

적합성피이드백을 이용한 전문검색시스템의 검색효율성 증진을 위한 연구

Enhancing Performance of Full-Text Retrieval Systems Using Relevance Feedback

문 성 빈(Sung Been Moon)*

□ 목 차 □

- | | |
|--------------------|-------------------------------|
| 1. 서 론 | 4.2 검색결과와 평가 |
| 2. 배경 | 4.3 단일용어자동색인 |
| 2.1 정보검색실험 | 4.4 데이터베이스,
질문서, 그리고 적합성판단 |
| 2.2 전문검색실험 | 5. 가 설 |
| 3. 적합성피이드백과 확률검색모형 | 6. 실험결과 |
| 3.1 확률검색모형 | 6.1 주효과와 상호작용효과 검증의 결과 |
| 3.1.1 이진독립모형 | 6.2 검색효율성의 평가 |
| 3.1.2 투포이슨독립모형 | 7. 결 론 |
| 4. 실험환경 | |
| 4.1 시스템 | |

초 록

본 연구의 주목적은 전문검색시스템의 문제점인 낮은 정확률을 향상시키는 검색기법을 제시하는 것이다. 높은 재현율을 유지하면서 정확률을 증진시키기 위하여 확률검색모형(이진독립모형과 투포이슨독립모형)에 기초를 둔 적합성피이드백을 이용하였고, 이것이 전문검색시스템의 검색효율성에 미치는 영향을 조사하였다.

ABSTRACT

The primary purpose of the study is to improve the low precision often found in full-text retrieval systems. In order to enhance the low precision of full-text retrieval while retaining its high recall, relevance feedback mechanisms based on probabilistic retrieval models (binary independence and two-Poisson independence models) were employed. This paper investigates the effect of relevance feedback on the performance of full-text retrieval systems.

* 연세대학교 문헌정보학과 강사
논문접수일: 1993년 11월 6일

1. 서 론

지난 십여년동안 온라인 상용시스템에는 전문 데이터베이스가 끊임없이 증가하여 왔고, 최근에 전문데이터베이스가 차지하는 비율은 주목할 만하다. 1989년 4,000여개의 온라인 데이터베이스 중에서 전문데이터베이스가 차지하는 비율은 34%인데, 이는 1980년의 5%, 1983년의 18%, 그리고 1986년의 25%에서 점차로 증가되어 온 것이다 (Tenopir, 1989). 전문데이터베이스는 "장서에 포함되어 있는 모든 문헌들의 문장, 단어, 그리고 모든 문자들을 컴퓨터에 저장시킴으로써 기계에 의해 이들을 탐색할 수 있도록 만들어진 데이터베이스"로 정의되고 있다(Blair와 Maron, 1985, p.289). 전문검색시스템은 전문데이터베이스의 자연언어 탐색을 가능하게 하며, 표제나 초록, 그리고 통제언어집을 이용하는 검색시스템의 대체물로서 인기를 모으고 있다.

Blair와 Maron은 자동전문검색시스템의 매력을 아래와 같이 이야기하고 있다. 첫째로, 정보 및 컴퓨터기술의 끊임없는 발전은 보다 빠르고 저렴하며, 신뢰도가 높고 사용하기 편리한 컴퓨터로의 접근을 가능하게 하고 있으며, 둘째로, 고용료가 비싸고 때로는 색인작업의 일관성을 유지하지 못하는 색인자가 필요없고, 셋째로, 정보원인 전문데이터는 출판을 위해 원고를 준비하는 과정에서 손쉽게 얻어질 수 있으며, 마지막으로, 문헌의 전문을 탐색함으로써 보다 정확하게 탐색어에 접근할 수 있다(Maron, 1988; Blair와 Maron, 1985).

그 동안 여러가지 검색모형을 이용하여 검색시스템의 검색효율성을 조사한 연구들은 많으나, 이러한 모형들의 전문검색시스템에 적용가능성과 검색효율성에 대한 연구는 극히 소수에 지나지 않

았다. 왜냐하면, 무엇보다도 실험을 위한 전문데이터베이스가 부족하기 때문이다. 그리고 이전에 행해졌던 연구들은 대체로 불리안(Boolean)모형을 적용하고 있는 온라인 상용시스템에 의존하고 있다. 반면에, 본 연구는 실험실내에서 392개 문헌의 전문데이터베이스와 함께 확률검색모형에 기초를 둔 이진독립모형(binary independence model: 이하 BI)과 투포이슨독립모형(two-Poisson independence model: 이하 TPI)을 이용하여 전문검색시스템의 검색효율성을 측정하였다.

일반적으로 정보검색시스템의 검색효율성은 정확률(precision)과 재현율(recall)에 의해 측정되는데, 정확률은 검색된 문헌중 적합문헌의 비율로 정의되며 재현율은 데이터베이스내의 적합문헌총수중 검색된 적합문헌의 비율로 정의된다. 전문검색시스템은 대체로 낮은 정확률과 높은 재현율을 보여주는데, 이러한 낮은 정확률은 전문검색시스템의 문제점으로 간주되고 있다.

본 연구는 전문검색시스템의 문제점인 낮은 정확률을 증진시킬 수 있는 검색기법을 개발하는데 주안점을 두었다. 정확률을 증진시키는 방안으로는 용어가중치 또는 적합성 피이드백을 이용하는 것이 효과적이라고 알려져 있다. 특히 적합성 피이드백은 정확률의 큰 손실없이 재현율을 증진시킬 수 있는 검색기법으로 간주되고 있다(Harper와 Van Rijsbergen, 1978). Ro(1988b)의 연구는 다양한 용어가중치 알고리즘을 이용하면 전문검색시스템의 정확률이 증진됨을 증명하였다. 그러나 이제까지는 적합성 피이드백이 과연 전문검색시스템에 적용가능한가에 대한 연구는 없었다.

본 연구는 세 개의 독립변인(문헌표현방식: 초록/표제, 전문:용어분산모형:이진독립, 투포이슨독립: 적합성 피이드백:유, 무)을 조절하여 이들

이 종속변수인 검색효율성에 미치는 영향을 연구하였다. 종속변수인 검색효율성은 평균정확률(Average Precision)로 측정하였는데 이는 세 지점의 재현율(25%, 50%, 75%)에서의 정확률을 합하여 셋으로 나눈 평균값을 의미한다. 이는 검색효율성의 단일가척도의 한 방법으로 널리 사용되고 있다.

2. 배 경

2.1 정보검색실험

지난 30년동안 문헌의 다양한 구성요소를 이용한 정보검색시스템의 색인 및 검색의 효율성이 많은 정보학자들에 의해 연구되었다. 색인 및 검색의 주목적은 문헌의 구성요소 및 문헌을 대표할 수 있는 디스크립터를 이용해 문헌을 표현하여, 정보이용자들이 그들의 정보요구에 적합한 문헌을 검색하도록 하는 것이다.

Sparck Jones(1981a)는 1958-1978년을 컴퓨터를 이용한 정보검색실험기간으로 보고 이 기간을 실험구성 및 환경에 따라 크게 1958-1968, 1968-1978의 두 기간으로 나누고 있다.

