

상이한 적합성 판정과 전문검색시스템의 평가에 관한 연구 *

Variations in Relevance Assessments and Evaluation of the Performance of Full-Text Retrieval System

문성빈(Sung-Been Moon)**

목 차

- | | |
|------------|-----------------------------|
| 1. 서 론 | 4.1 표준재현을 지점에서의 정확률 및 평균정확률 |
| 2. 이론적 배경 | 4.2 선정된 9개 질문서의 검색효율성 분석 |
| 3. 실험환경 | 5. 결 론 |
| 4. 연구결과 분석 | |

초 록

본 논문에서는 문헌의 전문을 대상으로 얻어진 4개의 상이한 적합성 판정 세트가 전문검색시스템의 검색효율성 평가에 미치는 영향을 살펴보았다. 각 적합성 판정 세트간의 검색효율성에는 주목할 만한 차이가 없는 것으로 밝혀졌다. 이는 적합성 개념에 대한 다양한 관점이 검색효율성의 평가에는 커다란 영향을 미치지 못하고 있음을 암시하는 것이다. 그러나, 적합성 정보를 효과적으로 이용하는 정교한 검색 기법인 적합성 피드백을 통합한 검색실험은 계속 연구되어야 할 과제로 제시하고 있다.

ABSTRACT

This study examined the extent to which variations in relevance assessments affect the evaluation of the performance of full-text retrieval system. Four sets of relevance judgments obtained by examining the full-text of documents were used to test the retrieval effectiveness. There was no noticeable difference in retrieval performance among the four relevance judgment sets. It implies that a variety of definitions of relevance has no effect on the evaluation of the performance of the full-text retrieval system. Further retrieval experiments on this topic incorporating relevance feedback, which is one of the sophisticated retrieval techniques using relevance information, are suggested.

* 이 논문은 1995년도 연세대학교 학술연구비의 지원에 의해 연구되었음.

** 연세대학교 문헌정보학과 조교수

■ 논문 접수일 : 1997년 10월 30일

1. 서론

정보검색시스템의 효율성을 평가하기 위한 척도로써 가장 많이 사용되는 정확률과 재현율은 적합성 개념을 그 기반으로 하고 있기 때문에 적합성(relevance)은 정보검색시스템의 효율성을 평가하는데 있어서 없어서는 안되는 기본적인 개념이다. 일반적으로 실험실내에서 행해지고 있는 정보검색실험을 위한 구성요소로서는 검색시스템, 문헌, 질문서, 그리고 문헌과 질문서간의 적합성 판정을 필수적인 것으로 보고 있다. 데이터베이스내의 문헌들과 질문서에 대한 적합성 여부는 그 주제에 대해 잘 알고 있는 주제전문가나 질문서를 작성한 이용자에 의해 미리 판정되어 문헌 및 질문서 파일과 함께 하나의 독립된 파일로써 저장되어 검색시스템의 평가를 위한 실험에 이용된다. 그러나 실험목적에 의해 사용되고 있는 질문서는 이용자의 실질적인 정보요구를 잘 반영하고 있지 못하며 문헌과 이러한 질문에 대한 적합성 판정은 실제 탐색환경에서 얻어질수 있는 적합성 판정을 대체할 수 없을 것이라는 우려가 있다. 이는 정보검색을 위한 실험실 연구의 부적절성에 대한 비판의 원인이 되기도 한다(Salton, 1992).

문헌과 질문간의 적합성 판정의 결과는 여러 가지 요소에 의해 영향을 받는데, 이는 적합성 판정의 과정, 즉 누구에 의해 행해졌는가와 또는 문헌의 어떤 요소를 판정의 기준으로 삼았는가에 따라 달라지곤 한다. 또한 질문서와 문헌간의 직접적인 적합성 관계와는 별도로 다른 외부 조건이 적합성 판정에 영향을 주기도 한다(Eisenberg, 1988; Eisenberg &

Barry, 1988). 즉, 적합성 판정은 질문서와 문헌간의 적합성 이외에 다른 요인들에도 매우 민감할 수 있음을 의미하는 것이다. 그러므로 심사자들간의 적합성 판정의 불일치는 자주 발견되었고 이것이 검색효율성의 평가에 미치는 영향은 몇몇 정보과학자들의 관심을 끌게 되었다.

이를 대표할 수 있는 선행연구로서 Burgin (1992)과 Lesk & Salton(1968)은 서로 다른 과정에 의해 얻어진 적합성 판정 세트는 검색효율성의 변화에 중대한 영향을 미치지 않는다고 보고하고 있다. 그러나, 이들의 검색실험은 문헌의 초록/표제를 이용하여 행해졌으며, 적합성 판정의 실질적인 대상이 되었던 전문(full-text)을 검색대상으로 이용하였을 경우, 이것이 검색효율성의 평가에 미치는 영향에 대해서는 아직까지 연구된 바가 없었다. 본 논문은 전문을 대상으로 얻어진 4개의 상이한 적합성 판정 세트가 전문 검색시스템의 효율성 평가에 미치는 영향을 연구하는데 그 초점을 두었다.

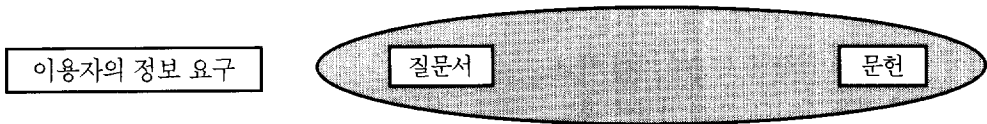
2. 이론적 배경

적합성 개념은 지난 30년간 끊임없이 논의되어 왔고 최근에는 많은 정보학자들에 의해 재조명되고 있다(Harter, 1996; Park, 1994; Harter, 1992; Jane, 1994; Schamber, 1994; Schamber et al, 1990; Swanson, 1986; Saracevic, 1975). 적합성에 대한 정의 및 논의는 너무 다양하고 선행연구에서 구체적으로 논의되고 있기 때문에 본 논문

에서는 크게 두 그룹으로 나누어 간단히 언급해 보고자 한다. 하나는 객관적인 관점에서 본 것이고 다른 하나는 주관적인 관점에서 본 적합성이다(Swanson, 1986). <그림 1>에서와 같이 주관적인 적합성(subjective relevance)은 문헌과 정보요구자의 직접적인 관계로 정의될 수 있는데 이는 곧 “이용자 지향적”인 관점에서 보는 것이고, 객관적인 적합성(objective relevance)은 이용자의 정보요구를 표현하고 있는 질문서와 문헌과의 관계를 나타내는 것으로 “시스템 지향적”인 관점에서 보는 것이다. “시스템 지향적”인 관점은 주제에 대한 적합성 여부가 질문서에 출현하는 용어가 문헌에 출현하는 주제어와 일치되는 정도에 의해 결정될 수 있다고 하여 “주제와의 관련성(topicality)”으로 표현하기도 한다. 반면에 “이용자 지향적”인 관점은 적합성은 고정적이거나 정적인 개념이 아니고 동적인 개념이며 상황에 따라 변화할 수 있다는 것이다. 즉 “적합성은 다차원적인 개념으로 내적(인지적) 또

는 외적(상황적)요인에 지배를 받는 것이고 ; 인간의 동적인 판정과정에 기반을 둔다는 것이다”(Schamber et al, 1990, p.755).

