

OPAC에서 서명단어탐색의 문헌순위화에 관한 연구*

An Experimental Study on Ranking Output of Title Word Searching in the Boolean OPAC Systems

노정순(Jung-Soon Ro)**

초 록

본 연구는 불질의 기반의 OPAC에서 정렬과 적합성 순위화알고리즘의 효과를 분석하고, 순위화의 성능을 평가하는 척도를 연구하였다. 연구결과 출판년도순 정렬이 저자명순, 서명순, 출판사순 정렬보다 우수하였으나 유의한 차이는 아니었다. 용어빈도수에 기반을 둔 퍼지나 DNF 모델보다는 위치정보에 기반을 둔 알고리즘의 순위화 성능이 더 우수하였다. 6개의 순위화 성능평가척도간에 차이는 없었으며, 순위화시스템에서 정확률 P보다 분별력이 높은 단순 척도의 유용성이 확인되었다.

ABSTRACT

The characteristics of the short document representatives and short queries of OPAC systems need the different ranking algorithms from IR systems. This study tested and analyzed the effectiveness of four sorting schemes and four ranking algorithms and the six effectiveness measurements for the ranked Boolean OPAC systems.

The sorting by publication year was better but without significant difference. The cover density ranking was significantly better than the frequency-based ranking of the Fuzzy or DNF models. The simple effectiveness measurement based on the average rank of relevant documents retrieved was as good as the others and better than the precision P.

키워드 : 정렬, 순위화, OPAC, 불논리검색, 시스템효과평가척도, ranking, sorting,
Boolean searching, effectiveness measurement

* 이 논문은 2000년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

** 한남대학교 문헌정보학과 교수(jsr@mail.hannam.ac.kr)

■ 논문 접수일: 2001년 3월 26일

■ 게재 확정일: 2001년 6월 5일

1 서 론

1.1 연구목적

온라인 도서관목록(Online Public Access Catalog : OPAC)이 독점적으로 장서목록의 수단어 된지 수십년이 지났지만 OPAC의 성능향상을 위한 연구는 정보검색(IR)시스템이나 인터넷 Web 탐색시스템에 비해 상대적으로 다양하지 못하였다. OPAC 연구는 주로 이용형태나 탐색실패의 원인을 구명하는데 집중되었다. 시스템의 탐색성능을 향상시키기 위한 첨단 정보검색 이론과 기술들은 주로 IR시스템에서 연구되어 발전해 왔고, OPAC은 IR시스템에서 일반화된 기능들을 채용하였다. 보다 나은 성능향상을 위해 탐색식을 자동으로 생성하거나, 입력된 탐색식의 오타를 자동으로 수정하고, 검색된 문헌을 적합성 순서로 출력하고, 시스템이 탐색식을 수정하여 탐색결과를 확대하거나 축소하는 기능들이 3세대 OPAC에서 요구되고 있다.

OPAC시스템의 문제는 크게 두 가지로 귀결된다. 한 건의 문헌도 검색되지 않는 것과 너무 많은 문헌이 검색되는 것이다. 적합문헌이 검색되지 않는 문제는 OPAC에서 탐색실패(no posting)의 원인을 규명하려는 많은 연구들로 그 원인과 해결책이 제시되었다. 그러나 부적합 문헌이 너무 많이 검색되는 문제를 탐색자의 탐색능력을 높이는 것 외에 시스템으로 해결하려는 노력은 많이 부족하였다.

저자명이나 주제명표목과 서명에서 불리언 연산자를 사용하여 단어단위 키워드탐색이 가능하게 된 2세대 OPAC에서 너무 많은 부적합

문헌의 검색은 여러 연구에서 보고되었다. 특히 저자나 서명으로 알고 있는 문헌을 탐색할 때보다 알지 못한 문헌을 관련주제로 탐색할 때 이 문제는 심각하였다. 잡지 기사보다 포괄적인 정보를 다루는 도서의 특성상 이용자는 많은 문헌을 원치 않음에도 불구하고 주제탐색이 너무 많은 문헌을 검색한다는 것이다. 켈리포니아대학의 MELVYL시스템에서 이용자들은 평균 77.5건의 문헌을 검색하였으나 9.1개만 디스플레이하였다(Larson 1986). 100건 이상 검색한 탐색의 37%와 1000건 이상 검색한 탐색의 52%는 주제색인을 사용한 탐색이었다. Lynch(1989)의 연구에서 주제탐색은 평균 179.6건을 검색하였으며, 7년 동안(1982-1986) OPAC 사용을 3개월 간격으로 분석한 Larson(1991)의 연구에서 주제탐색은 평균 156-350건을 검색하였다. Barbuto & Cevallos(1991)의 연구에서 OPAC이용자는 높은 정확률보다는 높은 재현율을 원하는 (comprehensive) 탐색을 원한다고 하지만, 실제 높은 재현율을 원하는 이용자의 54%는 1-20건의 문헌이 검색되기를 원하였다. Wiberly, Daugherty, & Danowski(1990) 연구에서 이용자들이 너무 많이 검색됐다고 생각하는 문헌 수는 중앙치(median) 15건이었고, 30건 이상이 검색되면 저항감을 느꼈다. 같은 저자의 1995년 연구에서는 평균 90건의 문헌을 검색하였으나 9건만 디스플레이한다고 보고되었다.

주제탐색 중에서도 주제명 키워드탐색보다 서명단어탐색에서 부적합 문헌의 검색은 심각한 것으로 보고되었다. Peters & Kurth(1991)의 연구에서 주제명 탐색은 평균 19개 문헌을 검색하였으나, 서명단어 탐색은 105개의 문헌

을 검색하였다. Carlyle(1996)의 연구는 서명 단어 탐색의 정확률(중앙치)은 7%이고, 90개의 탐색 중 74개의 탐색(82%)이 19% 이하의 정확률을 보였다고 보고하고 있다.

OPAC 시스템에서 LCSH를 유지하는데 많은 경비가 들지만 통제어 탐색을 제공하는 이유는 Voorbij(1998)의 연구가 잘 설명한다. 주제전문사서가 탐색을 대행하였지만, LCSH를 사용한 통제어 탐색은 86%의 재현율을 보인 반면 서명단어 탐색은 48%의 재현율을 보였다. 전체 적합문헌의 35.2%만이 주제명과 서명단어 탐색에서 모두 검색되었고, 51.8%는 주제명 탐색으로만, 13%는 서명단어 탐색으로만 검색되었다. 주제명 키워드의 38%만이 서명단어와 일치하였다.

이상과 같이 서명키워드 탐색은 주제탐색 대용으로 완벽하고 효과적으로 사용될 수 없음에도 불구하고 우리나라에서는 한글자료에 통제된 주제명을 부여하지 않기 때문에 대부분의 OPAC에서 주제명 탐색은 불가능하다. 대부분의 이용자(End User)는 분류번호를 정확히 알지 못하기 때문에 분류번호 탐색은 거의 이용되지 않는 것으로 보고되었다(0.3%(Hunter 1991) ~ 0.6%(Wallace 1993)). 이 때문에 요즘의 MAESTRO에서처럼 분류번호순 주제 열람기능이 대신 제공되기도 하나, 세분된 소주제목까지 안내되지 못하기 때문에 높은 정확률을 위한 탐색에는 도움이 되지 못하고 있다. 이와 같이 서명단어 탐색이 유일한 주제탐색의 접근점이 되고 있는 우리나라 OPAC 상황에서 서명단어 탐색으로 부적합문헌이 검색되지 않도록 정확률을 향상시키는 장치가 요구된다. 본 연구는 불질의탐색 환경의 OPAC에서 검색

된 문헌을 적합성 순위로 출력하여 탐색자들이 적합성이 높은 상위의 문헌을 원하는 만큼만 출력함으로써 높은 정확률을 얻게하려는 목적에서 순위화 알고리즘의 효과를 연구하였다.

불리언탐색결과를 순위화하려는 연구는 IR 분야에서 다양하게 연구되어 왔다. 그러나 OPAC 특히 한글문헌을 대상으로 하는 OPAC에서의 연구는 거의 이루어지지 않았다. OPAC에서 문헌의 대응물인 제목은 IR에서 문헌의 대응물인 초록이나 전문과 비교하여 길이가 짧고, 단어의 출현 빈도도 매우 낮다. 또한 탐색질의도 IR에 비해 길어도 짧고 특정성도 매우 낮다. 본 연구에서는 단행본도서의 특성(초록과 전문 대신 서명 사용)과 OPAC 탐색질문의 특성(낮은 특정성, 짧은 질문)에 적당한 순위화 알고리즘을 개발하는 것을 그 목적으로 하고 있다.