첫번째 기간(1958-1968)은 “색인언어 검증기간”으로 분류되고 있으며, 이 기간 중의 실험들은 대체로 Cranfield 1400 장서에 의존하고 있다. Cleverdon과 그의 동료들에 의해 행하여진 Cranfield 프로젝트는 색인언어연구의 효시로서 간주되고 있고 그들은 다양한 색인언어의 상대적 효율성을 폭 넓게 연구검토하였다. 특히 두번째 프로젝트는 33개의 색인언어(예를들면: 단일용어, 단일개념, 통제용어, 주제, 용어가중치, 표제/초록, ..등, 또는 이들의 조합)의 상대적 효율성을 조사

하였다. 또한 여러가지 수동색인언어의 효율성을 연구하고 자동색인방법의 가능성을 제시하였다. 이러한 연구들은 자동색인과 수동색인방법을 비교연구하고 검색효율성에서 차이점을 보고하였는데 자동색인방법은 수동색인방법을 능가하거나 최소한 동등한 효율성을 가져다 준다고 하였다.

두번째 기간(1968-1978)은 자동화된 시스템하에서 적합성가중치(relevance weight)를 이용 및 수정할 수 있는 적합성 피이드백에 관한 연구의 시대라고 특징짓고 있다(Sparck Jones, 1981a). 이 기간동안에 많은 정보학자들이 적합성 정보가 유용할 때의 검색효율성에 대한 관심을 갖고 확률이론에 기초를 둔 다양한 적합성가중치 부여방법을 연구하였다. 이들은 이용자로부터 얻어진 적합성 정보를 다음 검색단계에 통합시키는 것은 검색효율성에 크게 공헌한다고 결론지었다. 처음 20년간의 연구는 표제나 초록 그리고 인공적으로 만들어진 디스크립터에만 의존하였는데 몇몇 정보학자들은 이것들만으로는 검색목적에 충족시킬 수 없음을 논의하였다. 특히 표제용어를 자연언어 그대로 문헌표현을 위해 사용하는 것은 적절하지 않음이 지적되었다. 왜냐하면, 저자들이 인상적인 표제를 만들기 위하여 때때로 이해하기 어렵고, 환상적이고, 강한 인상을 주는 단어를 사용하려는 경향이 있어서, 표제가 그들이 나타내고자 하는 참주제를 적절하게 표현하지 못하는 경우가 있기 때문이다. Salton(1970) 또한 문헌의 내용분석을 위해서 문헌의 표제만을 이용하는 것은 문헌초록이나 문헌전문을 이용하는 것보다 효과적이지 못하다고 하였다. 반면, 몇몇 정보학자들은 특히 과학문헌의 표제는 문헌표현을 위해 이용할 가치가 있다고 주장하기도 한다. 그들은 과학문헌의 표제는 KWIC(Keyword in Context)의 출현 이후 그 길

이가 점차 길어지고 있으며, 따라서 그것들의 주제반영도는 사회 및 인문과학의 표제들에 비해 훨씬 크다고 보고하고 있다. 이러한 연구들은 표제가 점차로 중요한 정보전달체로 인식되어지고 있고 KWIC의 출현 이후 정보성을 띠지 못하는 표제들은 점차 사라져 가고 있음을 보여 주는 것이다.

이제까지는 문헌의 초록이 색인 및 검색을 위한 가장 중요한 정보원으로 인식되어 왔다. 초록의 목적은 문헌의 내용을 간략하게 요약함으로써 독자가 문헌의 전문을 읽어보지 않고도 그 주제를 파악할 수 있도록 하는 데 있다. 문헌의 초록은 반드시 정보성을 가지고 있어야만 한다. 초록은 검색 실험을 위해 문헌을 표현하기 위한 도구로서 많이 사용되어 왔기 때문에 초록의 유용성에 대한 논의는 더 이상 할 필요가 없을 것이다. 많은 정보학자들은 문헌검색에 있어서 초록과 표제의 상대적 효율성에 대한 연구결과를 보고하여 왔다. 대부분의 정보학자들은 문헌표현은 초록을 이용해서 이루어져야만 한다고 했지만, 몇몇 정보학자들은 색인과 검색을 위해 문헌의 본문을 일부 추가시킴으로써 보다 나은 검색효율성을 얻을 수 있다고 하였다.

Rath et al.(1961)은 문헌검색을 위해 표제만을 이용하는 것은 잘못된 것이며 초록 혹은 문헌의 본

문 일부를 포함시켜야만 한다고 했고, Salton과 Lesk(1968)는 검색을 위해 표제에 초록을 첨가시킴으로 검색효율성이 실질적으로 향상되었지만 전문으로 확대하는 것에 대해서는 경고하고 있다. Barker et al.(1972)은 화학초록데이터베이스를 이용하여 표제와 초록의 상대적 효율성을 연구하였는데 그들은 표제로부터 초록으로의 확대 및 본문의 일부를 포함시키는 것은 보다 나은 재현율을 가져다 준다고 하였다.

반면에 Cleveland et al.(1984)은 비불리안모형(Non-Boolean Model)을 이용하여 전문을 이용하였을 때와 그렇지 아니하였을 경우의 상대적 효율성을 연구하였다. 그들은 비불리안 탐색모형인 Goffman의 간접방법(Indirect Method)을 이용하였을 때에 전문을 사용하지 않는 것이 최고의 검색효율성을 가져다 준다고 보고하였으며, 이는 전문을 이용한 색인이 꼭 필요하지 않음을 암시하고 있다.

2.2 전문검색실험

본 연구에서는 1978년 이후부터 현재까지를 이전의 두 기간(1958-1978)과는 구별하여 이를 "전문검색시스템 실험기간"으로 간주하고자 한다. 이 기간동안에는 많은 연구자들이 전문검색시스템

〈표 1〉 이전의 전문검색실험결과

연구자	정확률		재현율	
	전문	초록	전문	초록
Tenopir(1985)	0.18	0.36	0.74	0.19
Ro(1988a)	0.14	0.59	0.83	0.18
Blair & Maron(1985)	0.79	N/A	0.20	N/A
Mckinin et al(1991)	0.37	0.62	0.75	0.41

평가에 많은 관심을 보여 주었다. 최근 정보기술의 발전은 정보학자들로 하여금 전문표현 및 자연언어탐색을 가능하게 하는 전문검색시스템의 잠재력을 검토할 수 있게 하였다(Mckinin et al, 1991; Ro, 1988a, 1988b; Tenopir, 1985; Blair와 Maron, 1985). <표 1>에서는 여러 학자들이 보고한 이전의 전문검색실험결과를 요약하였는데 이는 전문검색시스템이 전반적으로 높은 재현율과 낮은 정확률을 가져다 주고 있음을 보여주고 있다. 한편, 규모가 큰 전문데이터베이스를 이용한 첫번째 연구로서 간주되는 Blair와 Maron의 실험은 연구결과가 이전의 다른 연구결과와는 상반되고 있는데 이는 주목할 만하다.

앞에서도 언급하였듯이 전문검색시스템에서는 데이터베이스에 있는 문헌의 모든 단어들 컴퓨터에 의해 탐색되어질 수 있으므로, 문헌에 할당된 색인용어들이 질문의 용어들과 매치(match)될 수 있는 기회가 많아진다. 그러므로 전문검색시스템에서는 위의 <표 1>에서 보여준 바와 같이 상대적으로 높은 재현율과 낮은 정확률을 얻게 된다. 전문검색시스템은 색인의 망라성(indexing exhaustivity)을 최대화시키기 위한 수단인 하나이다. 이는 문헌의 개념을 표현한 색인용어의 수에 의해 측정될 수 있는데 이를 증가시키는 재현율을 높일 수 있는 방법으로 간주되고 있다. 그러나 문헌표현을 위해 전문을 이용하는 것은 문헌의 주제표현에 별로 유용하지 못한 수백개의 색인용어를 추가하게 되어 시스템내에 “정보잡음(noise)”을 생성하게 하는 것과 같다고 한다(Blair and Maron, 1990). 즉 이러한 “정보잡음”은 정확률을 떨어뜨리는 요인이 될 수 있는 것이다.