정보검색 효율성의 평가를 위한 적합성 개념은 일반적으로 질문서와 문헌간의 내용상의 일치, 즉 문헌이 질문의 주제를 다루고 있는 정도로 측정되었다. 그러나 실제의 적합성 판정은 시스템에 질문을 제시한 이용자의 주관적 판정으로 이루어지는 것으로서 특정한 질문에 대해 검색된 같은 문헌이라도 이용자에게 따라 적합문헌으로 평가될 수도 있고 부적합문헌으로 평가될 수도 있다. 또한 같은 이용자에게 동일한 문헌을 제시하여도 제시된 시점에 따라 적합성 판정이 달라질 수 있다. 이 때문에 주제의 일치정도에 따르는 적합성 개념만으로는 이용자의 정보요구 만족도를 충분히 측정할 수 없다는 비판이 제기되어 온 것이며, 탐색전이나 탐색도중의 이용자의 지식상태와 그것의 변화를 고려한 보다 주관적이고 이용자 지향적인 적합성 정의가 요구되기도 하였



(a) 객관적인 적합성



(b) 주관적인 적합성

<그림 1> 적합성 개념

다(Harter, 1996; Harter 1992). Harter는 이를 심리적 적합성 (psychological relevance)이라고 표현하고 있는데, 검색된 정보가 심리적 자극 (psychological stimulus)으로 작용하여 어떤 개인의 인지상태에 변화를 유발할 때, 그 사람에게 적합한 것으로 판정된다는 것이다.

몇몇 학자들은 적합성을 정의하는데 있어서, 서로 다른 종류의 적합성 개념을 설명하기 위해 각각 다른 용어를 사용하거나, 적합성이라는 용어를 다른 모든 개념을 포함하는 다차원적인 개념으로 정의하기도 한다(Schamber et al, 1990, p.758). 적합성의 개념을 다른 용어 - 유용성 (utility), 적절성 (pertinence), 만족도 (satisfaction) 등 - 로 대체함으로써 적합성의 일반적인 개념과 조금 다른 해석을 제공할 수는 있지만, 개념에 대한 끊임없는 논의가 근본적인 문제를 해결할 수는 없을 것이다. 그러므로, 이러한 다양한 적합성 개념이 검색 효율성의 평가에 미치는 영향을 살펴볼 수 있는 실증적인 연구가 더 필요하다.

Harter(1996)는 "검색효율성의 평가를 위한 모형이 서로 다른 적합성 개념에 의해 지대한 영향을 받지 않을 것이다"라는 가설에 정보 검색 실험은 더 이상 의존할 수 없다고 주장하였다. 그러나, 이용자의 정보요구와 관련된 모든 변수를 고려하여 정보검색 실험을 수행하기란 결코 쉬운 과정이 아니며 이는 거의 불가능한 것으로 볼 수 있다. 즉, 고정되지 않고 동적인 적합성 개념을 검색시스템의 효율성 평가에 통합시키기란 어렵다는 것이다. 이는 이용자 지향적인 적합성 개념이 필수적으로 고려되어야만 한다는 것을 알면서도 이를 대

할 수 있는 명확한 대안이 없었기 때문에 지난 30년 동안 전통적인 연구방법에 의존하였고 이용자 지향적인 적합성 개념을 검색실험에 통합한 연구 결과가 없었던 것을 보면 알 수 있는 것이다.

본 논문에서는 질문서를 직접 작성한 소아과 전문의를 포함하여 그의 동료교수 9명, 4명의 포스트 박사 과정 연구원, 온라인 탐색의 경험이 풍부한 의학 분야의 서지전문가 등 적합성을 판정하는 심사자의 유형(교육 수준, 주제에 대한 지식 정도, 경험 등)을 독립변수로 하여 얻어지는 상이한 적합성 판정 세트를 이용하였을 때 이들이 보여주는 상대적인 검색 효율성을 비교·분석하였다. 이는 적합성 여부를 판정하는 심사자를 최종 이용자로 보고, 이들의 교육 수준, 주제에 대한 지식정도, 경험 등을 이용자와 관련된 변수로 간주하고 이들이 따라 달라질 수 있는 적합성 판정을 검색 효율성의 평가에 반영한 것이다.

3. 실험 환경

본 연구에서는 SMART 시스템과 CF1239 데이터베이스의 일부인 CF392 전문데이터베이스를 이용하여 검색실험을 실시하였다. CF392 전문데이터베이스는 1974-1979 사이에 출간된 392개의 문헌으로 구성되어 있는데 이들은 미국의 국립의학도서관에 의해 MEDLINE화일에서 "Cystic Fibrosis(CF)"라는 주제용어로 색인되어 있다. CF는 소아의 호흡기질환, 즉 폐질환의 일종으로 알려져 있다. CF392의 문헌들은 CF1239의 일부분을 선정

하여 전문데이터베이스로 만들어졌는데 CF1239 중 초록을 가지고 있는 784개의 문헌들을 우선 선정하고 그들로부터 임의로 하나씩 건너서 문헌을 선정하였다. 392개의 문헌들의 전문(표제, 초록, 그리고 본문 : 표나 그림에 나오는 단어들도 포함하였으나 참고문헌은 제외하였다)은 컴퓨터를 이용하여 전문데이터베이스로 입력되었다(문성빈, 1993).

일반적으로 정보검색실험에서는 하나의 적합성 판정 세트를 이용하여 검색효율성을 평가하여 왔다. 그러나 CF 데이터베이스는 앞에서 언급한 문제들을 연구할 수 있는 여러 개의 적합성 판정 세트를 가지고 있다. 세트 A는 질문서를 직접 작성하였으며 Cystic Fibrosis의 연구에 관여하고 있는 소아과 교수에 의해 적합성 판정이 이루어진 것이고, 세트 B는 질문서를 작성한 교수의 동료 교수 9명에 의해 적합성 판정이 이루어졌다. 세트 C는 소아호흡기 분야의 포스트 박사과정에 있는 4명의 연구원에 의해 얻어진 것이며, 세트 D는 의학분야에서 온라인 탐색가로서 폭넓은 탐색 경험을 갖고 있는 서지전문가에 의해 얻어진 것이다. 즉 심사자의 유형을 교육수준, 주제에 대한 지식정도, 경험 등에 의해 구분한 것이다. 질문서-문헌들간의 적합성 판정을 위해 위의 네 그룹의 주제전문가들은 적합성 정도를 “매우 적합한”, “조금 적합한”, “전혀 적합하지 않은” 등으로 구분하였다. 그러나 본 연구에서의 적

합문헌으로 간주된 문헌은 “매우 적합한”이라고 판정되었을 경우에 해당하는 것이다. 이러한 기준에 의거하여 얻어진 각 적합성 판정 세트의 적합문헌의 수는 <표 1>에 나타나 있다.

