SDI체제에 대한 탐색결과를 고찰한 것이었고, 서명과 초록을 보고 적합성을 예측하는 것은 여러번 고찰된 것이다. 表 4에서는 6가지 색인지 모두에서 나온 결과를 요약하고 있다. 表 4에서의 결과는 4가지의 가능한 結果 중에서 2가지를 기록한 것이며 다른 두가지는 같은 방법에 의해 계산될 수 있다. 그리고 表 4에서 우리는 적합성 예측의 전체적인 결과와 주제분야, 원문, 비원문에 따라 측정된 適合性 豫測率을 알 수 있고, 또한 원문은 적합성 예측이 항상 완전한 것임을 알 수 있다.

이상의 두 주제분야(情報科學, 圖書館學)는 몇가지 차이점을 나타내고 있다. 즉, 색인지를 보고 적합하다고 판단했을 때는 적합성 예측(relevance predictability)에서 거의 차이가 없었으나 색인지에

서 적합할 수 있을 것이라는 판단을 내린 경우에는 정보과학에서 거의 항상 부적합했고 도서관학에서는 종종 적합한 것으로 나타났다. 이것은 탐색자들이 모두 圖書館學科 학생들이어서 그들이 잘 알지 못하는 주제에 대해서는 절대 안전한 판단방법을 훨씬 더 많이 사용했기 때문이다.

유감스럽게도 表 4의 색인지와 주제분야에 대해 요약해 놓은 데이터는 너무 標本數(sample size)가 작아서 그렇게 큰 의미를 주지는 못한다. 그러나 이 결과는 예상했던 경향의

일부를 보여주고 있다. 즉 BS와 LL은 다른 것보다 탐색성도가 낮다. 그 첫째 이유는 그 抄錄이 英語로 되어 있지 않기 때문이고, 둘째 이유는 그 색인지에는 적합성을 豫測하기 위한 것으로서는 索引表題語와 書名뿐이기 때문이다. LISA와 ISA에서는 적합성 예측성도가 가장 좋고 RZI와 CCA도 이러한 적은 수의 표본 크기에 대해서는 적합성 예측의 결과가 좋은 편이다 (다음 호에 계속)

情報管理研究 Vol. 11, No. 2

印刷・發行 1978. 4. 25.

編輯 情報管理研究會

發行 韓國科學技術情報센터

서울特別市東大門區清涼里洞206-9

電話 (965) 6211~18

定價 600원 年間購讀料 3,000원