경제적 혹은 실용적인 측면에서 볼 때 전문에서 전문이 아닌 문헌표현방식으로서의 압축은 저장 및

검색을 위한 비용을 절감하게 한다. 하지만, 이러한 요소들은 빠르고 저렴한 컴퓨터로의 접근을 가능하게 하는 미래의 정보사회에서는 심각하게 고려할 필요가 없게 될 것이므로, 오히려 앞으로의 시스템은 전문에 의해 동반될 수 있는 정보잡음을 제거하거나 감소시킬 수 있는 검색기법의 개발에 힘써야 할 것이다. 다음에는 지난 십여년간 행하여진 대표적인 전문검색실험의 예를 검토해 봄으로써 실험환경 및 연구결과를 관찰해 보았다.

1980년대 초에 미국화학협회(ACS)에 의해 수행된 여러 실험에서는 ACS 전문잡지 데이터베이스를 이용하였는데 통제어휘집이나 초록을 이용한 다른 방법들이 검색할 수 없는 많은 적합문헌들이 검색되었다고 한다(Durkins et al., 1980 ; Terrant et al., 1982 ; 1983).

Cleveland et al.(1984)은 비불리안(non-Boolean)탐색방법을 이용하여 전문과 초록, 참고문헌목록, 그리고 이들의 조합을 통해 얻어진 검색효율성을 조사하였는데 전문이 세번째, 그리고 표제/초록/참고문헌목록의 조합이 최고의 효율성을 가져다 준다고 하였다.

몇몇 전문검색실험은 더 많은 전문데이터베이스가 유용해짐에 따라 비교적 큰 규모의 데이터베이스로 확대되었다(Mckinin et al., 1991 ; Blair and Maron, 1985). 이러한 실험들은 정보학자들로부터 많은 관심을 끌게 되었다.

규모가 큰 전문데이터베이스를 이용한 첫번째 실험으로 알려진 Blair와 Maron(1985)의 연구는 정보학자들이나 법률도서관인들에게 그리고 LEXIS나 WESTLAW와 같은 전문검색시스템에게도 중요한 의미를 부여하고 있다. 왜냐하면 그들의 연구결과가 낮은 재현율(20%)을 보여주고 있는데, 이는 이전의 전문검색실험의 연구결과와

는 상반되기 때문이다. 이들의 연구결과로 인하여 많은 법률문헌탐색자들과 검색시스템(Computer-Assisted Legal Research: CALR)의 관리자들과 탐색도구로서의 전문검색시스템의 적합성을 재고하게 하였다. 이것은 또한 LEXIS나 WESRLAW와 같은 CALR시스템에서 전문검색시스템의 적용성에 대한 논란을 일으키게 했다.

한편 Mckinin et al.(1991)의 전문검색실험은 역시 실제환경에서 큰 상용데이터베이스를 이용하여 행하여졌는데 이는 약 89종의 잡지로부터 100,000문헌을 포함하고 있다. 연구결과는 큰 규모의 데이터베이스일지라도 전문검색시스템은 높은 재현율과 낮은 정확률을 가져옴을 확인하였으며, 이것은 Blair와 Maron의 실험결과를 뒷받침해 주지 못하고 있다. Mckinin et al.은 이러한 결과는 서로 다른 종류의 문헌이용 및 재현율의 정의에 의해 야기될 수도 있다고 한다. Mckinin et al.은 전문검색시스템을 이용하였을 때에 총 89개의 탐색 중 33개의 탐색결과가 표제나 초록을 이용한 것보다 5개 이상의 더 적합한 문헌을 검색한 것으로 보고하고 있는데 이는 Tenopir(1985)의 발견을 뒷받침해 주는 것이다. Mckinin et al.은 또한 전문검색시스템에서 몇몇 적합한 문헌이 왜 검색되고 있지 않는가를 보여 주었는데 이는 너무 제한된 탐색전략과 자연언어에 포함되어 있는 문제점들 때문이라고 한다.

Mckinin et al.과 Blair와 Maron은 전문검색시스템에서 너무 많은 문헌이 검색되는 문제점을 발견하고 이는 데이터베이스의 규모가 커짐에 따라 생긴다고 하였다. 두 경우 모두 이를 피하기 위해 매우 제한적인 탐색문을 이용하였는데, 이 문제는 불리안모형 대신 확률모형을 이용하여 제거될 수 있다고 한다(Maron, 1988). Mckinin et al.은 자연

언어를 이용한 검색에서 생기는 문제점을 지적하였는데 이는 같은 생각을 표현하는데 매우 다양한 단어가 사용되고 있으며, 더욱 광의의 의미를 갖는 용어, 축약어 또는 동의어 그리고 어떤 부류에 소속됨을 나타내는 용어들을 쓰는 것이 문제를 발생시킨다고 보고하고 있다. 왜냐하면 그들은 문헌의 전문에 암시적으로 나타나지만 명확하게 표현되고 있지 않기 때문이다.

Tenopir(1985)는 Havard Business Review라고 불리는 잡지의 전문을 이용하여 그것의 검색효율성을 초록이나 표제 또는 통제어휘집을 이용하였을 때와 비교 분석하였다. 전문탐색을 통해 다른 방법으로는 검색될수 없는 더 많은 적합한문헌을 검색할 수 있음을 발견하였다. 그녀는 전문탐색에 의해 검색된 문헌의 부분적 특성을 관찰하였는데, 첫째로 전문에 사용된 용어의 특정성(specificity)의 정도는 질문의 용어들과 매치(match)될 수 있는 기회를 높이고, 두번째로, 전문의 이용은 통제어휘집의 부족한 면을 보충하고 있으며, 세째로, 몇몇 개념들은 본문에 명확하게 표현되어 있지 않고 암시되어 있었으며, 마지막으로, 전문에는 동의어를 많이 사용하고 있어 불완전한 탐색전략을 보완해 주기도 한다. Tenopir는 본문에 나타나는 용어의 출현빈도를 이용하거나, 전문과 통제어휘집을 동시에 이용한 검색은 정확률을 증진시키는 수단일 것이라고 말한다.

Ro(1988a, 1988b)는 두가지 실험결과를 발표하였는데, 첫번째 실험에서는 전문검색이 다른 검색방법과 비교하였을 때에 높은 재현율과 낮은 정확률을 가져온다는 것을 확인하였고, 두번째 실험에서는 29개의 용어가중치 알고리즘이 정확률을 증진시키고 있음을 보여주었다. 그녀는 실험결과가 Blair와 Maron의 실험결과와 다른 이유를 아래와

같이 설명하였다: 1) 재현율에 대한 정의가 다르고, 2) 법률문헌은 Havard Business Review의 문헌과 비교하였을 때에 특성상 더욱 많은 동의어를 사용하고 있다. 결론적으로, Ro는 비용절감이 요구되지 않는다면 가장 많은 적합문헌을 검색하여 주는 전문을 이용한 포괄적 탐색이 바람직하지만, 적합성정도가 비교적 높은 문헌을 검색하고자 할 때는 통제어휘집이나 초록을 이용한 간단한 탐색(brief search)을 추천하고 있다.