1.2 연구과제

1. OPAC의 서명단어탐색의 성능은 어떠한가?
2. OPAC 이용자들은 몇 권의 도서를 대출하기 원하는가? 검색된 적합문헌과 대출 희망도서는 차이가 있는가?
3. 탐색결과를 정렬시킬 경우, 저자명순, 서명순, 출판시순, 출판년도순 중 가장 좋은 정렬방법은 무엇인가?
4. 탐색결과를 문헌의 적합성기중치로 순위화시킬 때 가장 좋은 순위화 알고리즘은 무엇인가?
5. 어느 순위화방법이 가장 좋은가를 측정하는데 적당한 척도는 무엇인가?

6. 적합문헌에서 탐색용어는 어디에 출현하는가?

1.3 연구가설

1. 탐색결과와 저자명순, 서명순, 출판사순, 출판년도순 정렬방법의 정확률에는 차이가 없을 것이다.
2. 정렬방법과 문헌순위화기법의 정확률에는 차이가 없을 것이다.
3. 정확률 측정정도 간에는 차이가 없을 것이다.

2 선행연구

2.1 불리언시스템에서 용어 가중치 기반의 순위화

문헌의 순위화알고리즘에 대한 연구는 IR 분야에서 많이 수행되었다. 비불리언 환경에서 순위화방법은 노정순(1999)에 의해 포괄적으로 리뷰되었다. 불리언 환경에서 확장된 불리언기법은 Fox et. al.(1992)에 의해 리뷰되었으며, 한글 DB를 대상으로는 P-Norm 검색에 의한 순위화기법이 연구되었다(이효숙 1993; 고미영, 정영미 1999). OPAC에서 순위화기법은 비불리언 환경에서는 OKAPI, CHESHIRE, MUSCAL, CITE와 같은 시스템에서 사용되었으나, 불리언 환경에서는 퍼지이론을 적용시킨 츠쿠바대학의 FACOM Fair-I가 보고되었다(김현희 1994).

IR 분야에서 불질의로 검색된 문헌을 순위

화시키는 연구에서 문헌의 순위는 연산자 AND, OR, NOT의 처리에 따라 결정되었다. 벡터모델을 응용한 혼합불리언시스템(예, SIRE, NORIS 등)에서 검색된 문헌의 가중치는 모든 용어가중치의 합으로 결정되었다(Noreault & McGill 1977, Bing & Selmer 1980). Angione(1975)는 OR로 연결된 용어에는 가중치 1을, AND로 연결된 용어에는 가중치 2를 부여하고, 문헌의 가중치가 3 이상인 문헌만을 검색하였다.

논리합정규형(Disjunctive Normal Form)을 이용한 불리언시스템에서 불질의는 논리적으로 항등한 논리합(OR)으로 연결된 형태의 탐색식으로 변환된다. 예를 들어, 탐색식1 (A or B) AND C는 탐색식2 (A and C) OR (B and C) 혹은 탐색식3 (A not B and C) or (A and B and C) or (not A and B and C)와 같은 논리합정규형으로 표현된다. Bookstein(1978)은 탐색식1을 탐색식 2와 같은 논리합정규형으로 표현하여 검색된 문헌에 OR로 연결되는 아톰(Atom)을 모두 만족시키는 문헌에서부터 한 개만 만족시키는 문헌까지 내림차순으로 순위를 부여하였다. <표 1>과 같은 행렬도에서 아톰 1(A and C)는 D1과 D2를, 아톰2(B and C)는 D1과 D3을 검색한다. D1은 두 개의 아톰을 모두 만족시키기 때문에 가중치 2를, D2와 D3는 가중치 1을 갖게 된다.

그러나 Radecki(1982)의 연구에서 탐색식1은 탐색식3으로 더 세분되고, 검색된 문헌은 세 아톰 중 어느 하나만을 만족시킨다. 아톰1은 D2를, 아톰2는 D1을, 아톰3은 D3을 검색한다. DNF로 검색된 문헌의 순위화는 아톰의 적합성확률에 의해 결정된다. 단어 A와 B, C

〈표 1〉 문헌-용어 행렬도

문헌 \ 용어	A	B	C	D	E
D1	0.2	0.5	0.1		
D2	0.7		0.2	0.1	
D3		0.4	0.3		0.2

의 적합성 가중치가 각각 0.2, 0.6, 0.1이라고 하면 아톰1(D2), 아톰2(D1), 아톰3(D3)의 가중치는 각각 0.3, 0.9, 0.7이 됨으로써 DNF로 검색된 문헌은 D1, D3, D2의 순서로 순위화된다.

퍼지(Fuzzy) 개념을 기반으로 한 불탐색은 하나의 파셋내에 OR로 연결된 용어의 가중치는 최대값을 취하고, AND로 연결된 용어의 가중치는 최소값을 취하여 문헌의 순위를 결정한다. 탐색식1로 검색된 문헌을 퍼지이론으로 순위화시키면 OR은 최대값, AND는 최소값을 갖기 때문에 D1은 0.1, D2는 0.2, D3은 0.3의 가중치를 갖게되고, 검색된 문헌은 D3, D2, D1의 순서로 순위화된다. 퍼지개념에 기반하여 Fox & Sharat(1986)가 개발한 MMM모델은 불연산을 융통성있게 하기 위해 최대값과 최소값 문헌가중치를 선형으로 결합하여 질의-문헌의 유사도를 구한다. OR질의에는 최대값에, AND질의에는 최소값에 비중을 더 주기 위해 일반적으로 상관계수 C_1 에 C_2 보다 큰 값을 준다.

$$\text{Sim}(D, Q_{\text{OR}}) = C_1 \times \text{최대값} + C_2 \times \text{최소값}$$

$$\text{Sim}(D, Q_{\text{AND}}) = C_1 \times \text{최소값} + C_2 \times \text{최대값}$$

Mitra et. al.(1998)의 연구에서는 OR로 연

결된 용어는 퍼지모델처럼 최대값을 취하나, AND로 연결된 용어나 파셋은 합(Σ)으로 처리되었다. Robertson(1978)은 퍼지개념에 불확실성을 고려하여 a, b가 독립적일 때 확률이론(x 가 A와 B를 갖을 확률 = x 가 A를 갖을 확률 \times x 가 B를 갖을 확률)를 적용시켰다.

$$A \text{ and } B = W_a \times W_b$$

$$A \text{ or } B = W_a + W_b - (W_a \cdot W_b)$$

$$\text{Not } A = 1 - W_a$$

그러므로 〈표 1〉의 예에서 탐색식A로 검색된 D1, D2, D3은 각각 0.06, 0.14, 0.12의 문헌가중치를 갖기 때문에 검색된 문헌은 D2, D3, D1의 순서로 랭크된다.

P-Norm은 문헌과 불질의의 유사도를 거리로 측정하는 기능에 근거한다. 문헌과 질의(A or B)의 유사도는 P라는 특별계수의 통제아래 점(0.0)에서 질의(A or B)까지의 거리, 질의(A and B)의 유사도는 점(1.1)에서 질의(A and B)까지의 거리로 측정된다(A와 B의 가중치는 0~1 사이값). 따라서 n차원의 벡터공간에서 문헌과 질의의 유사도는 다음과 같이 정의된다.

Sim(D, Q_(q1 or ...qn))=

$$\left[\frac{d_1^p q_1^p + d_2^p q_2^p + \dots + d_n^p q_n^p}{q_1^p + q_2^p + \dots + q_n^p} \right]^{1/p}$$

Sim(D, Q_(q1 and ...qn))=

$$1 - \left[\frac{q_1^p (1-d_1)^p + q_2^p (1-d_2)^p + \dots + q_n^p (1-d_n)^p}{q_1^p + q_2^p + \dots + q_n^p} \right]^{1/p}$$

P=1이면 AND도 OR과 같은 유사도값을 갖게 된다. Salton, Fox, & Wu의 연구(1983)에서 P=2일 때 검색효과가 가장 좋았다. Savoy(1997)는 SIRE, 퍼지이론, Robertson의 기법, P-Norm 등을 비교하였다. SIRE나 퍼지기법보다 Robertson의 기법과 P-Norm이 우수했으며, P=1, 2, 5일 때 P-Norm이 Robertson 기법보다 우수하였다. P=1일 때 P-Norm은 가장 우수하였다.