<표 1>에 의하면, 탐색 중개인으로서 의학분야 서지전문가가 주제분야의 전문가보다 더 많은 문헌을 적합한 문헌으로 판정 내리는 것으로 나타나고 있다. 서로 다른 적합성 판정 세트간의 적합성의 일치 정도를 조사한 결과 전문가들에 의해 평가된 세트 A, B, C간에는 상대적으로 일치 정도가 높은 반면에 세트 A, B, C와 의학분야 서지전문가에 의해 평가된 세트 D간에는 일치 정도가 낮은 것으로 나타났다(Burgin, 1992).

본 연구에서 사용되고 있는 적합성 판정 세트의 적합성 판정의 과정은 다음과 같다.

“주제전문가와 의학 분야의 서지담당자는 각 질문서를 MEDLINE 탐색을 위해 사용된 것으로 간주하였고 전체 문헌 집단은 각 질문서에 의해 검색된 것으로 보고있다. 적합성 판정을 위해 심사자는 그 내용을 파악하기 위해 문헌의 전문을 훑어 보았으며, 문헌에 대한 적합성 판정은 본문에 출현하는 표나 그림에 의해 영향을 받기도 하였다. 몇몇 경우에는 문헌의 본문에는 표현되어 있지 않지만 표나 그림에 내포되어 있는 정보에 의해 적합성 판정이 내려지기도 하였다”(Shaw et al., 1991, p. 352).

이는 질문서의 용어가 적합문헌의 초록이나 표제뿐만 아니라 본문에서도 출현하지 않을 수 있다는 가능성을 암시하고 있다. 검색실험

<표 1> 각 적합성 판정 세트에서 “매우 적합한”이라고 판정된 적합문헌의 수

	세트 A	세트 B	세트 C	세트 D
CF392	408	340	355	449

의 결과는 실험 시스템의 특성보다도 질문이 작성되고 적합성 판정이 만들어지는 조건에 의해 좌우될 수 있다고 한다. 최상의 검색효율성은 적합성 판정이 단순히 문헌의 초록이나 표제에서 출현하고 있는 용어나 어구에 의해 이루어졌을 때 얻어질 수 있으며 최악의 검색효율성은 적합성 판정이 주제전문가에 의해 문헌의 본문에 함축되어 있는 미묘한 의미에 의존하여 행해졌을 때 얻어질 수 있다고 한다 (Swanson, 1988).

CF392 데이터베이스는 앞에서도 언급하였듯이 적합성 판정이 문헌에 출현하는 표나 그림에 의해 영향을 받았기 때문에 질문서의 용어가 한번도 출현하지 않는 적합문헌을 포함하고 있는 것으로 나타났다. 즉 적합문헌과 질문서의 유사도가 "0"인 적합문헌이 있음을 의미하는 것이다. 실제로 적합성 판정 세트 A, B에서는 질문서 40과 91이 각각 유사도가 "0"인 2개의 적합문헌을 갖고 있으며, 세트 C에서는 질문서 91이 유사도가 "0"인 1개의 적합문헌, 세트 D에서도 2개의 질문서 29와 94가 각각 유사도 "0"인 2개의 적합문헌을 갖고 있음이 확인되었다. 질문서와 유사도가 "0"인 적합문헌들은 검색순위가 매우 낮아 전반적으로 검색효율성을 저하시키는 요인이 되고 있다. CF392는 CF1239의 일부이므로 질문서의 수 또한 100개로부터 각 세트별로 84, 80, 84, 87개로 감소되었는데 이는 적합문헌이 존재하지 않는 질문서는 제외되었기 때문이다.

검색실험을 위해 SMART 시스템을 이용하

여 문헌과 질문서의 텍스트를 대상으로 단일 용어으로써 색인어의 어간을 추출하였다. 용어의 가중치는 질문용어의 역문헌빈도(Inversed Document Frequency)와 용어의 문헌빈도(Document Frequency)를 가장 많이 출현한 용어의 문헌빈도로 표준화(Augmented Normalized Term Frequency)시킨 것을 곱하여 아래와 같이 구하였다.

$$\text{용어 가중치} = \text{역문헌빈도(IDF)} * (0.5 + 0.5 * (\text{TFki} / \text{max_TFi}))$$

여기서 TFki는 용어 k의 문헌 i에서의 문헌빈도를 의미하며 max_TFi는 문헌 i에서 최대로 출현한 용어의 빈도 수를 나타내고 있다. 이에 대한 유용성은 이미 Salton과 Buckley (1988)에 의해 밝혀진 바 있다.

4. 실험결과 분석

4.1 표준재현율 지점에서의 정확률 및 평균 정확률

본 연구에서 초록과 표제, 또는 전문을 색인과 검색대상으로 하였을 때의 상대적인 검색효율성을 측정하여 비교하였다. 이를 위한 척도로는 표준재현율¹⁾ 지점에서의 정확률 및 평균정확률(Average Precision)²⁾을 이용하였는데 4개의 적합성 판정 세트에 의해 얻어진 결

1) 0.00, 0.05, 0.10, 0.20, 0.25 ... 1.00 등 21개의 재현율 지점

2) 0.25, 0.5, 0.75의 재현율지점에서의 정확률을 더하여 평균을 구함.

〈표 2〉 표준재현율 지점에서의 정확률 및 평균정확률

재현율	세트 A	세트 B	세트 C	세트 D
0.00	0.5909	0.5617	0.5484	0.6147
0.05	0.5840	0.5604	0.5472	0.6124
0.10	0.5735	0.5505	0.5381	0.5853
0.15	0.5506	0.5220	0.5252	0.5739
0.20	0.5319	0.5090	0.5085	0.5734
0.25	0.5199	0.5052	0.4808	0.5501
0.30	0.4865	0.4775	0.4610	0.4908
0.35	0.4579	0.4432	0.4382	0.4574
0.40	0.4488	0.4423	0.4299	0.4522
0.45	0.4364	0.4387	0.4155	0.4381
0.50	0.4347	0.4385	0.4146	0.4351
0.55	0.3426	0.3632	0.3336	0.3376
0.60	0.3423	0.3623	0.3305	0.3371
0.65	0.3384	0.3592	0.3227	0.3246
0.70	0.3105	0.3215	0.2937	0.2878
0.75	0.3087	0.3165	0.2908	0.2823
0.80	0.2962	0.3064	0.2766	0.2614
0.85	0.2827	0.3026	0.2697	0.2499
0.90	0.2813	0.3002	0.2679	0.2489
0.95	0.2755	0.2993	0.2668	0.2452
1.00	0.2745	0.2985	0.2663	0.2446
평균정확률	0.4211	0.4201	0.3954	0.4225