3. 적합성 피드백과 확률검색모형

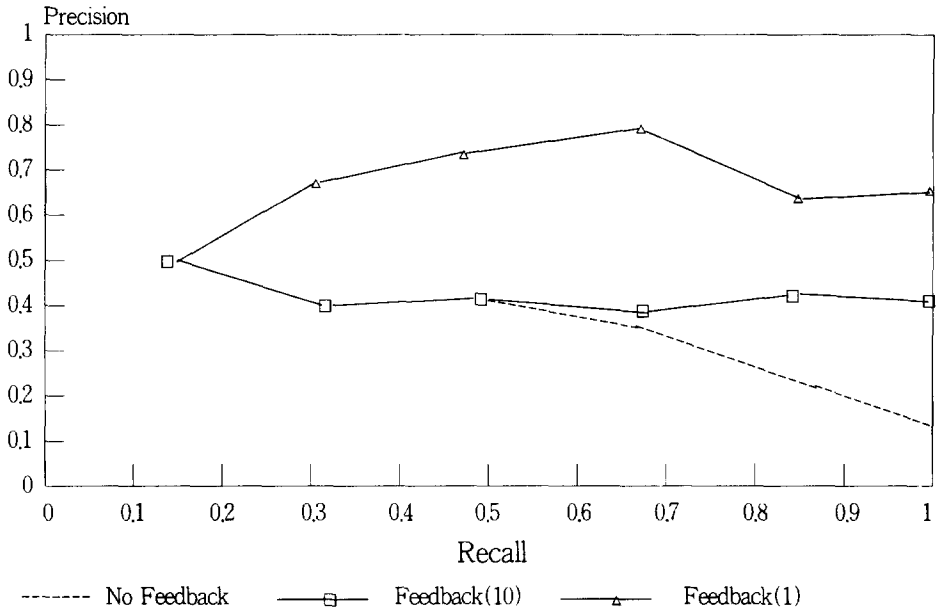
적합성 피드백은 검색효율성을 높이는 검색 기법으로 이미 알려져 있으며 검색효율성을 증진시키는 데에 기여하고 있는 것으로 보여지고 있다. 적합성 피드백은 질문서를 자동으로 재작성시키는 과정으로 사용이 용이하고 상당히 효과적인 방법이다(Salton and Buckley, 1990). 다양한 적합성 피드백 방법이 검증되고 각 방법의 장단점은 이미 많은 정보학자들에 의해 논의되었다(Salton and Buckley, 1990 ; Losee, 1988 ; Robertson, 1986 ; Salton and Rijsbergen, 1983 ; Croft and Harper, 1979 ; Harper and Rijsbergen, 1978).

정보이용자들은 광범위한 정보요구를 갖고 시스템에 접근하는데, 그 요구들은 질문서로 표현되어진다. 이러한 초기질문서는 매우 임의적이며 때때로 이용자들은 그들이 갖고 있는 문제점조차도 정확하게 표현할 줄을 모른다. Belkin(1982)은 이러한 정보이용자들의 상태를 “비정상적인 지식의 상태(anomalous state of knowledge)”라고 불렀고, Bookstein(1983)은 불확실성(uncertainty)이 정보검색과정에서 본질적인 것이라고 말했으며,

정보검색은 연속적인 배움의 과정이라고 했다. 그것은 또한 Swanson(1979)에 의해 시행착오의 과정이라고 불리웠다. 불확실성과 이용자의 정보요구의 비정상(anomaly)을 감소시키기 위해서 정보 검색시스템과의 연속적인 상호작용(interaction)을 가능하게 하는 검색전략이 필요하고 이를 통해 검색과정중에 불완전했던 초기질문서를 보완하여 재작성해 나갈 수 있는 것이다. 적합성 피드백은 이러한 상호작용을 정교한 방법으로 처리하고 있다.

적합성 피드백과정에서 시스템 이용자들은 검색된 문헌들에 대한 적합성평가를 하게 하는데 이는 “적합한” 혹은 “부적합한”으로 구분되어지고 있다. 정보검색시스템은 이렇게 얻어진 적합성 정보를 이용하여 새로운 질문서를 작성하게 되는데 이는 다음의 검색과정을 위해 사용된다. 본 연구에서 채택된 적합성 피드백의 기본원리는 아래와 같다.

적합성 피드백에서는 배움표본(learning sample)을 이용하는데 배움표본이란 이전의 검색과정에서 검색되는 문헌들을 의미한다. 이들은 이용자의 적합성판단과 함께 질문서용어들의 가중치를 수정하는데 이용된다. 만약에 용어가 검색된 적합문헌에서 발견되면 그 용어의 가중치는 증가하고 그렇지않으면 가중치는 감소하게 되는 것이다. 이전의 연구에서 실험자들은 적합성정보를 얻기 위해 보다 큰 배움표본인 10 혹은 20개의 문헌을 이용하였는데, 이는 보다 적은 피드백의 반복과 유사도 측정에 걸리는 시간을 단축시키기 위해서 였다. 그러나 Losee(1988)와 Smeaton(1984)는 한개의 문헌을 배움표본으로 하는 것이 적합성 피드백후 최고의 검색효율성을 가져온다는 실험결과를 보여 주었다. 이는 가능한 적은 규모의



〈그림 1〉 배움표본의 문헌수가 검색효율성에 미치는 효과 (질문서 94)

배움표본과 함께 보다 자주 피이드백을 가짐으로써 최대한의 적합성정보를 얻을 수 있다는 것을 암시하고 있다. 이런 이유로 본 연구에서는 한개의 문헌을 배움표본으로 선택하였는데, 이는 이용자가 검색된 문헌의 적합성을 검사하고 피이드백을 제공한다는 관점에서 볼 때에 가장 현실적이라고 보여진다. 보다 적은 배움표본의 검색효율성을 보여 주기 위해 CF392의 질문서 94를 선정하였고 이는 위의 〈그림 1〉에 나타난다.

3.1 확률검색모형

정보이용자들은 갖고 있는 정보요구를 충족시키고자 문제점을 갖고 정보검색시스템에 접근하지만 이들이 시스템이나 데이터베이스에 대하여 갖고 있는 지식은 불충분한 경우가 많다(Salton

and Buckley, 1990). 이러한 불확실성에 효과적으로 대처하기 위해 개발된 확률검색모형은 성공적인 것으로 평가되고 있다(Fuhr, 1992 ; Van Rijsbergen, 1992 ; Bookstein, 1983). 특히 확률검색모형에 기초를 둔 적합성 피이드백은 검색과정 중의 불확실성을 감소시키는 방법의 하나로써 인정받고 있다. 다른 검색모형들과 비교하였을 때에 확률검색모형에 기초를 둔 적합성 피이드백의 잇점은 질문서의 용어가중치가 정교한 방법에 의해 직접적으로 산출된다는 것이다(Salton and Buckley, 1990).

전문검색시스템이 가지고 있는 문제점을 해결하기 위해 확률원칙이 전문검색시스템의 설계에 적용되어야만 한다고 Maron은 제안하고 있다(Maron, 1988). Maron은 적합성(relevance)과 문헌의 특성사이에는 확률관계가 있다고 한다. 특히

전문문헌에는 적합성을 예측할 수 있는 많은 단어들이 포함되어 있는데 이러한 실마리를 이용하여 적합할 확률(probability of relevance)을 계산하고 이에의해 장서안에 있는 모든 문헌들의 검색순위를 매길 수 있도록 한다. 이것을 “확률등급원리(probability ranking principles)”라고 부른다 (Robertson, 1977).