불탐색에서 가중치를 사용하여 검색된 문헌을 순위화한 여러 방법을 비교평가한 국내 연구로는 이효숙(1993), 신은자, 정영미(1998), 고미영, 정영미(1999) 등의 논문이 있다. 이효숙(1993)은 P-Norm(p=2, 색인어=이진값, 탐색어의 가중치=적합성가중치(F4)와 역문헌빈도가중치)과 적합성가중치기법(AND, OR, NOT 구별 없이 모든 색인어의 가중치 ((di × (F4)공식)의 합), DNF(만족시키는 Atom의 수로 문헌을 순위화)를 비교하였다. 재현율에서는 차이가 없었으나, 정확률 측면에서는 영문 DB에서는 DNF < P-Norm < 적합성가중치의 순으로, 한글 DB에서는 DNF < 적합성가중치 < P-Norm의 순으로 좋은 성능을 보였다. 이용자가 판정한 순위와 가장 비슷한

순위를 보인 것은 영문·한글DB 모두에서 DNF < P-Norm < 적합성피드백 순위였다. P-Norm에서 탐색어의 가중치로 적합성가중치를 사용할 때가 역문헌빈도가중치를 사용할 때보다 더 좋은 결과를 보였으나 통계적으로 유의수준의 것은 아니었다. 그러나 신은자, 정영미(1998)의 피드백 연구는 초기탐색에서 DNF(용어의 가중치는 적합성가중치 (di × log($\frac{N-n+0.5}{n+0.5}$)))가 P-Norm (P=2, 탐색어가중치=역문헌빈도 log N/n)보다 우수하다고 보고하였다.

2.2 위치정보 기반의 순위화

블리언시스템에서 AND연산자 대신 인접연산자(n, r, w, adj, with 등)는 용어 간의 거리나 탐색필드를 제한함으로써 정확률을 향상시키는 도구로 사용되고 있다. 블리언시스템의 인접연산자의 기능에서 유추하여 비블리언시스템에서도 용어간의 거리와 출현하는 질의 용어수를 기반으로 문헌을 순위화하려는 노력이 수행되었다(Keen 1991, 1992; Clarke, Cormack & Kuntz 2000).

Clarke, Cormack & Kuntz(2000)는 web에서 1~3개 단어로 이루어진 짧은 질문으로 검색된 문헌을 단어간의 거리를 기반으로 순위화하는 것이 OKAPI의 적합성확률모델이나 벡터코사인모델, WCM순위화보다 우수하다고 보고하였다. 이 연구에서는 문헌에 담긴 질문용어의 종류 갯수로 먼저 문헌을 그룹화하고 같은 그룹내의 문헌은 탐색어 사이의 거리로 가중치를 계산하였다. 예를 들어 비블리언 탐색식 (A, B, C) 세 단어를 탐색하였을 때 먼저

3단어가 모두 출현하는 문헌과, 2단어가 출현하는 문헌, 1단어가 출현하는 문헌으로 문헌을 순위화한 후, 다시 단어가 출현하는 거리로 다음과 같이 순위화하였다.

1. 3단어 모두 출현하는 문헌그룹에서 문헌의 가중치

$$(-A-B-C) : (2, 6) = \frac{K}{5}$$

$$(-AB-CB-C-A) = (2, 5)(6, 10) = \frac{K}{4} + \frac{K}{5}$$

2. 2단어가 출현하는 문헌그룹에서 문헌의 가중치

$$(A-B-) : (1, 4) = \frac{K}{4}$$

$$(-A-B-A) : (2, 4)(4, 7) = \frac{K}{3} + \frac{K}{4}$$

$$\begin{aligned} (-A-B-A-B) : (2, 4)(4, 6)(6, 9) \\ = \frac{K}{3} + \frac{K}{3} + \frac{K}{4} \end{aligned}$$

3. 1단어가 출현하는 문헌그룹에서 문헌의 가중치

$$(A-A) = (1, 1)(4, 4) = 2K$$

3 실험설계

3.1 정렬평가 실험

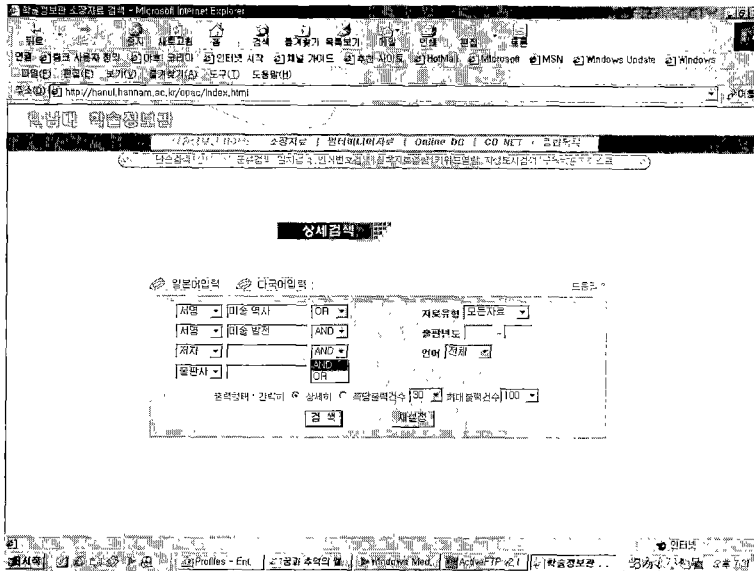
본 연구는 두 개의 실험과정으로 이루어졌다. 첫째는 현행 불리언시스템에서 서명순, 저자명순, 출판사순, 출판년도순 정렬을 비교평가하기 위한 것이다. 4가지 정렬이 가능한 한 남대학교 OPAC시스템(MAESTRO)이 사용되었다.

2000년 2학기 한남대학교 문헌정보학과 3

학년 교과목 “색인과 시소러스”를 수강한 학생 31명이 탐색질의를 가지고 탐색을 수행하였다.

탐색질의는 최소 1개 이상의 불연산자를 사용하도록만 요구되었지만 MAESTRO시스템이 갖는 구조적 문제때문에 더 제한되었다. 첫 문제는 MAESTRO의 DNF 질의의 표현이다. 즉 A and (B or C)라는 탐색식은 (A and B) or (A and C)로 바꾸어 탐색하여야 한다. MAESTRO에서 스페이스(space)는 AND 기능을 하기때문에 “미술 AND (역사 OR 발전)”이라는 불리언탐색식은 <그림 1>의 화면에서와 같이 입력되어야 한다. <그림 1>에서 보는 바와 같이 입력화면에서 OR로 연결할 수 있는 아톰 수는 4개로 제한되고 있다. 따라서 (A or B) and (C or D or E)와 같은 탐색식은 6개의 아톰으로 조합되므로 탐색이 불가능하다. 또 다른 문제점은 AND와 OR 연산자만 사용가능하다는 점이다. 따라서 본 연구에서 사용된 탐색식은 AND와 OR만 사용되고 NOT은 사용될 수 없었다. 실험에 참가한 이용자들은 DNF 탐색형식에 관해서는 충분히 숙지되었다.

탐색은 상세검색, 자료는 단행본으로, 검색건수는 100건으로, 쪽당 출력은 30건으로 제한하여 서명탐색을 수행하도록 요구되었다. 검색된 최대 100건의 문헌은 각각 서명, 저자명, 출판사명의 오름차순과 출판년도의 내림차순으로 정렬하여 간략정보 리스트를 출력하도록 하였다. 단행본탐색으로 검색된 문헌중 학위는 문은 제외하고, 단행본 도서만 대상으로 상위 20개 문헌의 순위를 다시 부여하고 적합성을 판정하도록 하였다. 같은 책이 소장위치가 달라 다른 레코드로 기록되어 2번 이상 검색된 단행본 도서는 한 건으로 취급되었다. 질문당



〈그림 1〉 MAESTRO 상세 검색 화면

검색된 문헌이 20개 미만일 경우에는 검색된 모든 문헌을 대상으로 하였다.

적합성은 상세정보(MARC 레코드)를 사용하여 판정하도록 하였다. 순위리스트에서 적합 문헌에 체크(V)한 후, 적합문헌중 대출희망문헌에 다시 동그라미(Ⓢ)를 치는 방법으로 부적합, 적합, 대출희망으로 기록하도록 하였다. 이용자가 평가한 적합성은 연구자에 의해 검토되었는데 몇몇 오류가 발견되었다. 전혀 관련없는 문헌이 적합문헌으로 판정되거나, 동일문헌이 서명순 리스트에서는 “대출희망”으로 평가되었지만 출판년도순 리스트에서는 “적합”으로 평가되는 등 정렬별 리스트에서 다르게 판정되는 경우가 있었다. 대출희망여부는 적합문헌수에 따라 영향을 받을 수 있다고 이해되었다. 예를 들어 20권중 15권이 적합문헌일 경우와 5권이 적합문헌일 경우 특정 적합문헌 A를 대출희망할 확률은 5권의 경우가 15권의 경우

보다 상대적으로 높을 것으로 이해되었다. 그러나 정렬기법이나 순위화기법을 비교하는 연구에서는 통일된 적합성 정보가 필요하기 때문에 연구자의 객관적 판단에 의해 적합성 판정이 조정될 필요가 있다고 판단되어 이용자의 초기 판정을 고려하여 연구자에 의해 적합성은 다시 판정되었다. 이용자와 연구자의 판정결과 30% 이상의 문헌에서 이견이 보인 질문은 분석에서 제외시켰다. 분석대상이 된 총 20건의 질문중 5개의 질문에서 31건의 적합성판정이 수정되었으나 이는 전체 1404건(=351건×4개 정렬)중 2.2%에 해당되는 수준이었다(적합성이 수정된 하나의 문헌이 3개 정렬리스트에 공통으로 출현한 경우 3건으로 계산했음).