과는 위의 〈표 2〉와 같다

〈표 2〉에서 보여주고 있는 검색효율성의 차이가 유의한지를 검증하기 위해 여러 가지 통계적 방법을 적용할 수 있겠지만 본 연구에서는 Sparck Jones(1974)의 정의를 사용하여 연구결과를 분석하였다. Sparck Jones는 검색효율성의 차이가 5% - 10%에 있을 때는

“주목할 만한(noticeable)”, 10% 이상일 때는 “실질적인(material)” 차이라고 정의하였는데 이것을 적용하여 연구결과를 분석하면 어느 적합성 판정 세트 간의 평균정확률도 “주목할 만한” 차이가 없음을 보여주고 있다. 또한 〈표 3〉에서 보여주고 있는 Burgin(1992)의 연구 결과도 초록/표제를 이용하여 4개의 적합성

〈표 3〉 각 적합성 판정 세트의 평균정확률(Burgin, 1992, p623)

	Set A	Set B	Set C	Set D
초록/표제	0.3797	0.3451	0.3490	0.3314
초록	0.3440	0.3137	0.3117	0.2960

세트로부터 얻어진 검색효율성 간에는 “주목할 만한” 차이가 없음을 보여주고 있다. 그러나 세트 A가 다른 세트에 비해 상대적으로 높은 평균정확률을 보여주고 있다. 그러나 전문을 이용한 본 연구의 결과인 〈표 2〉에서는 세트 C보다는 조금 높지만 세트 A, B, D가 거의 동일한 검색 효율성을 보여주고 있다. 결과적으로, 이는 초록/표제 또는 전문을 이용하였을 경우 4개의 상이한 적합성 판정 세트로부터 얻어진 검색효율성은 상호간에 “주목할 만한” 차이가 없음을 보여주고 있는 것이다.

4.2 선정된 9개 질문서의 검색효율성 분석

일반적으로 검색효율성의 측정 및 평가는 〈표 2〉과 같이 전체 질문에 대한 평균으로써 표현하기도 한다. 그러나 이러한 방법은 주어진 조건하에서 개별적인 질문서가 어느 정도의 검색 성능을 발휘할 수 있는지를 평가하는 데는 적합하지 못하다. 개별적인 질문서의 검색성능에 대한 분석의 유용성은 이미 밝혀진 바 있다(Burgin, 1992 ; Salton, 1970). 그러므로 본 연구에서는 각 질문서가 주어진 환경에서 얻을 수 있는 상대적인 검색효율성을 평가하고 관찰하기 위해 적합문헌을 4개 이상 가지고 있으며 세트간의 평균정확률의 차이가

약 20% 이상 되는 질문서 6개(5, 15, 16, 20, 34, 43)를 〈표 5〉와 〈표 6〉에서 분석대상으로 선정하였다. 그리고, 예외로 질문서 39와 57은 선정기준에는 미달되지만 세트 B와 D가 다른 세트와 비교해 적합문헌 수가 너무 적기 때문에 그 타당성을 살펴보기 위해 포함시켰다. 또한 질문서 94는 세트 B가 3개의 적합문헌을 가졌지만 다른 세트와 비교하였을 때 평균정확률의 차가 크게 나타나는 이유를 살펴보고자 포함시켰다.

아래의 〈표 4〉에서는 적합성 정도와 이에 대한 정의, 그리고 이에 따라 심사자가 부여한 점수가 나타나 있다.

〈표 4〉 적합성 정도와 그 정의 및 점수

적합성 정도	정의	점수
매우 적합한	문헌이 질문서의 요구에 정확하게 일치.	2
조금 적합한	문헌이 주체적으로 적합하나 질문서의 요구에 정확하게 일치하지 못함.	1
부적합한	문헌이 질문서의 요구에 부적합함.	0

그리고, 아래의 〈표 7〉 - 〈표 15〉는 각 질문서의 특성을 살펴보기 위해 질문서의 텍스트와 질문서들에 대해 네 심사자 그룹 중에 한 그룹에 의해서는 어떻게 판정되고 있는가를

〈표 5〉 각 질문서의 평균정확률

질문서	세트 A	세트 B	세트 C	세트 D	질문서	세트 A	세트 B	세트 C	세트 D
1	0.7778	0.7778	0.1624	0.1667	51	0.1781	0.2007	0.1651	0.1736
2	-	-	-	-	52	-	-	-	-
3	0.1000	0.0905	0.0905	0.3333	53	-	-	-	0.2500
4	-	-	-	-	54	0.6579	0.6333	0.8009	0.7593
5	0.4156	0.1187	0.3972	0.3864	55	0.0294	0.0294	0.0294	0.0221
6	-	-	-	-	56	-	-	-	-
7	-	-	-	0.0286	57	0.2000	0.4762	0.4000	1.0000
8	1.0000	-	0.5000	1.0000	58	0.3191	0.2469	0.2644	0.2133
9	0.2500	-	-	0.0833	59	0.0607	0.2092	0.0895	0.1476
10	0.3725	0.7059	0.7255	0.7255	60	0.0197	0.0036	0.0156	0.0465
11	1.0000	1.0000	1.0000	0.9167	61	0.7037	1.0000	0.0556	0.4621
12	0.7778	0.7778	1.0000	0.1667	62	0.0944	0.0824	0.1176	0.0763
13	-	0.0345	0.0345	0.0056	63	0.3361	0.3949	0.2778	0.2222
14	0.0652	0.0911	0.0667	0.0667	64	0.0290	0.0290	0.0290	0.0290
15	0.3289	0.1778	0.2152	0.4444	65	0.0558	0.0477	0.0583	0.1250
16	0.0721	0.0511	0.1420	0.2134	66	0.5725	0.5620	0.5620	0.2725
17	0.3651	-	0.3651	0.6984	67	-	-	-	-
18	0.3580	-	0.0370	0.0370	68	0.3333	0.3333	-	0.8889
19	0.0388	0.0388	0.0345	0.0194	69	1.0000	1.0000	1.0000	1.0000
20	0.3393	0.6250	0.3512	0.3942	70	0.3750	0.3750	0.3750	0.3750
21	0.5694	0.5000	0.4667	0.6444	71	-	-	-	-
22	-	-	0.6703	0.6884	72	0.6759	0.0139	0.6759	0.7778
23	0.0357	0.0357	0.0357	0.0476	73	0.8889	0.8889	0.8889	0.8889
24	0.1528	0.1528	0.1528	0.1239	74	-	-	-	-
25	0.1042	0.0161	0.0161	0.1109	75	1.0000	1.0000	1.0000	1.0000
26	0.7778	1.0000	0.7222	0.6389	76	1.0000	1.0000	1.0000	0.7222
27	0.4000	0.3333	0.4000	0.4000	77	0.0905	0.1122	0.1122	0.1104
28	-	-	-	0.6190	78	0.0291	0.0632	0.0401	0.0388
29	-	-	0.0175	0.1066	79	0.2431	0.2431	0.2431	0.2500
30	0.5000	-	-	1.0000	80	-	-	-	-
31	0.3444	0.3464	0.3444	0.4444	81	1.0000	1.0000	0.5000	1.0000
32	1.0000	1.0000	0.8889	0.8889	82	0.3869	0.5714	0.6250	0.6250
33	0.1118	0.0613	0.0613	0.0506	83	0.1429	0.1667	0.2308	-
34	0.7143	0.8571	0.8095	0.3182	84	0.5000	0.5000	0.5000	-
35	0.0172	0.0086	0.0086	0.0086	85	0.0316	0.0247	0.0316	0.0454
36	0.0104	0.0104	0.0104	-	86	0.0488	0.0417	0.0417	0.6944
37	0.5529	0.4353	0.4647	0.5301	87	0.4210	0.4210	0.4210	0.4210
38	0.8889	1.0000	1.0000	0.6806	88	0.0833	0.0833	0.0833	0.0833
39	0.1270	0.0204	0.1347	0.3822	89	0.3333	0.2614	0.3333	0.2963
40	0.0925	0.0088	0.1274	0.1373	90	0.4524	0.8667	0.6984	0.9333
41	0.6190	0.6190	0.8571	0.8333	91	0.1564	0.1852	0.1281	0.1470
42	0.8095	0.8095	1.0000	0.8095	92	0.6108	0.8571	0.3160	0.5000
43	0.2810	0.3025	0.0588	0.2908	93	1.0000	1.0000	1.0000	0.5597
44	0.6443	0.6308	0.6036	0.6268	94	0.6944	0.1500	0.6944	0.4451
45	1.0000	0.7407	-	0.7407	95	0.5000	-	0.5000	0.0081
46	0.6989	0.6806	0.6989	0.6989	96	0.0909	0.0909	0.0909	0.0909
47	0.1633	0.1633	0.1176	0.0254	97	0.3333	0.3333	0.2000	0.2000
48	0.5111	0.5111	0.5111	0.5354	98	0.0963	0.0963	0.0963	0.0508
49	0.7111	0.7778	0.7250	0.7374	99	-	0.5000	0.5000	-
50	0.5000	1.0000	1.0000	1.0000	100	1.0000	1.0000	1.0000	1.0000