3. 1. 1 이진독립모형(Binary Independence Model)

이진독립모형은 문헌내에서 질문용어의 출현여부에만 관심을 갖고 용어의 출현시는 ‘1’ 그리고 출현하지 않을 때에는 ‘0’이라고 문헌벡터상에 표시한다. 용어분산의 통계적 독립성을 가정하면서 검색을 위한 문헌의 순위는 아래 공식에 의해 결정된다(Bookstein, 1983; Losee, 1988).

$$\frac{\text{확률}(d/\text{rel})}{\text{확률}(d/\text{nonrel})} = \frac{\prod_{i=1}^n p_i^{d_i} (1 - p_i)^{(1-d_i)}}{\prod_{i=1}^m q_i^{d_i} (1 - q_i)^{(1-d_i)}}$$

여기서 p_i, q_i 는 i 번째 질문용어가 적합문헌이나 부적합문헌에서 출현할 확률을 각각 나타내고 d_i 는 ‘0’ 혹은 ‘1’로 표현되며 i 번째 질문용어가 적합문헌 및 부적합문헌에서의 출현여부를 보여주고 있다. 위의 표현은 문헌가중치를 계산하기 위해 양변에 \log 를 취하면 아래와 같이 전환되어진다.

$$\log \frac{\text{확률}(d/\text{rel})}{\text{확률}(d/\text{nonrel})} = \sum_{i=1}^n d_i * \log \frac{p_i / (1-p_i)}{q_i / (1-q_i)} + C$$

본 실험에서, 초기검색때에는 적합성정보가 없다고 가정하고 모든 용어의 p_i 값은 동일하게 1/2를 부여하였는데 이는 초기검색시 모든 용어들의 p_i

값은 일정하다고 가정하였기 때문이다. p_i 는 결국 모형의 파라미터인 S'/R' 로써 표현될 수 있는데, S' 는 질문용어가 출현하고 있는 적합문헌의 수이고, R' 는 장서내에 그 질문에 적합한 문헌의 총수를 의미한다. 각 용어를 위한 p_i 의 값은 검색과정 중에 적합성 피드백에 의해 수정되어진다. 반면에 ‘ n_i/N ’ (즉 장서안에서 질문서의 i 번째 용어가 출현하는 비율), 이것은 IDF(Inverse Document Frequency)의 역수인데 용어의 q_i 값으로 부여되며 검색과정중 불변한다. Croft와 Harper(1979)는 부적합문헌내에서 용어의 분산은 전체장서안에서의 것과 거의 일치한다고 가정하고 있다.

p_i 를 ‘1/2’로 그리고 q_i 를 ‘ n_i/N ’으로 치환한 후 문헌가중치는 초기검색때에 아래 공식에 의해 계산된다.

$$\sum_{i=1}^n d_i * \log \frac{p_i / (1-p_i)}{q_i / (1-q_i)} = \sum_{i=1}^n d_i * \log \frac{(1/2) / (1-(1/2))}{(n_i / N) / (1-(n_i / N))} = \sum_{i=1}^n d_i * \log \frac{(N - n_i)}{n_i}$$

본 실험에서 초기검색후에 p_i 의 값($S'=1$, 그리고 $R'=2$)은 검색된 적합문헌으로부터 얻어진 적합성 정보와 n_i/N 를 통합함으로써 아래 원칙에 의해 실제값에 접근하게된다.

$$p_i = \frac{1+n_i/N+s_i}{2+r_i}$$

여기서 s_i 는 i 번째 질문용어가 출현하고 있는 검색된 적합문헌의 수를 나타내며 r_i 는 검색된 적합문헌의 수를 나타낸다. 여기서 n_i/N 는 초기 검색 후에 첨가되어지는데 이는 교정인수(Correction factor)로서 보다 근접한 p_i 의 값을 측정하는데 도움을 준다. 예를 들면, 만약에 검색된 문헌이 적

합하고 i번째 용어를 갖고 있으면 $S'=2, R'=3$ 이 된다. 그러므로 p_i 는 $(2+n_i/N)/3$ 가 되고 그 용어가 적합문헌에서 출현할 확률이 증가됨을 보여준다. 만약에 검색된 문헌에 그 용어가 출현하지 않았을 때에는 p_i 는 $(1+n_i/N)/3$ 이 되어 그 값이 감소되고 있음을 보여준다. 이 실험에서는 오직 적합문헌만이 적합성 피드백을 위해 사용되고 있기 때문에 q_i , 즉 n_i/N 는 전 검색과정을 통해 아래와 같이 변한다.

$$q_i = \frac{n_i - S_i}{N - r}$$

그러므로 문헌가중치를 위한 공식은 적합성 피드백 동안에 아래와 같이 표현된다.

$$\sum_{i=1}^n d_i * \log \frac{\frac{(1+n_i/N+s_i)/(2+R)}{1-(1+n_i/N+s_i)/(2+r)}}{\frac{n_i - S_i}{1-(n_i - S_i)/(N-r)}}$$

위의 공식은 다음의 검색단계를 위해 장서안에 검색되지 않고 남아있는 문헌들의 검색순위를 결정하는데 사용된다.

3. 1. 2 투포이슨독립모형(Two-Poisson Independence Model)

투포이슨독립모형은 Bookstein과 Swanson (1974)에 의해 제안되어 Harter(1975a; 1975b)에 의해 연구되었다. 이 모형은 장서내에서 비주제용어(non-specialty words)는 문헌의 적합성에 상관 없이 임의로 분산되어 있으며, 그 분포는 단일 포이슨함수에 의해 표현되어질 수 있다고 가정하고 있다.

$$\text{확률}(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

하지만, 주제용어(specialty words)는 적합문헌들과 부적합문헌들 사이에 서로 다른 분포를 가지고 있다는 가정에 기초를 두고 있으며 그 분포는 아래와 같이 투포이슨 함수에 의해 표현된다.

$$\text{확률}(k) = \pi \frac{e^{-\lambda_1} \lambda_1^k}{k!} + (1-\pi) \frac{e^{-\lambda_2} \lambda_2^k}{k!}$$

여기서 π 는 장서내 적합문헌의 비율을 나타내며, λ_1, λ_2 는 각각 적합문헌과 부적합문헌내에서의 용어의 평균출현빈도를 나타낸다. 적합문헌내에서 용어의 평균출현빈도는 부적합문헌내에서 용어의 평균출현빈도보다 크다고 가정하고 있다 (즉, $\lambda_1 > \lambda_2$). 이진독립모형에서는 적합성가중치를 결정하기 위하여 용어의 출현여부만을 고려하는 반면에, 투포이슨독립모형에서는 용어출현빈도를 고려하여 적합성가중치를 계산한다. 아래와 같이 투포이슨용어분산을 가정하며 문헌의 검색순위를 위한 공식은 아래와 같이 결정된다 (Bookstein, 1983; Raghavan et al., 1983; Losee, 1987; Losee, 1988).

$$\frac{\text{확률}(d/\text{rel})}{\text{확률}(d/\text{nonrel})} = \frac{\prod_{i=1}^n \frac{e^{-\lambda_1} \lambda_1^k}{k!}}{\prod_{i=1}^n \frac{e^{-\lambda_2} \lambda_2^k}{k!}}$$

양변에 log를 취하고 나면, 위의 표현은 아래와 같이 변환된다.

$$\sum_{i=1}^n d_i * \log(\lambda_1 / \lambda_2) + \sum_{i=1}^n (\lambda_2 - \lambda_1)$$

결국, 우변은 상수가 되고 좌측의 λ_1, λ_2 는 각각 $1/1$ ($S=1, R=1$)과 n_i/N 로 치환되어 초기 검색을 위한 공식은 아래와 같이 된다.

$$\sum_{i=1}^n d_i * \log \frac{N}{n_i}$$

그리고 적합성 피이드백 후의 문헌가중치를 위한 표현은 아래와 같이 변환될 수 있다.

$$\text{문헌가중치} = \sum_{i=1}^n d_i * \log \frac{\frac{I + (f_i/N) + s_i}{I + r}}{\frac{f_i - s_i}{N - r}}$$

이전독립모형에서의 'n/N'와 같이 'f_i/N'는 교정인수(Correction factor)의 역할을 하는데, 여기서 'f_i'는 i번째 질문용어의 정서내에서 출현하는 총빈도수를 나타낸다.