3.2 실험2 : 순위화평가 실험

두 번째 실험은 적합성순위화 알고리즘을 평

가하기 위해 실행되었다. 실험1의 4가지 정렬 방법중 가장 정확률이 높은 정렬방법으로 판정된 출판년도순 정렬로 검색된 문헌(최대 20건)을 MARC형식으로 출력하여 문헌에 출현하는 탐색단어의 위치와 빈도수를 기반으로 하는 4가지 알고리즘으로 문헌의 순위를 다시 부여하였다.

문헌에 출현하는 탐색용어의 위치와 빈도수는 기본적으로 MAESTRO에서 사용하는 자동 색인추출기법에 기반하여 산출되었다. MAESTRO에서 사용된 형태소분석기는 내장된 명사사전과 동사사전에 일치하는 단어가 있으면 키워드로 추출하고 띄어쓰기 단위로도 추출하기 때문에 “음식으로 본 서양문화”라는 서명에서 “음식”과 “문화”의 위치는 (1, 4), “그림 놀이를 통한 어린이 심리치료”에서 “놀이”와 “치료”의 위치는 (2, 6), “놀이치료”나 “놀이 치료”에서는 인접용어(1, 2)로 간주하였다. 불용어는 인정하지 않아 “본”, “통한”, “대한” 등을 한 단어로 계산하였으나 “및”은 제외하였다.

1. R1(필드+빈도수)

MAESTRO는 DNF형식으로 탐색질문을 수행하는 시스템이기 때문에 A and (B or C)는 (A and B) or (A and C)와 같이 2개의 아톰으로 입력된다. R1은 하나의 아톰 내에 출현하는 용어의 위치에 따라 4등급으로 나눈 후, 같은 그룹에 속한 문헌은 탐색용어의 출현빈도수로 순위화한다.

- 1등급 : 아톰 내 AND로 조합된 용어가 모두 245필드의 서명 혹은 부서명 (\$a와 \$b)에 함께 출현하면서 인접한 문헌
- 2등급 : AND로 조합된 용어가 모두 245필

드의 서명 혹은 부서명에 출현하나 인접하지 않은 문헌

3등급 : AND로 조합된 용어가 245를 제외한 하나의 서명관련 필드(440, 500, 490, 730, 8XX)에 함께 출현하거나, 내용주기 필드(505)에서 같은 권(vol)이나 장(chapter)의 서명에 함께 출현한 문헌

4등급 : AND로 조합된 용어가 서로 다른 필드에 출현하거나 505필드에서 서로 다른 권호나 장에 출현하는 문헌
같은 등급에 속하는 문헌은 문헌 내 탐색용어의 총 출현빈도수로 가중치를 준다.

2. R2(문헌 내 탐색용어의 총빈도수)

문헌 내에 출현하는 탐색용어의 총출현빈도수로 순위화한다. R1에서 등급을 무시하고 출현빈도수만 사용한 것과 같다.

3. R3(DNF기반 가중치)

페이지론에 기반을 두어 연산자 AND로 연결된 두 용어는 최소값을 선택하나, OR로 연결된 용어는 최대값을 선택하지 않고 그 합(sum)을 취함으로써 DNF화하였다. (A or B) and C로 검색된 문헌에 A 1번, B 1번, C가 2번 출현한다면 OR로 연결된 용어는 최대값을 취하면 이 문헌의 가중치는 1이나, 합을 취하면 2가 된다. 이것은 A and C와 B and C가 모두 출현하므로 ATOM=2의 가중치를 갖는 것과 같다. OPAC문헌의 특성상 하나의 용어는 서명에 두 번 이상 출현하기 어렵고, OR로 연결된 용어가 각각 모두 출현한다면 한번만 출현한 문헌보다는 더 적합문헌일 가능성이 높을

것이다는 가정 때문이다.

4. R4(필드+거리)

Clarke의 위치정보 순위화기법을 응용한 방법이다. AND로 연결된 용어의 출현필드에 따라 등급으로 나누고, 등급 내 문헌은 위치정보로 순위화한다.

1등급 : AND로 연결된 모든 용어가 서명 245필드에 출현하는 문헌(R1의 1~2등급 문헌). 245필드에 출현하면 K=1로 가중치를 계산하고, 기타 서

명 관련 필드에도 출현하면 $K = \frac{1}{2}$ 로 계산된 가중치를 추가한다.

2등급 : 245필드를 제외한 다른 서명 관련 필드에 AND로 조합된 용어가 모두 출현할 경우 혹은 505필드에서 같은 권호의 서명에 출현할 경우(R1의 3등급 문헌). 1등급과 같은 가중치방법을 사용한다(K=1).

3등급 : AND로 연결된 용어가 각각 서로 다른 서명 관련 필드에 출현할 경우 혹은 505필드에서 서로 다른 권호의

〈표 2〉 용어의 출현위치에 따른 4가지 가중치 비교

탐색식	용어 출현 빈도			문헌의 가중치				
	245(\$a-\$b)	440	501	R1	R2	R3	R4	
A or B	(A-)			1등급 1	1	1	1등급 1	
	(A-A)			1	2	2	2	1 (1+1)
	(A-B)			1	2	2	2	1 (1+1)
	(A-)	(B-)		1	2	2	2	1 (1+1/2)
		(A)		3	1	1	1	3 1/2
A B	(ABAB)		(A)	1등급 5	5	2	1등급 (1/2 + 1/2 + 1/2)	
	(AB--A)			1	3	3	1	1 (1/2 + 1/4)
	(A-B)	(A--B)		2	4	4	2	1 (1/3) + (1/2 × 1/4)
	(A-)	(B)		3	2	2	1	3 0
		(B)	(A)	4	2	2	1	3 0
(A B) or (A C)	(AB--B)		(B)	1등급 4	4	1	1등급 1/2	
	(ABA-C)			1	4	4	2	1 (1/2 + 1/2 + 1/3)
	(A-B)	(A-C)		2	4	4	2	1 (1/3) + (1/2 × 1/3)
		(A--C)		3	2	2	1	2 1/4
		(A)	(C)	4	2	2	1	3 0
A B C	(ABC)			1등급 3	3	1	1등급 1/3	
	(A-BCC)	(ABC)		2	7	7	2	1 (1/4) + (1/2 × 1/3)
		(AB-C)		3	3	3	1	2 (1/2 × 1/4)
	(A-B)	(C)		4	3	3	1	3 1/3
	(A)	(B-C)		4	3	3	1	3 (1/2 × 1/3)
	(B)	(C)		4	3	3	1	3 0

서명에 출현한 문헌(R1의 4등급).

1) 두 용어로 조합된 탐색식 (A and B)의 경우, 505필드 내에 출현하면가중치는 1점, 기타 다른 필드에 각각 출현하면 가중치는 0점.

2) 세 용어로 조합된 탐색식 (A and B and C)의 경우, 하나의 필드에 두 용어가 출현하고 다른 필드에 나머지 하나가 출현하면 1등급에서와 같은 방법으로 두 용어의 가중치만 부여한다. 세 용어가 모두 다른 필드에 출현하면 가중치는 0점.

(표 2)는 4가지 문헌순위화방법별 문헌의 가

중치가 탐색식과 용어의 출현위치에 따라 어떻게 산출되는지를 보여준다.

3.3 성능측정척도

성능은 6가지 척도에 의해 평가되었다. P, P1, P3, P4에서는 부적합(0점), 적합(1점)의 2진 적합성판정이, P2와 P5에서는 부적합(0점), 적합(1점), 대출희망(2점)으로 3진 가중치를 사용하였다. 3진 적합성척도에서의 적합(1점)과 대출희망(2점)은 2진 척도에서는 모두 적합(1점)으로 간주되었다.