〈표 6〉 각 질문서의 적합문헌 수

질문서	세트 A	세트 B	세트 C	세트 D	질문서	세트 A	세트 B	세트 C	세트 D
1	2	2	2	1	51	31	34	22	28
2	0	0	0	0	52	0	0	0	0
3	1	2	2	1	53	0	0	0	1
4	0	0	0	0	54	11	6	12	11
5	8	7	5	7	55	1	1	1	2
6	0	0	0	0	56	0	0	0	0
7	0	0	0	1	57	5	7	5	1
8	1	0	1	2	58	24	17	18	17
9	1	0	0	1	59	13	8	12	23
10	2	2	3	4	60	2	1	3	2
11	2	2	2	3	61	2	1	1	3
12	2	2	1	1	62	5	3	2	5
13	0	1	1	1	63	6	6	5	2
14	5	3	4	3	64	2	2	2	2
15	10	8	9	13	65	9	5	7	15
16	9	7	10	14	66	3	3	3	5
17	2	0	2	2	67	0	0	0	0
18	2	0	1	1	68	1	1	0	2
19	4	4	3	2	69	1	1	1	1
20	6	4	6	5	70	4	4	4	4
21	3	1	2	4	71	0	0	0	0
22	0	0	2	3	72	2	1	2	2
23	4	4	4	5	73	2	2	2	2
24	2	2	2	2	74	0	0	0	0
25	4	1	1	3	75	1	1	1	1
26	2	1	2	3	76	1	1	1	2
27	2	1	2	2	77	6	8	7	8
28	0	0	0	4	78	3	5	3	4
29	0	0	1	5	79	3	3	3	1
30	1	0	0	2	80	0	0	0	0
31	4	2	4	6	81	2	2	1	1
32	1	1	2	2	82	5	3	4	4
33	6	5	5	7	83	1	2	3	0
34	5	4	4	5	84	1	1	1	0
35	2	1	1	1	85	3	3	3	6
36	1	1	1	0	86	2	1	1	2
37	12	10	11	14	87	4	4	4	4
38	2	1	1	4	88	1	1	1	1
39	9	1	9	15	89	1	2	1	2
40	8	3	8	8	90	3	4	3	5
41	3	3	4	3	91	28	26	23	31
42	3	3	2	4	92	8	4	5	4
43	7	7	5	7	93	1	1	1	3
44	34	32	26	32	94	6	3	6	10
45	1	2	0	2	95	2	0	2	1
46	4	3	4	4	96	1	1	1	1
47	5	5	2	3	97	2	2	1	1
48	3	3	3	3	98	2	2	2	3
49	10	12	11	8	99	0	1	1	0
50	1	2	2	2	100	1	1	1	1

〈표 7〉 질문서 5와 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	9	17	46	105	142	150	159	233	259	277	288	291	360	평균 정확률
Set A (8)		2	2	2	0	2	2	2	0	2	1	0	2	0	0.4156
Set B (7)		2	2	2	0	1	1	2	0	2	2	0	2	0	0.1187
Set C (5)		2	2	2	1	2	2	1	0	1	1	0	1	0	0.3972
Set D (7)		0	1	2	2	0	2	1	2	1	2	2	0	2	0.3864
검색순위		60	159	1	46	24	2	178	69	75	40	105	8	83	

보여주기 위해 각 표에 적합성 정도에 대응하는 〈표 4〉의 점수를 포함시켰다. 또한 이 문헌들의 검색순위도 보여 주고 있다.

평균정확률을 보여 주고 있다.

4.2.1 질문서 5

질문서 5는 세트 B가 다른 세트에 비해 낮은 평균정확률을 보여 주고 있다. 이에 대한 주요 원인은 다른 세트들이 검색순위 1, 2위인 문헌 46과 150을 적합문헌으로 판정하여 초기에 검색한 반면에 세트 B는 문헌 150을 "조금 적합한"이라고 판정하여 적합문헌으로 포함시키지 못하였기 때문에 상대적으로 낮은

4.2.2 질문서 15

질문서 15의 경우에는 세트 D는 다른 세트 (A, B, C)보다 많은 적합문헌을 갖고 있는 것으로 나타났다. 각 세트에 의해 얻어진 평균정확률은 세트 D가 0.4444, 세트 A가 0.3289로 세트 B, C보다 10-20% 이상의 차이를 보여 주고 있는데 이는 세트 A, D가 검색순위 1위인 문헌 164를 "매우 적합한"이라고 평가한 반면에 세트 B, C는 "조금 적합한"이라고 판정하여 세트 B, C는 이를 적합문헌으로써 포

〈표 8〉 질문서 15와 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	22	34	64	119	134	164	223	229	230	243	322	365	375	387	391	평균 정확률
Set A (10)		2	2	2	2	2	2	2	2	1	1	1	0	2	0	2	0.3289
Set B (8)		1	1	2	2	2	1	2	2	1	1	2	0	2	0	2	0.1778
Set C (9)		2	1	2	2	2	1	1	2	2	0	2	0	2	0	2	0.2152
Set D (13)		2	2	2	2	2	2	2	2	2	2	1	2	2	2	0	0.4444
검색순위		21	13	25	22	12	1	15	72	9	11	134	65	4	37	45	

함시키고 있지 못하기 때문이다. 질문서 5와 15의 결과는 검색효율성이 초기에 검색된 적합문헌들에 의해 의존하고 있음을 보여주고 있다.