4. 실험환경

이전에 행해졌던 정보검색시스템의 효율성에 대한 연구는 각기 다른 실험환경때문인지 일관성 없는 실험결과를 보여주고 있다. 정보검색실험결과에 영향을 미치는 몇몇 요소들이 있는데 이는 검색시스템, 데이터베이스, 질문서, 적합성판단이 이들에 해당한다(Robertson, 1981).

검색시스템에 의해 산출될 수 있는 검색결과들은 여러가지 방법에 의해 평가되어 질 수 있는데 이는 대체로 정확률과 재현율에 의해 이루어지고 있다. 여기서는 본 연구의 실험환경을 소개하여 평

가방법과 실험결과와 통계적 검증방법을 설명하고 있다.

4.1 시스템

본 연구의 모든 실험들은 스마트(SMART)시스템에 의해 수행되었다. 스마트시스템은 미국 코넬대학에서 최초로 1960년대에 개발되었고 지난 30년동안 정보검색실험도구로서 많이 사용되어 왔다. 사용된 시스템은 1980년 UNIX시스템을 위해 개발된 것이다. 이 시스템에 앞에서 언급한 두 가지(BI와 TPI)의 적합성 피이드백 과정이 본 연구자에 의해 첨가되었다.

문헌과 질문서들은 스마트시스템에서 쓰이고 있는 단일용어 자동색인방법에 의해 어미를 제거한 형태로 표현되었다(예를 들면 library와 librarian은 같은 단어로 취급되었다). 스마트시스템은 21군데의 표준재현율(standard recall : 0.00, 0.05, 0.1, 0.15, ... 0.95, 1.0)에서의 정확률을 산출해 낸다.

4.2 검색결과와 평가

본 연구에서 검색과정은 각 질문서에 해당되는

〈표 2〉 적합성 피이드백 유무에 따른 문헌의 순위 변화

문헌	적합성피이드백	
	무	유
A	1	1
B	5	2
C	9	5
D	12	8
E	20	10

〈표 3〉 적합성 피이드백 유무에 따른 정확률의 변화

재현율	정확률	
	피이드백 전	피이드백 후
0.2	1.00	1.0
0.4	0.40	1.0
0.6	0.33	0.6
0.8	0.33	0.5
1.0	0.33	0.5

적합문헌들이 전부 검색될 때까지 계속되는데 검색결과는 적합문헌의 유사도에 따른 순위에 근거하여 정확률과 재현율을 계산함으로써 평가되어지고 있다. 본 연구의 주 관심사는 적합문헌들의 순위가 서로 다른 조건하에서 어떻게 변화하는가에 두고 있다. 5개의 적합문헌이 어떤 질문서에 주어졌다고 가정할때 5개 적합문헌(A,B,C,D,E)의 순위는 유사도에 따라 장서내에서 〈표 2〉와 같이 나타날 수 있다. 주어진 표에 의하면, 정확률과 재현율은 〈표 3〉과같이 계산될 수 있다.

각 질문서는 서로 다른 수의 적합문헌들을 갖고 있기 때문에 대부분의 질문서는 표준재현율의 각 지점(예 : 0.00, 0.05, ..., 0.95, 1.00)에 해당되는 정확률값을 갖고 있지 못하다. 일반적으로 재현율과 정확률에 의해 표현되는 그림은 초기에는 톱니바퀴와 같은 형태를 하고 있지만 스마트시스템은 보간법(interpolation)을 이용하여 점을 따라 선을 비교적 매끄럽게 만들어주고 있다. 스마트시스템은 100%의 재현율에서의 정확률값을 나타내는 지점으로부터 시작하여 왼쪽으로 평행선을 그어 보다 높은 정확률값을 가질 수 있는 재현율 지점에서 멈추는 보간법을 이용하고 있다(Ide, 1971, p.344). 보간법에 의해 만들어진 곡선을 가지고 각 표준재현율 지점에서의 각 질문서에 해당되는 보

간된 정확률(interpolated precision)이 산출되었을 때에, 그 정확률값들은 합하여져 총 질문서수로 나뉘어지고 평균값을 갖게 된다. 21개의 표준재현율 지점에서의 정확률값이외에도 몇몇 단일가척도방법이 사용되고 있는데 그 중의 하나는 평균정확률(average precision)과 10, 30, 50, 70, 100개의 문헌들이 검색된 후의 재현율과 정확률을 보는 것이다.

본 연구에서는 이러한 값들을 근거로 하여 전반적인 검색효율성을 평가하였다. 특히 10, 30, 50, 70, 100 문헌이 검색된 후의 정확률과 재현율의 값은 비교적 높은 재현율에서의 정확률값의 변화를 관찰하는데 도움이 되고 있다.

4.3 단일용어 자동색인

이제까지 다양한 문헌표현의 방식이 제안되어왔고 그들에 대한 검색효율성이 연구되어왔다. 이러한 것들로는 1) 자연언어 그대로의 단일용어, 2) 어미제거와 같은 최소통제에 의해 생성된 단일용어, 그리고 3) 색인자나 컴퓨터에 의해 만들어진 어구 및 통제어휘집 사용 등이다.

이전 연구들의 실험결과는 최소한 통제와 함께 생성된 단일색인용어가 다른 색인시스템들보다 검색효율성에 있어 우위에 있는 것으로 보고하고

있다(Cleverdon, 1967 ; Keen, 1973). 두번째 크랜필드 프로젝트(Cranfield Project)도 자연언어 그대로의 단일색인용어가 동의어를 포함하거나 어미를 제거한 후에 최고의 색인언어로 됨을 확인하였다(Cleverdon, 1967). 반면에 Aberystwyth 프로젝트는 자연언어 그대로의 단일용어가 아무런 통제없이도 다른 것들과 비교하였을 때에 더 나은 검색결과를 보여준다고 하였다(Keen, 1973).

Salton과 Zhang(1986)은 용어간의 관계에 대한 그럴듯한 이론이 개발될 때까지 구현하기 쉽고, 바람직한 검색효율성을 가져다주는 단일용어 자동 색인방식을 고수하는 것이 최선의 방법이라고 말한다.