P : 검색된 최대 상위 20건의 문헌중 적합문헌수

P1 : Leighton & Srivastava(1999)의 연구에서 사용한 First-n P를 응용한 척도다. First-n P에서는 20개 문헌을 3개로 그룹지었지만, 본 연구에서는 4개로 그룹지어, 1등~5등 문헌중 적합문헌에는 10점, 6~10등 문헌중 적합문헌에는 8점, 11등~15등 문헌중 적합문헌에는 6점, 16~20등 문헌중 적합문헌에는 4점을 부여함으로써 등수간의 차이를 더욱 민감하게 하였다. 검색된 단행본이 20개가 되지 않을 때는 부족한 문헌수(×4점)을 분모에 반영하였다.

$$\frac{(1-5\text{등 적합문헌수})10+(6-10\text{등 적합문헌수})8+(11-15\text{등 적합문헌수})6+(16-20\text{등 적합문헌수})4}{140-(\text{부족문헌수})4}$$

P2 : P1의 공식에서 적합문헌에는 1점, 대출희망문헌에는 2점의 가중치를 준다.

$$\frac{(1-5\text{등 적합문헌수})10+(1-5\text{등 대출희망문헌수})10 \times 2 + \dots + (16-20\text{등 적합문헌수})4 + (16-20\text{등 대출문헌수})4 \times 2}{(140-(\text{부족문헌수})4)2}$$

P3 : 노정순(2000)이 제안한 적합문헌의 평균등수에 기반한 단순척도(r/e).

$$\frac{\text{적합문헌수}(\gamma)}{\text{적합문헌의 평균순위}} = \frac{\gamma \times \gamma}{\sum \text{적합문헌순위}}$$

P4 : P1보다 등수간의 차이를 더욱 민감하게 하기 위해 모든 순위에 서로 다른 점수를 부여한다. 1등 적합문헌에 20점, 2등 적합문헌에 19점, ... 20등 적합문헌에 1점을 부여한다.

$$\frac{\sum \text{적합문헌 점수}}{210 - \sum \text{부족문헌 점수}} = \frac{21 \times \text{적합문헌수} - \sum \text{적합문헌 순위}}{210 - \sum \text{부족문헌의 점수}}$$

P5 : P4의 공식에 적합문헌에는 1점, 대출희망문헌에는 2점을 부여한다.

$$\frac{\sum \text{적합문헌 점수} + \sum \text{대출희망문헌 점수}}{(210 - \sum \text{부족문헌의 점수})2} = \frac{(21(\text{적합문헌 수}) - \sum \text{적합문헌 순위}) + (21(\text{대출희망문헌수}) - \sum \text{대출희망문헌순위})}{(210 - \sum \text{부족문헌의 점수})2}$$

〈표 3〉은 15개의 문헌이 검색된 사례에서 6개 성능평가척도의 계산을 보여준다.

4 자료분석 및 가설 검증

실험에 참가한 31개의 탐색중 탐색요구와 탐색식이 다른 탐색(예, 문화를 문학으로 오타), 다른 학생과 동일 탐색을 수행한 탐색, 검색된 모든 문헌이 적합문헌(P=100%)인 탐색(예, 조기 교육), 이용자와 탐색자 간의 적합성 판정에서 30% 이상의 차이를 보인 탐색을 제외하고 20건의 탐색결과가 분석되었다. 20건의 질문은 모두 인문·사회분야 질문이었다. 〈부록 참조〉.

DNF형식으로 표현된 20개의 탐색식에는 총 38개의 아톰과 18개의 OR이 사용되어, 평

균 1.9개의 아톰이 0.9개의 OR로 연결되었다. A B or A C 형식의 탐색이 11개로 가장 많았고, A B C가 3개, A B, A B or C, A B or C D, A B or C or D, A B or A C or A D, A B C or A B D가 각각 1개씩 사용되었다.

학위논문과 중복된 문헌을 포함하여 20개 질문으로 총 1896건, 질문당 평균 94.8건(최고 804건, 최소 7건)의 문헌이 검색되었다. 7개의 질문에서는 단행본 중 학위논문과 중복된 단행본 도서를 제외하고 단행본 도서가 20건 미만이 검색되었다.

4.1 정렬필드별 성능(가설1) 검증 및 분석

〈표 4〉에서 보는 바와 같이 20개 질문의 상위 20개 문헌의 총수는 351건으로, 질문당 평

〈표 3〉 15개 단행본 도서가 검색된 사례에서 성능측정도 계산의 예

순 위	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
적합성	⊙	⊙	⊙		v	⊙		⊙	⊙	v	⊙			v	v
P	$\frac{11}{15} = 73.3\%$														
P1	$\frac{4 \times 10 + 4 \times 8 + 3 \times 6}{140 - 5 \times 4} = \frac{90}{120} = 75\%$														
P2	$\frac{3 \times 10 \times 2 + 1 \times 10 \times 1 + 3 \times 8 \times 2 + 8 + 1 \times 6 \times 2 + 2 \times 6 \times 1}{240} = 62.5\%$														
P3	$11 / (84 / 11) = \frac{11 \times 11}{84} = 1.44$														
P4	$\frac{20 + 19 + 18 + 16 + 15 + 13 + 12 + 11 + 10 + 7 + 6}{210 - (5 + 4 + 3 + 2 + 1)} = \frac{21 \times 11 - \sum \text{적합문헌순위}}{210 - 15} = \frac{21 \times 11 - 84}{195} = 75.4\%$														
P5	$\frac{(21 \times 4 - (5 + 10 + 14 + 15)) + 2(21 \times 7 - (1 + 2 + 3 + 6 + 8 + 9 + 11))}{390} = 65.1\%$														

균 17.55건이었다. 이 중 질문당 평균 8.9권의 적합문헌이 검색되었으며, 그 중 5.36권이 대출희망문헌이었다. 질문당 평균 적합문헌은 출판년도순이 9.80건, 서명순이 8.65건, 출판사순이 8.75건, 저자순이 8.40건이었다. 적합문헌 중 대출희망문헌은 출판년도순 6.20건, 서명순 5.30건, 출판사순 5.15건, 저자명순 4.80건이었다. 적합문헌수, 대출희망문헌수, P~P5의 정확률 척도 모두에서 출판년도가 가장 우수하였고, 그 다음 서명, 출판사순으로 우수했으며, 저자명순이 가장 나빴으나, 일원배치 분산분석(One-way ANOVA) 결과 유의한 차이($\alpha=0.05$)는 아니었다.

20개 질문 중 출판년도순은 9개 질문에서, 서명순은 4개 질문에서, 출판사순은 4개 질문에서, 저자명순은 3개 질문에서 가장 좋은 결과를 제공하였다. 질문당 가장 좋은 정렬방법을 분석한 이 분석에서는 P3~P5의 평균이 사

용되었다(이유는 5장에서 설명). 동일 저자명이나 출판사명이 한글과 한자에 따라 별도로 정렬되는 문제가 발견되었으나 분석에서는 무시되었다. 탐색질의가 최신정보에 민감하지 않는(예, 문헌이론이나 미술역사, 다도(차문화) 등) 주제를 제외하고 출판년도순 정렬이 가장 좋은 결과를 보였다.

4.2 적합성 순위화 알고리즘의 성능 (가설2) 검증 및 분석

순위화 알고리즘의 성능은 4개의 정렬방법 중 가장 우수한 결과를 제공한 출판년도순 정렬방법으로 검색된 문헌을 대상으로 테스트되었다. P는 출판년도순 정렬에서와 같으므로 P1~P5의 5개 성능평가척도로 4개의 순위화 알고리즘이 평가되었다.

〈표 4〉가설1 검증 요약

	서명	저자	출판사	년도	평균	F	P
상위 20문헌 총수	351	351	351	351	351		
질문당 평균 상위 20문헌수	17.55	17.55	17.55	17.55	17.55		
적합문헌 총수	173	168	175	196	178		
질문당 평균 적합 문헌수	8.65	8.40	8.75	9.8	8.9	.478	.699
대출희망문헌 총수	106	96	103	124	107.25		
질문당 평균 대출희망 문헌수	5.3	4.8	5.15	6.20	5.36	.789	.504
P	51.39	50.14	51.89	57.14	52.64	.363	.780
P1	50.23	48.05	50.19	56.35	51.20	.514	.674
P2	40.57	38.29	40.38	46.14	41.34	.593	.621
P3	.97	.94	.99	1.11	1.00	.551	.649
P4	51.81	49.14	51.55	58.12	52.65	.538	.658
P5	42.58	39.43	41.95	48.24	43.05	.663	.577

〈표 5〉가설2 검증 요약

	출판년도	R1	R2	R3	R4	F	P
P1 (향상률)	56.35	59.94 (+6.4)	55.18 (-2.1)	55.07 (-2.3)	60.44 (+7.3)	.317	.866
P2	46.14	50.66 (+9.8)	45.35 (-1.7)	46.19 (+0.1)	49.47 (+7.2)	.316	.866
P3	1.11	1.42 (+27.9)	1.08 (-2.7)	1.11 (0)	1.47 (+32.4)	4.919	.001**
P4	58.12	66.77 (+14.9)	56.71 (-2.4)	53.38 (+0.4)	67.90 (+16.8)	1.275	.285
P5	48.24	55.39 (+14.8)	47.07 (-2.4)	48.35 (+0.2)	56.36 (+16.8)	1.128	.348

〈표 5〉에서 보는 바와 같이 20개 질문의 평균을 분석한 결과, 4개의 순위화 알고리즘 중 R4와 R1로 순위화시킨 결과가 P1~P5 모든 측면에서 순위화시키지 않은 것과 비교해 성능을 향상시켰다. R4가 P1, P3, P4, P5 측면에서 가장 우수하였으나, P2 측면에서는 R1이 더 우수하였다. R2는 P1~P5 모든 측면에서 순위화시키기 전보다 나쁜 결과를 제공하였다. 일원배치분산분석(One-way ANOVA) 결과 순위화시키기 전과 4개의 순위화 결과 사이에는 P3 측면에서 유의한 차이($\alpha=0.001$)를 보이고 있으나, 다른 척도로는 유의한 차이를 보여주지 못하였다. P3에서 R4는 +32.4%, R1은 +27.9%, R2는 -2.7%의 향상률을 보이고 있다.