4.2.3 질문서 16

질문서 16의 검색효율성은 전반적으로 낮은 것으로 나타났고 질문서에 대한 적합문헌의 수는 세트 D가 14개로 가장 많으며 또한 세트 D는 검색순위 2위인 문헌 288을 초기에 검색함으로써 다른 세트에 비해 상대적으로 높은 검색효율성(0.2134)을 보여 주고 있다. 반면에 나머지 세트들의 검색효율성은 아주 낮은 것으로 나타났다. 이는 주제전문가 그룹들에 의해 적합하다고 판정된 적합문헌의 검색순위가 대체로 낮아 검색효율성을 저하시키는 요인이 되었기 때문이다. 위의 표에서는 서지전문가에 의해 적합한 문헌으로 판정 받은 문헌의 검색순위가 주제전문가들에 의해 적합한 문헌으로 판정받은 문헌의 검색순위보다 높은 것으로 나타났다. 특히 문헌 288은 세트 A, B, C에서는 “부적합한”으로 판정되어 적합

한 문헌으로서 포함되지 못한 반면에 세트 D에서는 “매우 적합한”으로 판정되어 있다.

4.2.4 질문서 20

질문서 20에 대한 적합문헌의 수는 각 세트가 거의 비슷하며, 특히 주목할 만한 것은 각 세트가 거의 동일한 문헌(242, 365, 383, 384)들을 “매우 적합한” 문헌으로 포함하고 있어 적합문헌의 판정에 있어서도 높은 일치도를 보이고 있다. 그 중 B가 상대적으로 가장 높은 평균정확률(0.6250)을 보여 주고 있는데, 그 이유는 나머지 세트 A, C, D가 한두 개의 적합문헌을 더 가지고 있지만 이들의 검색순위(13, 28, 35)가 낮기 때문에 세트 B와 비교하였을 때 오히려 검색효율성을 저하시키는 요인이 되고 있다.

4.2.5 질문서 34

질문서 34에 대한 검색효율성은 세트 D가 다른 세트에 비해 훨씬 낮은 것으로 나타났는데 이는 검색순위 2인 문헌 298을 “조금 적합

〈표 9〉 질문서 16과 문헌들에 대한 각 세트의 적합성 판정

문헌 세트 (적합문헌수)	32	39	56	75	123	144	147	201	207	215	230	243	246	288	293	334	365	387	평균 정확률
Set A (9)	2	2	2	2	2	1	2	1	2	0	1	1	2	0	2	0	0	1	0.0721
Set B (7)	2	1	2	2	2	2	1	1	1	0	1	1	1	0	2	0	0	2	0.0511
Set C (10)	2	2	2	1	2	2	2	0	2	0	2	0	2	0	1	0	0	2	0.1420
Set D (14)	2	0	2	0	2	2	2	2	2	2	2	2	2	2	0	2	2	0	0.2134
검색순위	72	162	153	100	2	166	45	7	4	107	13	22	37	2	121	40	83	10	

〈표 10〉 질문서 20과 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	65	161	242	365	383	384	392	평균 정확률
Set A (6)		2	2	2	2	2	2	1	0.3393
Set B (4)		1	1	2	2	2	2	1	0.6250
Set C (6)		2	1	2	2	2	2	2	0.3512
Set D (5)		2	1	2	2	2	2	0	0.3942
검색순위		13	35	1	4	8	44	28	

한” 문헌으로 판정하여 적합문헌에 포함시키지 못하였고 검색순위 91위인 문헌 69를 다른 세트는 “부적합한” 문헌으로 판정하였으나 세트 D는 “매우 적합한” 문헌으로 간주하여 검색효율성을 저하시키는 요인이 되었다.

질문서 16에서는 세트 D가 다른 세트가 부적합하다고 판정한 검색순위 2위인 문헌을 적합문헌으로 판정하여 검색효율성을 높인 반면에, 질문서 34에서는 검색순위가 91위인 문헌을 적합문헌으로 포함시켜 검색효율성을 저하시키는 대조적인 결과를 초래하였다.

4.2.6 질문서 39

질문서 39에 대한 검색효율성은 세트 D가 A, B, C에 비해 높은 것으로 나타났다. 주제 전문가 그룹에 의해 얻어진 세트 A, B, C의 평균정확률은 각각 0.127, 0.0204, 0.1347로 아주 낮게 나타났다. 반면에 온라인 탐색가(서지전문가)에 의해 얻어진 세트 D는 상대적으로 0.3822의 높은 평균정확률을 보여주고 있다. 그 이유는 다른 세트와는 달리 검색순위가 3, 4, 6위에 있는 문헌(281, 35, 370)들을 적합문헌으로 판정하여 초기에 검색할 수 있었기 때문이다. 특히 주목할 만한 것은 주제전문

〈표 11〉 질문서 34와 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	24	69	83	126	199	298	평균 정확률
Set A (5)		2	0	2	2	2	2	0.7143
Set B (4)		2	0	1	2	2	2	0.8571
Set C (4)		1	0	2	2	2	2	0.8095
Set D (5)		2	2	2	2	2	1	0.3182
검색순위		6	91	41	1	7	2	

가 그룹에 의해 얻어진 적합성세트 B가 한 개의 적합문헌(69)만을 가지고 있으며, 이것은 또한 검색순위가 49번째로 초기에 검색되지 못하였다는 것이다. 세트 B는 다른 그룹에 의해서는 “매우 적합한”이라고 판정되었으며 검색순위도 높은 몇몇 문헌(35, 281)들을 “부적합한”이라고 간주하였으며, 검색순위 1위인 문헌 169를 “조금 적합한”이라고 판정함으로써 적합문헌으로 포함시키지 못하였다. 이는 결과적으로 0.0204라고 하는 매우 낮은 평균 정확률을 얻게 되는 요인이 되었다. 질문서 39에 대한 세트 B의 적합성 판정의 타당성에 의문을 제기할 수 있을 것 같다.