Bhattacharyya(1974)는 기초과학(예, 화학, 물리, 식물, 동물, 지질학)등에 속하는 특정한 주제 분야들의 용어적 혹은 정보적 구조를 연구하였다. 그는 검색과 색인에 있어서 과학분야에서의 자연언어의 효율성을 연구하였는데, 순수과학(soft science)과 응용과학(hard science)사이의 구별은 그 주제의 용어적 일관성을 관찰함으로써 행하여질 수 있다고 하였다. 그는 용어적 일관성을 "어떤 개념과 그에 대응하는 용어간의 관계에 있어서의 안정성(stability)"으로 정의하고 있다. 어떤 개념의 용어적 일관성은 아래와 같이 표현되어 질 수 있다.

$$t = \frac{1}{s}$$

여기서는 "s" 는 그 개념에 해당되는 동의어의 수를 의미하고 있다. 어떤 분야에 n개의 개념이 있다고 가정하면 그 분야의 용어적 일관성은 전체적으로 아래와 같이 측정될 수 있다.

$$T = \frac{1}{n} \sum_{i=1}^n \frac{1}{s_i}$$

위의 등식은 T의 값이 클수록 그 분야의 용어적 일관성이 높다는 것이다. 그의 결론은 통제되지 않은 형태로서의 단일용어의 이용은 응용과학에서 성공적이라는 것이다. 왜냐하면 그 주제의 용어적 일관성이 높기 때문이다(Bhattacharyya, 1974). Bonzi(1984) 역시 이론적 혹은 구체적인분야에서의 용어적 일관성에 대해 연구하였는데 구체적인분야의 용어들이 이론적(추상적)분야의 것들보다 일관성이 있다는 것을 뒷받침해주고 있다. 본 연구에서 이용된 CF392데이터베이스는 응용과학에 속하는 의학문헌들을 다루고 있다.

4.4 데이터베이스, 질문서, 그리고 적합성 판단

본 연구의 실험을 위한 데이터베이스로는 CF1239의 일부인 CF392전문데이터베이스가 이용되었다. CF392데이터베이스는 1974-1979 사이에 출간된 392개의 문헌으로 구성되어 있는데 이들은 미국의 국립의학도서관에 의해 MEDLINE 화일에서 "Cystic Fibrosis"라는 주제용어로 색인되어 있다. CF392의 문헌들은 CF1239의 일부분을 선정하여 전문데이터베이스로 만들어졌는데 CF1239중 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.

최초 CF1239의 질문서들에 대한 질문서-문헌간의 적합성판정은 노스캐롤라이나 주립대학의 의과대학교수인 Dr. Robert Wood와 몇몇 동료교수, 그리고 포스트박사과정에 있는 연구원들이 문헌의 전문을 보고 행하였다. 질문서-문헌들간의

〈표 4〉 CF392 데이터베이스의 용어 통계

	초록/표제	전문
평균용어수	47.2	388.3
표준편차	18.3	161.0
최소용어수	12	84
최대용어수	122	1133

적합성판단을 위해 위의 세 그룹의 주제전문가들은 “매우 적합한”, “조금 적합한”, “전혀 적합하지 않은”으로 구분하였으며 따라서 적합문헌은 세 그룹중 최소한 두 그룹이 “매우 적합한”이라고 판정하였을 경우에 해당되는 것이다. CF392가 CF1239의 일부이므로 질문서의 수 또한 100개로부터 83개로 감소되었는데 이는 적합문헌이 존재하지 않는 질문서를 제외한 것이다. 질문서당 최소 적합문헌수는 한개인데 이는 23개의 질문서에 해당되고 최대 적합문헌의 수는 33개로 질문서 44에 해당된다. 각 질문서당 평균 적합문헌수는 4.66이고 평균용어수는 6.84이다. CF392데이터베이스의 용어 통계는 스마트시스템에 의해 생성된 고유어간에 근거를 두었는데 〈표 4〉에 나타난다.

5. 가 설

본 연구에서는 아래와 같은 귀무가설(Null Hypotheses)들이 진술되었고 이에 대한 통계적 검증이 실시되었다.

가설 1:

적합성 피이드백을 이용한 전문검색시스템과 이용하지 않은 시스템과는 검색효율성에 있어 차이가 통계학적으로 유의하지 않다.

가설 2:

두번째 가설은 네개의 부가설로 나누어질 수 있다. 3개의 부가설(2.1, 2.2, 그리고 2.3)은 세개의 독립변인에 대해 행하여 졌는데 주효과에 대한 검증이라고 불리운다. 그리고 나머지 하나(2.4)는 상호작용효과에 대한 검증이라고 한다.

부가설 2.1:

초록/표제와 전문의 검색효율성에 있어서의 차이는 통계학적으로 유의하지 않다.

부가설 2.2:

이진독립모형과 투포이슨독립모형의 검색효율성에 있어서의 차이는 통계학적으로 유의하지 않다.

부가설 2.3:

적합성 피이드백 유무의 검색효율성에 있어서의 차이는 통계학적으로 유의하지 않다.

부가설 2.4:

2.4.1) 두개의 독립변인: 문헌표현방식과 용어 분산모형 사이에는 상호작용효과가 없다.

2.4.2) 문헌표현방식과 피이드백유무사이에는 상호작용효과가 없다.

〈표 5〉 스마트시스템의 t-검증결과의 예

Significance Test - Comparison between

Method(A):Learning Sample Size(1)

Method(B):Learning Sample Size(10)

NULL hypothesis:There is no significant difference between two methods, that is, A=B

Conclude that Method (A) is better than Method (B) with SIGNIFICANCE if Prob (NULL hyp) < .05 (that is, the value of the Deviate > 1.645:the critical value of t is 1.645 at $\alpha=0.05$).

Conclude that Method (A) is better than Method (B) with HIGH SIGNIFICANCE if Prob (NULL hyp) < .01 (that is, the value of the Deviate > 2.326:the critical value of t is 2.326 at $\alpha=0.01$).

Measure	Deviate	Significant?
Recall-Prec 3pt Ave	1.6503	YES
Recall at 10 Docs	1.5889	NO
Recall at 30 Docs	0.4223	NO
Recall at 10 Docs	1.3924	NO
Recall at 30 Docs	0.0000	NO

2.4.3) 용어분산모형과 피드백유무사이에는 상호작용효과가 없다.

2.4.4) 3개의 독립변인들 사이에는 아무런 상호 작용효과가 없다.

치가 1.645($\alpha=.05$)보다 크고 2.326($\alpha=.01$)보다 작으면 "Yes", 2.326보다 크면 "High"라 하고 그 밖에는 "No"라고 하였다.

6. 실험결과

위의 귀무가설 1의 검증을 위해 스마트시스템에서 제공하는 t-검증법을 사용하였다. 그리고 나머지 귀무가설은 Three Repeated Measure of ANOVA를 이용하여 검증하였다. 스마트시스템의 t검증의 결과의 예가 〈표 5〉에 나타나 있다.

위의 예는 방법(A)의 평균정확률이 방법(B)의 평균정확률과 매우 틀리다는 것을 보여 주고 있다. 왜냐하면 검증통계치(Test Statistic)인 1.6503이 1.645($\alpha=.05$)보다 크기 때문이다. 위의 보기에서 보여진 대로 통계적 검증을 위한 결정은 아래와 같은 규칙에 의해 결정되어 지는데 만약에 검증통계

6.1 주효과와 상호작용효과 검증의 결과

세 독립변인들(문헌표현방식: 초록/표제 그리고 전문, 피드백 유무, 용어분산모형: BI와 TPI)의 주효과와 상호작용효과에 대한 검증결과가 〈표 6〉에 나타나 있다.