20개의 질문별 분석에 의하면 2개의 질문(질문 6, 질문 7)에서 4개의 순위화 알고리즘은 우수하지 못하였다. 다른 두 질문(질문 19, 질문 20)에서는 P3~P5 측면에서, 또 다른 한 질문(질문 18)에서는 P5 측면에서 평가할 때 순위화는 정렬보다 좋은 결과를 제공하지 못하

였다. 이 다섯 질문 중 네 질문은 정확률 P 측면에서 2, 3, 4, 7위의 높은 P를 가져온 질문으로, 평균 14건의 문헌을 검색하여 11.1건의 문헌이 적합문헌이었다. 모든 검색된 문헌이 적합문헌이면 순위화가 불필요하듯이, 정확률이 높은 질문보다는 낮은 질문에서 순위화가 더 효과적임을 보여주었다.

4개 순위화 알고리즘간의 효과는 P1~P5의 성능척도에 따라 차이를 보였다. 9개의 질문에서 4개 순위화 알고리즘간의 성능서열에는 P1~P2와 P3~P5 두 그룹간에 차이를 보였다. P3~P5가 P1~P2보다 판별력이 좋은 정확률 척도로 판단되어(다음장 참조) P3~P5의 평균으로 분석했을 때, R4는 10개의 질문에서, R1은 6개 질문에서, R2와 R3는 각각 1개의 질문에서 가장 좋은 결과를 제공했으며, 2개 질문에서 R1과 R4가 공동으로 가장 좋은 결과를 제공하였다. R2와 R3이 가장 좋은 순위화 결과를 제공한 것으로 나타난 질문15와 질문7을 제외하고 18개 질문에서 R4와 R1이 가장 좋은 순위화 결과를 제공하였다.

질문15에서 R1과 R4가 R2보다 좋은 결과를 제공하지 못한 이유는 “아동 심리 OR 놀이 치료”로 검색된 아동심리에 관한 문헌들이 부적합 문헌으로 판정되었기 때문이다. 이용자는 아동심리가 245필드에 인접 출현한 1등급 문헌들을 부적합 판정하고 놀이치료에 관한 문헌만 적합 판정하였다.

4.3 성능척도간의 차이(가설3) 검증 및 분석

〈표 6〉은 20개의 탐색질문으로 검색된 상위 20문헌을 4개의 정렬과 4개의 순위화 알고리즘으로 각각 출력한 결과를 피어슨의 상관계수로 6가지 척도간의 유사도를 측정 한 결과이다. 6가지 척도는 모두 유사한 것으로 나타났다($\alpha = 0.001$). 3진 적합성 판정을 이용한 P2와 P5는 서로 가장 유사도가 높았다. 2진 적합성 판정을 이용한 P1과 P3, P4 사이에서 P4는 P1과 P3 모두와 유사도가 가장 높았고, P3은 P4와, P1은 P4와 유사도가 가장 높았으나, P1과 P3 간의 유사도는 상대적으로 덜 높았다. P1과 P2, P4와 P5는 같은 공식에 적합성 판정만 2진과 3진으로 적용한 것이므로 유사도가 높

았으나, 같은 판정을 사용한 P1과 P4, P2와 P5 만큼은 아니었다.

20개의 질문 중 11개의 질문에서 P1~P5의 다섯 척도 모두에서 4개 순위화 사이의 성능은 동일하였으나, 9개 질문에서는 척도에 따라 순위화 성능에 차이를 보였다. 〈표 7〉은 5개의 순위화성능척도 P1~P5의 분별력을 분석하기 위하여 P1~P5에 따라 4개 알고리즘의 성능서열이 서로 다른 9개의 질문의 순위화기법에 따른 출력 순위와 P1~P5의 값을 요약한 것이다.

질문1에서 R1은 3개의 대출희망문헌을 1, 2, 3위에, 다른 3개의 대출희망문헌은 4~7위(평균 5.5위)로 순위화시켰으나, R4는 대출희망문헌 1개만 1위로 출력했을 뿐 기타 대출희망문헌 5개와 부적합문헌 14개는 판별하지 못하였다. 그럼에도 P1과 P2는 R4가 R1보다 우수하다고 평가하였다.

질문9에서는 R4는 적합문헌 3개를 5, 6, 8위에, 대출희망은 15위로 출력하였으나, R1은 적합문헌 3개는 모두 8.5위에, 대출희망은 16.5위에 출력하였다. R4가 R1보다 우수한 결과임에도 P1과 P2는 R1이 같거나 더 우수하다고 평가하였다.

질문10에서 R1과 R4는 3개의 부적합문헌

〈표 6〉 성능척도간의 유사도

	P값	P1	P2	P3	P4	P5
P값						
P1	.938**					
P2	.882**	.946**				
P3	.840**	.901**	.854**			
P4	.949**	.953**	.900**	.953**		
P5	.903**	.916**	.952**	.911**	.951**	

을 최하위 18~20등으로 출력함으로써 R2와 R3보다 우수하나, P2는 R3이 가장 우수한 것으로 잘못 평가하고 있다.

질문14에서 R1은 대출희망문헌을 1.5, 1.5, 4.5, 4.5, 11위에, 적합문헌은 7, 8위로 출력하였으나, R4는 대출희망문헌을 2.5, 2.5, 5, 5, 11위에, 적합문헌은 9, 10위에 출력하였다. R1이 R4보다 우수하나 P1과 P2는 동등하게 평가하였고, P3~P5는 R1이 더 우수하다고 평가하였다.

질문16에서 10~17위로 출력되는 부적합문헌 외에 또 다른 부적합문헌 1건을 R1과 R4 모두 6.5위로 출력시키나, R1은 대출희망도서를 1, 2.5, 4, 6.5위에, R4는 3.5, 3.5, 3.5, 8위에 출력시켰다. P1과 P2는 R4가 더 우수한다고 잘못 평가하고 있다.

질문17에서도 부적합문헌 1건이 R1에서는 3.5위에, R4에서는 5.5위에 출력하기 때문에 R4가 더 우수해 보이나, P1과 P2는 R1이 더 우수하다고 평가하고 있다.

질문18에서 R4는 전체 검색된 문헌 6개 중 부적합문헌 1개를 5.5위에 출력하였으나, R2는 적합·부적합문헌을 판별하지 못하였다(평균 3.5위). 그러나 P1~P2는 R2가 R4보다 더 우수하다고 평가하였다.

질문19에서 R1은 검색된 16개의 문헌 중 2건의 부적합문헌을 7.5위, 16위에 출력시키나 R2는 모두 10위에 출력시켰다. R1이 R2보다 우수해 보이나, P1과 P2는 R2가 더 우수하다고 평가하고 있다. 이상과 같이 R2와 R3 중 어느 것이 더 우수한 순위화결과를 제공하는지 명확하지 않은 질문7을 제외한 나머지 8개의 질문에서 P1~P2보다는 P3~P5가 더 좋은 순위

화 판별력을 가지고 있는 것으로 설명되었다.

4.4 적합문헌에서 탐색용어의 출현 분석

연구과제6을 규명하기 위해 출판년도순으로 출력된 351건의 문헌 중 적합문헌 196건에서 탐색용어의 출현위치가 분석되었다.

1) 부서명

서명탐색시 부서명(245필드 \$b)은 본서명(245필드 \$a)과 분리될 필요가 있는가? 전체 20건의 질문으로 검색된 351건의 문헌 중 탐색용어가 245필드에 출현한 R1의 1등급과 2등급 문헌은 227건(적합문헌은 171)이었다. 이 중 199건(적합문헌은 154건)의 문헌에서 탐색용어는 본서명(\$a)에 출현하였다. 탐색용어가 본서명(\$a)과 부서명(\$b)에 분산 출현한 문헌은 총 28건(적합문헌은 17)이었다. 3건의 문헌(모두 적합문헌)에서는 \$a와 \$b에 인접출현하였고, 13건의 문헌(적합문헌은 8건)에서는 \$a와 \$b에 떨어져 출현하였으며, 12건의 문헌(적합문헌은 6건)에서는 \$b에만 출현하였다. 특히 2개의 AND로 조합된 4개의 질문에서 3개의 탐색용어가 모두 245필드에 출현한 문헌은 총 26건(적합은 25건)이었는데, \$a에만 출현한 문헌은 18건(적합문헌 18)이었고, \$a와 \$b에 산재된 문헌은 8건(7건이 적합문헌)이었다.