4.2.7 질문서 43

질문서 43의 경우에는 전반적으로 평균정확률이 낮은 것으로 나타났다. 특히 세트 C는 검색순위 1위인 문헌 132를 “조금 적합한”이라고 판정하여 적합문헌인 문헌 132의 초기검색에 실패하였으므로 낮은 평균정확률(0.0588)을 보여주고 있다. 나머지 세트 A, B, D도 적합문헌 132, 308을 제외한 나머지

적합문헌들의 검색순위가 높지 않아 비교적 낮은 평균정확률을 보여주고 있다. 그러나 세트 A, B, D간의 적합성 판정의 일치도가 그리 높지 않음에도 불구하고 검색효율성은 커다란 차이가 없는 것은 주목할 만하다.

4.2.8 질문서 57

질문서 57의 경우에 세트 D가 검색순위 1위인 문헌 1만을 적합문헌으로 간주하고 있으며 이를 처음에 검색함으로써 100%의 평균정확률을 보여주고 있다. 반면에 주제전문가의 세트 A는 다른 세트가 “매우 적합한”이라고 판정한 문헌 1을 적합문헌으로 포함시키지 못한 결과 평균정확률(0.2000)이 다른 것에 비해 상대적으로 낮은 것으로 나타났다.

앞에서 다룬 질문서 39의 경우에는 주제전문가에 의해 얻어진 세트 B가 한 개의 문헌만을 적합문헌으로 간주하였으나 그것의 검색순위가 낮아 검색효율성을 저하시키는 요인이 되었지만 질문서 57의 경우에는 대조적이다. 즉 온라인 탐색가(서지전문가)에 의해 얻어진 적합성 판정 세트 D는 한 개의 문헌만을 적합

〈표 12〉 질문서 39와 문헌들에 대한 각 세트의 적합성 판정

문헌 세트 (적합문헌수)	문헌 번호																		평균 정확률
	35	69	74	101	103	126	127	133	169	191	192	205	224	273	281	299	370	389	
Set A (9)	0	2	2	2	0	0	2	2	2	0	1	2	0	2	1	0	1	1	0.1270
Set B (1)	0	2	1	1	0	0	1	1	1	0	1	1	0	1	0	0	1	0	0.0204
Set C (9)	0	2	2	2	0	0	2	1	2	0	0	2	0	2	0	0	2	0	0.1347
Set D (15)	2	0	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	0.3822
검색순위	4	49	20	36	27	69	42	57	1	63	16	58	46	91	3	50	6	85	

〈표 13〉 질문서 43과 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	42	52	89	132	142	173	300	307	308	378	평균 정확률
Set A (7)		2	2	2	2	1	2	1	1	2	2	0.2810
Set B (7)		2	2	1	2	2	2	1	1	2	2	0.3025
Set C (5)		2	2	2	1	1	2	0	0	2	1	0.0588
Set D (7)		2	2	0	2	0	2	2	2	2	1	0.2908
검색순위		68	67	100	1	29	60	21	50	3	14	

문헌으로 가지고 있으나 그것의 검색 순위가 높아 1.0의 평균정확률을 보여 주었다.

4.2.9 질문서 94

질문서 94의 경우에는 세트 A와 C의 적합성 판정의 결과가 동일하였으며 평균정확률은 0.6944로 나타났다. 반면에 세트 B는 가장 적은 적합문헌을 가지고 있고 다른 세트와 비교하였을 때 적합성 판정의 일치도도 낮다. 평균정확률도 0.15로 낮은 검색효율성을 보여 주고 있는데, 그 이유 중의 하나는 검색순위 1, 2, 4위의 문헌 93, 92, 371을 “조금 적합한”

문헌으로 판정하여 적합문헌에 포함시키고 있지 못하기 때문이다. 또한 세트 D는 가장 많은 적합문헌을 가지고 있지만 4개의 “매우 적합한” 문헌(36, 174, 331, 360)들은 주제전문가 그룹들에 의해서 “부적합한”으로 판정되어 상반된 적합성 판정의 결과를 보여 주고 있다. 그러나 검색순위가 높은 적합문헌(92, 93, 371)들을 초기에 검색할 수 있어 평균정확률은 0.4451로서 그다지 낮지 않다.

〈표 14〉 질문서 57과 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호	1	68	181	248	279	294	318	319	320	361	평균 정확률
Set A (5)		1	2	2	1	2	1	1	1	2	2	0.2000
Set B (7)		2	2	2	2	1	2	2	1	2	1	0.4762
Set C (5)		2	1	2	1	1	1	2	2	2	1	0.4000
Set D (1)		2	1	0	0	0	0	0	0	0	0	1.0000
검색순위		1	20	10	5	30	8	9	7	14	17	

5. 결 론

정보검색 시스템의 효율성을 측정하기 위한 척도로 이용되고 있는 정확률과 재현율은 적합성 개념을 그 기반으로 하고 있기 때문에 적합성 개념에 대한 다양한 정의가 논의되어 왔고 이러한 다양한 정의가 검색효율성의 평가에 어떠한 영향을 미치는 가는 많은 학자들의 관심을 끌고 있다. 정보검색 시스템의 검색효율성은 일반적으로 각 질문서에 의해 얻어진 평가척도를 합해 질문서의 총수로 나눈 평균으로 표현된다. 질문서는 검색실험에서 간과하기 쉬운 이용자의 정보추구 행위의 관찰을 가능하게 하기 때문에, 최초 질문서를 어떻게 작성하는가와 질문서와 문헌간의 적합성 판정이 어떻게 이루어지는가 하는 것이 검색결과 내적 및 외적 타당성에 커다란 영향을 미친다(Shaw et al. 1991, p362).

CF 데이터베이스는 주제전문가에 의해 준비되고 주제분야 전문의와 연구원에 의해 검증된 질문서를 제공하고 있다. 이러한 주제전문가들은 또한 전문을 이용하여 엄밀하게 행하여 얻어진 적합성 판정 세트를 제공하고 있

다.