2) 245필드 이외의 서명관련필드

245필드 이외의 서명관련필드에 출현한 탐색용어 때문에 검색된 문헌은 124건으로 전체 351건 중의 35%를 차지하였다. 그러나 124건

중 적합문헌은 25건으로 전체 적합문헌의 12.8%(25/196)를 차지하였다. R1에서 3등급 문헌으로 평가된 문헌은 42건(적합문헌은 10), 4등급으로 평가된 문헌은 82건(적합문헌은 15건)이었다.

3등급문헌으로 검색된 적합문헌 중 5건에서 탐색용어는 245필드와 440 혹은 505필드에 분산되었고, 3건은 505필드에, 2건은 500필드에 출현하였다. 4등급으로 평가된 적합문헌 15건 중 5문헌에서 탐색용어는 245필드와 440 필드에, 2문헌에서는 245필드와 490필드에, 1문헌에서는 245필드와 500필드에, 1문헌에서는 245필드와 505필드에, 3문헌에서는 245필드와 830필드에, 3문헌에서는 505필드 내 다른 권(혹은 장)에 분산되어 있었다.

245필드의 기타 서명관련필드에 출현하는 용어까지 서명단어 검색에 포함시킴으로써 12.8%의 적합문헌을 더 검색하였으나 정확률은 75.3%($= (196-25)/(351-42-82)$)에서 55.8%($= 196/351$)로 하락시켰다. 245필드로 제한하여 서명탐색한다면 12.8%의 적합문헌을 덜 검색하나 정확률은 20% 향상시킬 수 있을 것이다.

5 연구결론 및 논의

본 연구에서 OROM의 MAESTRO를 사용한 서명단어탐색은 평균 52.64%의 정확률(P)을 보였다. 질문당 최고 20개의 문헌만 출력했을 때 8.9권이 적합문헌이었고 이용자는 적합문헌 중 5.36건만을 대출하기 원하였다.

서명단어탐색 결과 간략서지를 서명, 저자

명, 출판사, 출판년도 순서로 정렬하였을 때 유의한 수준은 아니었지만(귀무가설1 채택), 출판년도순이 가장 좋은 탐색결과를 보였다. 이 결과는 플로리다주립대 학생을 대상으로 OPAC의 화면설계에 관한 김미현(1998)의 실험실 실험(Laboratory Experimental)의 결과와는 상이한 것이었다. 실험용으로 만들어진 저자명탐색과 서명탐색, 주제명탐색의 결과 서명순 정렬이 출판년이나 저자명보다 성공률이 높은 것으로 보고되었다. 그러나 초록을 포함한 서지DB(CACM과 CISI)를 대상으로 정렬을 비교한 Savoy(1997)의 연구결과와는 동일하였다. Savoy의 연구에서 서명순 정렬은 출판년도 정렬보다 정확률 -8.4%(CACM)와 -1.3%(CISI)의 하락을 초래하였다. 김미현의 연구에서는 본 연구나 Savoy 연구와는 달리 저자명탐색과 주제명탐색까지 포함하였다.

IR 연구에서와 마찬가지로 본 연구에서 순위화 알고리즘은 저자, 서명, 출판사 혹은 출판년도에 의한 단순 정렬보다 정확률을 향상시켰다(P3 측면에서 귀무가설2는 기각). 탐색용어의 위치와 빈도를 이용한 순위화 알고리즘(R1과 R4)이 빈도만을 이용한 알고리즘(R2와 R3)보다 우수하였다. 그러나 OPAC에서 탐색용어의 단순빈도에 의한 순위화(R2)와 DNF(R3) 순위화방법은 초록을 대상으로 한 연구에서처럼 정확률을 향상시키지 못하였을 뿐만 아니라, 오히려 저하시키기까지 하였다(R2). 위치정보를 이용한 Clarke, Cormack & Tudhope(2000)의 순위화(R4)는 web전문(Full Text)에서와 마찬가지로 OPAC에서도 우수한 순위화 결과를 제공하였다. 초록을 포함하는 한글 정보과학분야 KT 테스트set을 대

상으로 상위 30건으로 문헌수를 제한하여 순위화기법을 비교 평가한 고미영, 정영미(1999)의 연구에서 필드가중치는 통계적 빈도기법보다 많은 적합문헌을 출력하였다. 필드가중치를 근접거리와 함께 사용한 순위화보다는 가중치(단어빈도×역문헌빈도)와 함께 사용할 때 가장 우수하였다. 본 연구에서는 필드가중치를 근접거리와 함께 사용한 R4가 가중치(단순빈도)를 함께 사용한 R1보다 조금 더 우수하였다.

6개의 순위화 평가척도간에 차이는 없었지만($\alpha=0.001$), P, P1, P2보다는 P3, P4, P5가 보다 판별력이 있는 것으로 분석되었다. Leighton & Srivastara(1999)의 First-n P는 노정순(2000)이 제안한 바와 같이 상위 20문헌을 1~5등, 6~10등, 11~15등, 16~20등으로 그룹화하여 각각 10점, 8점, 6점, 4점을 부여하는 것(P1)보다 1등부터 20등까지 모든 순위에 서로 다른 점수(20점부터 1점까지)를 부여하는 것(P4)이 더 판별력을 갖는 것으로 분석되었다. 노정순(2000)의 연구에서 제안된, 계산방법(적합문헌수/적합문헌의 평균 순위)이 간단한 척도 r/e (=P3)는 P4와 가장 유사한 좋은 판별력을 지닌 순위화 평가척도임을 알 수 있었다.

적합문헌에서 탐색용어의 출현위치를 분석한 결과 196건의 적합문헌 중 78.6%(154건)에서 탐색용어는 본서명에 출현했으며, 8.7%의 적합문헌(17건)에서는 부서명에, 12.8%(25건)의 적합문헌에서는 기타 서명에 출현하였다. 이는 서명단어 탐색시 모든 서명을 대상으로 하는 대신 본서명으로 제한하면 재현율은 21.4%(42/196) 저하하나 정확률은

21.54%(154/199 - 196/351) 향상되고, 본서명과 부서명으로 제한하면 재현율은 12.8%(=25/196) 저하하나 정확률은 19.5%(171/227 - 196/351) 상승됨을 의미한다. 본서명과 부서명으로 탐색을 제한하면 25건의 적합문헌은 검색이 안되지만 전체 부적합문헌 155건(=351-196) 중 124건이 검색이 안되기 때문에 순위화 알고리즘보다 높은 정확률을 향상 수단이 될 수 있다. 이는 상위 20문헌으로 제한했을 때의 데이터이기 때문에, 20건으로 제한하지 않고 100권 이상의 문헌이 검색됐을 때는 순위화보다는 245필드로 탐색을 제한하는 것이 보다 효과적이고 효율적인 탐색수단이 될 수 있음을 보여준다.

본 연구는 불논리 탐색문과 도치색인화일을 사용하는 현행 OPAC 시스템에서 검색성능을 향상시킬 수 있는 순위화 기법을 연구하였다. 본 연구의 결과는 OPAC 시스템에서 다음과 같이 응용될 수 있을 것이다.

- 1) 탐색결과를 정렬하는데 디폴트(default) 정렬방법을 출판년도순으로 지정.
- 2) 탐색결과를 정렬하는 옵션에 적합성순을 추가.
- 3) 적합성순으로 정렬할 경우에 빈도수보다는 위치정보에 기반하는 알고리즘 사용.
- 4) 질문당 검색된 문헌수가 평균 100권에 육박하는 OPAC환경에서는 서명탐색을 245필드로 제한하는 옵션 추가.

끝으로 본 연구는 인문·사회분야 20개의 탐색질의로 검색된 문헌을 100건으로 제한하여 정렬하였고, 출판년도순으로 정렬된 상위 20건의 문헌만을 대상으로 적합성 순위화를 수행하였기 때문에, 전체 주제를 대상으로 보

다 많은 탐색을 수행하고, 검색된 전체 문헌을 대상으로 정렬과 순위화를 수행하여 연구결과

를 일반화할 필요가 있을 것이다.

참고 문헌

- 고미영, 정영미. 1999. P-norm 검색의 문헌 순위화기법에 관한 실험적 연구. 『정보관리학회지』, 16(1): 7-30.
- 김미현. 1998. 온라인목록시스템의 사용성에 관한 연구(화면설계를 중심으로). 『정보관리학회지』, 15(1): 43-61.
- 김현희. 1994. 퍼지정보검색시스템 설계. 『명지대 인문과학연구논총』, 11(94. 1): 177-206.
- 노정순. 1999. 탐색결과에 근거한 자연어질의 자동확장 및 응용에 관한 연구 고찰. 『정보관리학회지』, 16(2) : 49-80.
- 노정순. 2000. 순위화시스템의 효과측정척도에 관한 연구. 『정보관리학회지』, 17(4): 67-81.
- 신은자, 정영미. 1998. 피드백 정보를 이용한 불논리 검색시스템의 성능 증진에 관한 실험적 연구. 『정보관리학회지』, 15(1): 129-148.
- 이효숙. 1993. 적합성 가중치 검색 및 P-norm 검색에 관한 연구. 박사학위논문. 이화여자대학교.
- Angione, Pauline V. 1975. "On the Equivalence of Boolean and Weighted Searching Based on the Convertibility of Query Forms." *JASIS*, 26: 112-124.
- Barbuto, D. M. & E. E. Cevallos. 1991. "End-user Searching: Program Review and Future Prospects." *RQ*, 31(winter 1991): 214-227.
- Bing, J. & K. S. Selmer. 1980. *A Decade of Computer and Law*. Oslo, Norway: Norwegian University Press.
- Blecic, D. D., J. L. Dorsch, M. H. Koenig, & N. S. Bangalore. 1999. "A Longitudinal Study of the Effects of OPAC Screen Changes on Searching Behavior and Searcher Success." *College & Research Libraries*, 60(6): 515-530.
- Bookstein, A. 1978. "On the Perils of Merging Boolean and Weighted Retrieval Systems." *JASIS*, 29: 156-158.
- Carlyle, Allyson. 1996. "Ordering Author and Work Records: an Evaluation of Collocation in Online Catalog Displays." *JASIS*, 47(7): 538-554.
- Chignell, M. H., J. Gwizdka, & R. C. Bodner. 1999. "Discriminating

- Meta-Search: a Framework for Evaluation." *Information Processing & Management*, 35: 337-362.
- Clarke, C. L. A., G. V. Cormack & E. A. Tudhope. 2000. "Relevance Ranking for One to Three Term Queries." *Information Processing and Management*, 36: 291-311.
- Fox, E. & S. Sharat. 1986. *A Comparison of Two Methods for Soft Boolean Interpretation in Information Retrieval*. Technical Report TR-86-1, Virginia Tech, Dept. of Computer Science.
- Fox, E., S. Betrabet, M. Koushik, & W. Lee. 1992. "Extended Boolean Models." In W. B. Frakes & R. Baeza-Yates(Ed.), *Information Retrieval: Data Structures & Algorithms*(pp.393-418). Englewood Cliffs, New Jersey: Prentice Hall.
- Hunter, Rhonda N. 1991. "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction Log Analysis." *RQ*, 30(Spring 1991): 395-402.
- Keen, E. M. 1991. "The Use of Term Position Devices in Ranked Output Experiments." *J of Documentation*, 47(1): 1-22.
- . 1992. "Term Position Ranking: Some New Test Results." *Proceedings of the 15th Annual International ACM SIGIR Conference on Research & Development in Information Science*, 66-76.
- Larson, Ray R. 1986. "Workload Characteristics and Computer System Utilization in Online Library Catalog." Ph. D. dissertation, Univ. of California, Berkeley.
- Larson, Ray R. 1991. "The Decline of Subject Searching : Long-Term Trends and Patterns of Index Use in an Online Catalog." *JASIS*, 42: 197-215.
- Larson, Ray R. 1992. "Evaluation of Advanced Retrieval Techniques in an Experimental Online Catalog." *JASIS*, 43: 34-53.
- Larson, R. R., J. Mc Donough, P. O'Leary, L. Kuntz, & R. Moon. 1996. "Cheshire II : Designing a Next-Generation Online Catalog." *JASIS*, 47(7): 555-567.
- Leighton, H. V. & J. Srivastava. 1999. "First 20 Precision Among World Wide Web Search Services (Search Engines)." *JASIS*, 50(10): 870-881.
- Lynch, C. A. 1989. "Large Database and Multiple Database Problems in

- Online Catalogs." In *OPAC and Beyond : Proceedings of a Joint Meeting of the British Library, DBMIST and OCLC, Dublin, Ohio: OCLC, 1989.*
- Mitra, M., A. Singhal, & C. Buckley. 1998. "Improving Automatic Query Expansion." *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Science*, 206-214.
- Noreault, K. M. & M. J. McGill. 1977. "Automatic ranked output from Boolean Searches in SIRE." *JASIS*, 28: 333-339.
- Peter, T. A. & M. Kurth. 1991. "Controlled and Uncontrolled Vocabulary Subject Searching in an Academic Library Online Catalog." *Information Technology and Libraries*, 10(sept. 1991): 201-211.
- Radecki, T. 1982. "A Probabilistic approach to Information Retrieval in Systems with Boolean Search Request Formulations." *JASIS*, 33(6): 365-370.
- Robertson, S. E. 1978. "On the Nature of Fuzz: A Diatribe." *JASIS*, 29: 304-307.
- Savoy, J. 1997. "Ranking Schemes in Hybrid Boolean Systems: a New Approach." *JASIS*, 48(3): 235-253.
- Salton, G., E. Fox, & U. Wu. 1983. "Extended Boolean Information Retrieval." *Communication of the ACM*, 26: 1022-1036.
- Voorbij, Henk J. 1998. "Title Key-words and Subject Descriptors: a Comparison of Subject Entries of Books in the Humanity and Social Science." *J. of Documentation*, 54(4): 466-476.
- Wallace, Patricia A. 1993. "How Do Patrons Search the Online Catalog When no one's Looking? Transaction Log Analysis and Implications for Bibliographic Instruction and System Design." *RQ*, 33(2): 239-252.
- Wiberley, S. E., R. A. Daugherty & J. A. Danowski. 1990. "User Persistence in Scanning Postings of a Computer-driven Information System: LCS." *Library & Information Science Research*, 12: 341-353.
- Wilberley, S. E., R. A. Daugherty & J. A. Danowski. 1995. "User Persistence in Displaying Online Catalog Posting: LUIS." *LRTS*, 39(3): 247-264.

〈부록〉 : 탐색질문과 탐색식

- 질문
1. 문학사상과 이론에 관한 연구
 2. 기독교 사회와 기독교 문화에 관한 자료
 3. 한국문학의 전체적인 흐름을 알 수 있는 단행본
 4. 일본문화와 일본문화의 개방에 관한 자료
 5. 교육경영과 행정에 관한 자료
 6. 한국의 경제정책 및 계획에 관한 자료
 7. 아동교육이나 유아교육에 관한 자료
 8. 한국사회나 문화에 관한 자료
 9. 미술의 역사와 발전에 관한 자료
 10. 동양철학과 서양철학에 대한 교양과목 레포트 작성을 위한 자료 수집
 11. 세상을 보는 눈을 넓히기 위해 세계 여행과 세계 문화에 관한 정보
 12. 우리나라의 과거 · 현재 · 미래의 보건정책이나 보건방안
 13. 중국의 정치와 경제
 14. 우리나라의 교육 정책
 15. 아동의 심리를 파악하고 이해하기 위한 방법으로, 놀이로써 치료하는 것은 어떤 영향과 효과를 나타내는가
 16. 일본문화와 한국문화의 차이점, 영향
 17. 각국의 차문화(다도)
 18. 기독교인의 자녀교육
 19. 각국의 음식문화
 20. 북한의 정치 · 경제 · 사회 · 문화에 관한 문헌(중국의 영향, 중국과 비교)

- 탐색식
1. 문학 사상 OR 문학 이론
 2. 기독교 사회 OR 기독교 문화
 3. 한국 문학 연구
 4. 일본 문화 OR 일본 대중문화
 5. 교육 경영 OR 교육 행정
 6. 한국 경제 정책 OR 한국 경제 계획
 7. 아동 교육 OR 유아 교육
 8. 한국 사회 OR 한국 문화

9. 미술 역사 OR 미술 발전
10. 동양 철학 OR 서양 철학
11. 세계 여행 OR 세계 문화
12. 보건 정책 OR 보건 방안
13. 중국 정치 경제
14. 교육 정책 OR 교육 제도
15. 놀이 치료 OR 아동 심리
16. 한국 일본 문화
17. 차 문화 OR 다도 OR 다례
18. 기독교 자녀 OR 그리스도 자녀 OR 크리스찬 자녀
19. 음식 문화 OR 식생활
20. 북한 중국