본 논문에서는 이러한 과정을 통해 얻어진 여러 개의 적합성 판정 세트를 검색실험에 이용하여 이들이 검색 효율성의 평가에 미치는 영향을 조사하였다. 각 적합성판단 세트의 검색효율성은 각 세트가 갖고 있는 질문서의 수에 의해 나누어진 평균 값, 즉 "21개 지점의 재현율 지점에서의 정확률과 평균정확률"에 의해 표현되고 있는데, 평균정확률과 25% 재현율지점 이후의 정확률에 있어서는 적합성 판정 세트간에 "주목할 만한" 차이가 없음을 보여주고 있다. 그러나 전체 질문서에 대한 평균을 이용하여 검색효율성을 비교하는 것은 개개 질문서의 특성을 전혀 고려할 수 없기 때문에 선정된 몇몇 질문서에 대한 분석을 면밀하게 행하였다. 결과적으로, 전체 질문서의 적합성 판정 세트간의 검색효율성에서는 "주목할 만한" 차이를 볼 수 없었지만 개별적인 질문서의 검색효율성을 비교하였을 때는 적합성 판정 세트간에 "주목할 만한 차이(noticeable difference)" 뿐만 아니라 "실질적인 차이(material difference)"도 있음을 발견할 수 있었다. 특히, 온라인 탐색가이며 서지전문가

〈표 15〉 질문서 94와 문헌들에 대한 각 세트의 적합성 판정

세트 (적합문헌수)	문헌 번호											평균 정확률
	36	92	93	160	174	237	331	360	371	380		
Set A (6)	0	2	2	2	0	2	0	0	2	2	0.6944	
Set B (3)	0	1	1	2	0	2	0	0	1	2	0.1500	
Set C (6)	0	2	2	2	0	2	0	0	2	2	0.6944	
Set D (10)	2	2	2	2	2	2	2	2	2	2	0.4451	
검색순위	34	2	1	20	63	10	19	309	4	15		

에 의해 얻어진 적합성 판정 세트(D)가 주제 전문가 그룹에 의해 얻어진 적합성 판정 세트(A, B, C)보다 더 많은 문헌을 적합한 문헌으로 판정하고 있었다. 이는 주제전문가보다 주제전문지식이 적은 서지전문가가 적합성 판정에 관대하다는 것을 암시하고 있다. 그러나, 전반적으로 세트 D의 적합문헌은 대체로 검색 순위도 높아 초기에 검색되고 있었다. 이는 25%의 재현을 지점까지 세트 D의 정확률 값이 가장 높게 나타나고 있고 다른 세트(B, C)와 “주목할 만한” 차이가 있는 것을 보면 알 수 있다.

질문서 39의 경우와 마찬가지로 질문서 94에 있어서 주제전문가 그룹에 의해 얻어진 적합성 판정 세트 B는 다른 세트와 적합성 일치 정도가 매우 낮은 것으로 나타났다. 이는 세트 B의 적합성 판정의 과정에 의문을 제기할 수 있는 것이다. 특히 질문서 39의 세트 A와 C는 9개의 적합한 문헌을 가지고 있는 반면에 세트 B는 단 한 개의 적합문헌을 가질 수 있다는 것은 이해가 되지않는 부분으로 남아 있다. 또한 세트 B는 9개의 질문서 중 5개의 질문서에서 문헌에 대한 부적절한 적합성 판정 때문에 낮은 검색효율성을 보여주고 있다.

선정된 몇몇 질문서의 상세한 분석결과

주제전문가에 의해 엄격한 적합성 판정과정, 즉 단순히 문헌에 출현하는 질문용어에 적합성 판정의 기준을 두기보다는 문헌의 본문에 함축되어 있는 의미에 그 기준을 두고 있어 오히려 검색효율성을 저하시키는 요인이 되고 있음을 확인하였다. 이는 일반적으로 통용되고 있는 상용시스템이 불 논리나 매칭함수에 의존하고 있어 동일한 결과를 가져오게 될 것이다.

서로 다른 적합성 판정 세트에 의해 얻어진 전반적인 검색효율성(즉, 전체 질문서의 평균으로부터 얻어진 평균정확률)은 전문을 이용하였을 때도 세트간에 커다란 차이가 없는 것으로 밝혀졌는데 이는 초록과 표제를 이용한 Burgin과 Lesk & Salton의 연구결과를 뒷받침하는 것이다. 즉, 이는 적합성 개념의 다양한 정의(주관적 또는 객관적)가 검색시스템의 평가에 그다지 영향을 미치지 않고 있음을 다시 한번 확인시키고 있는 것이다. 그러나 검색된 적합문헌의 적합성정보를 효과적으로 이용하여 질문서에 새로운 질문용어를 자동으로 추가할 수 있는 적합성 피이드백 방법을 통합시켰을 때 이러한 상이한 적합성 판정 세트가 검색효율성의 평가에 어떠한 영향을 미치는가는 계속 연구해야 할 과제로 남아 있다.

참고문헌

- 문성빈. (1993). 적합성 피드백을 이용한 전문검색시스템의 효율성 증진을 위한 연구. 정보관리 학회지, 10(2) : 43-67.
- Burgin, R. (1992). Variations in Relevance Judgements and the Evaluation of Retrieval Performance. *Information Processing & Management*, 28(5) : 619-627.
- Eisenberg, M.B. (1988). Measuring Relevance Judgments. *Information Processing & Management*, 24(4) : 373-389.
- Eisenberg, M.B., & Barry, C. (1988). Order Effect : A study of the Possible Influence of Presentation Order on User Judgments of Document Relevance. *Journal of the American Society for Information Science*, 39(5) : 37-49.
- Harter, Stephen P. (1996). Variations in Relevance Assessments and the Measurement of Retrieval Effectiveness. *Journal of the American Society for Information Science*, 47(1) : 37-49.
- Harter, Stephen P. (1992). Psychological Relevance and Information Science. *Journal of the American Society for Information Science*, 43(10) : 602-615.
- Janes, J.W. (1994). Other People's Judgments : A Comparison of Users' and Others' Judgments of Document Relevance, Topicality, and Utility. *Journal of the American Society for Information Science*, 45(3) : 160-171.
- Lesk, M.E. and Salton, G. (1968). Relevance Assessments and Retrieval System Evaluation. *Information Storage Retrieval*, 4 : 343-359.
- Park, T.K. (1994). Toward a Theory of User-Based Relevance - A Call for a New Paradigm of Inquiry. *Journal of the American Society for Information Science*, 45(3) : 135-141.
- Salton, G. (1992). The State of Retrieval System Evaluation. *Information Processing & Management*, 28(4) : 441-449.
- Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, 24(5) : 513-523.
- Salton, G. (1970). Automatic Text Analysis. *Science* 168 : 335-343.
- Saracevic, T. (1975). Relevance : A Review of and a Framework for the Thinking on the Notion in

- Information Science. *Journal of the American Society for Information Science*, 26(5) : 321-343.
- Schamber, L. (1994). Relevance and Information Behavior. *Annual Review of Information Science and Technology*, 29 : 3-48.
- Schamber, L., Eisenberg, M.B., Nilan, M.S. (1990). A Re-examination of Relevance : Toward a Dynamic, Situational Definition. *Information Processing & Management*, 26 (6) : 755-776.
- Shaw, W. M. et al. (1991). The Cystic Fibrosis Database : Content and Research Opportunities. *LISR* 13 : 347-366.
- Sparck Jones, K. (1974). Automatic Indexing. *Journal of Documentation*, 30 : 393-432.
- Swanson, Don R. (1988). Historical Note : Information Retrieval and the Future of an Illusion. *Journal of the American Society for Information Science*, 39(2) : 92-98.
- Swanson, Don R. (1986). Subjective versus Objective Relevance in Bibliographic Retrieval Systems. *Library Quarterly*, 56(3) : 389-398.