위에서 A는 문헌표현방식, B는 용어분산모형, 그리고 C는 피드백 유무를 의미한다. 두 독립변수 A(문헌표현방식)와 C(피드백 유무)에 대한 주효과 검증을 위한 귀무가설(Null Hypotheses)

〈표 6〉 Three Repeated Measure of ANOVA의 결과

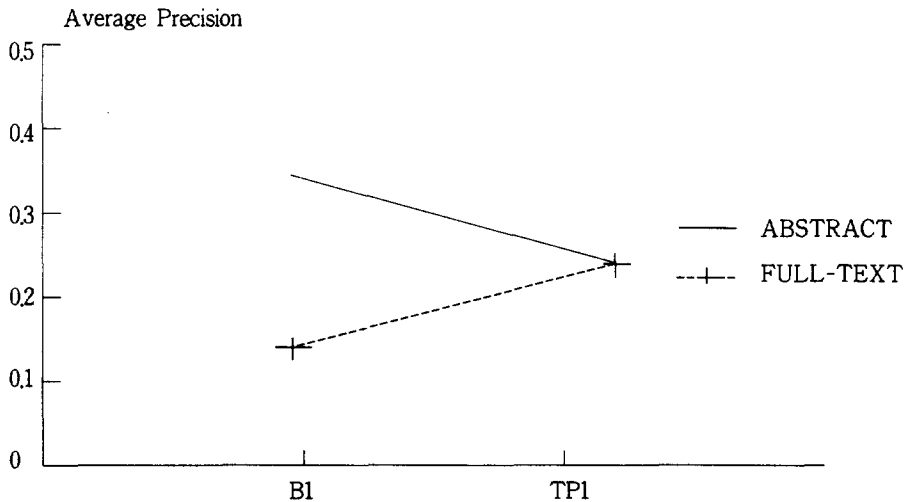
Source	DF	Anova SS	Mean Square	F Value	Pr>F
A	1	1.17571015	1.17571015	10.99	0.0014
B	1	0.00960869	0.00960869	0.15	0.6973
C	1	0.36429623	0.36429623	56.76	0.0001
A * B	1	1.30581480	1.30581480	17.55	0.0001
B * C	1	0.06682668	0.06682668	13.63	0.0004
A * C	1	0.00741627	0.00741627	1.67	0.1996
A * B * C	1	0.00074493	0.00074493	0.21	0.6481

은 받아 들여지지 않아, 주효과가 $\alpha=0.05$ 에서 유의한 것으로 나타났고, 두 번째 독립변인 B(용어분산모형)에 대한 주효과는 유의하지 않은 것으로 나타난다.

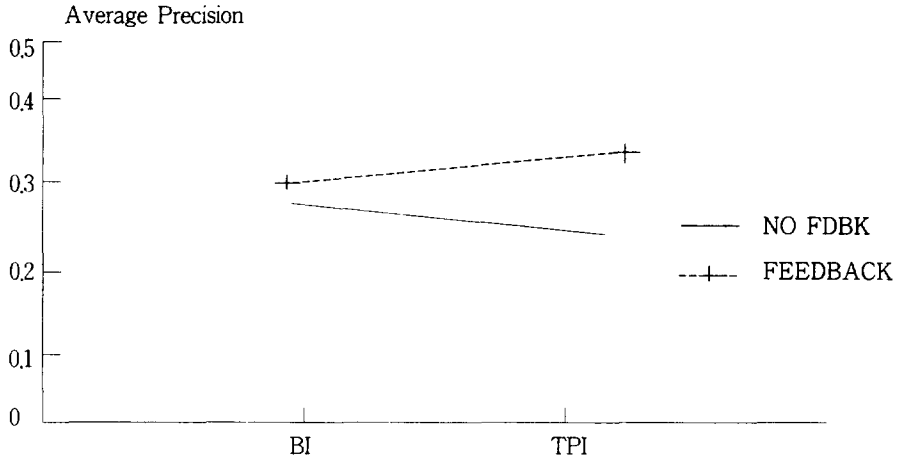
두 독립변인 A(문헌표현방식)와 B(용어분산모형)사이의 상호작용효과에 대한 검증은(가설 2.4.1) 통계학적으로 유의한 것으로 나타난다. 그것은 첫번째 독립변인(초록-표제 혹은 전문)의 영

향이 두번째 독립변인(BI와 TPI)에 걸쳐 동일하지 않음을 나타내고 있다. 두 독립변인들(A와 B)사이의 상호작용유형은 〈그림 2〉에 나타난다.

문헌표현방식이 용어분산모형에 같은 영향을 미친다면 두선은 평행선에 가까워야만 한다. 하지만 선들은 거의 교차하고 있다. 그것은 초록/표제의 검색효율성은 BI로부터 TPI로 옮길 때에 떨어지고 있는 반면 전문의 검색효율성은 향상되고 있



〈그림 2〉 문헌표현방식과 용어분산모형의 상호작용유형

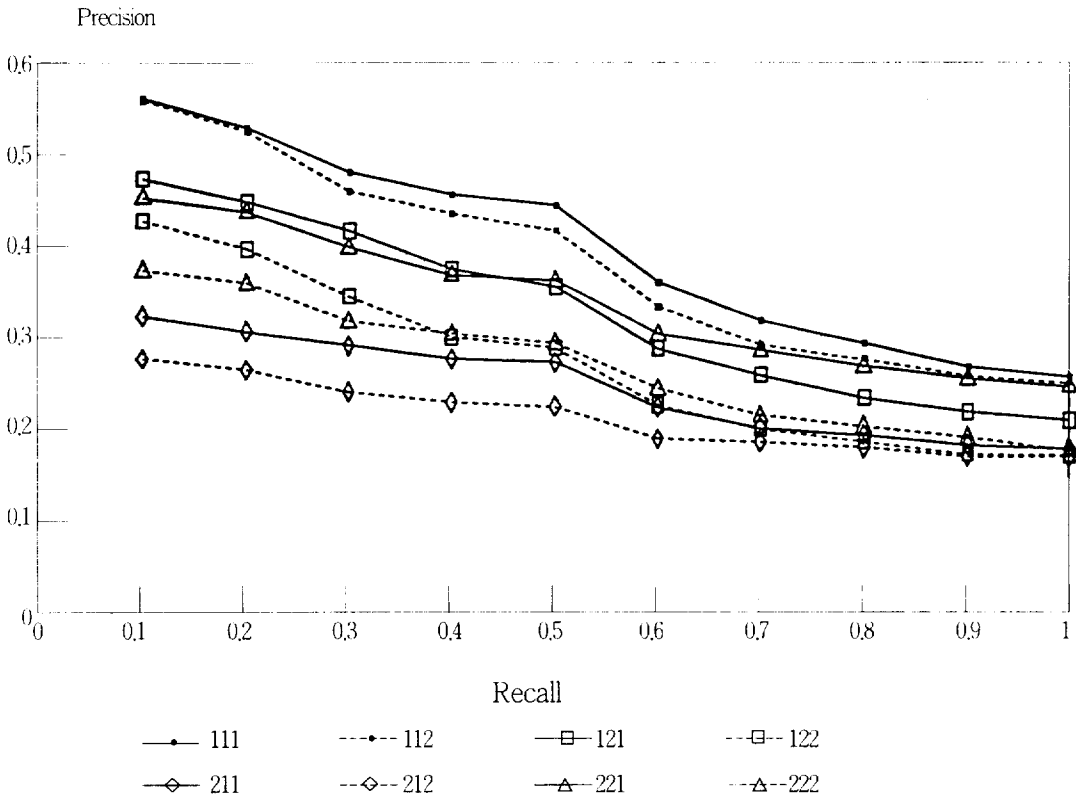


〈그림 3〉 용어분산모형과 적합성 피이드백의 상호작용유형

〈표 7〉 각 조건하에서의 검색효율성(83개 질문서의 평균치)

Recall	111	112	121	122	211	212	221	222
0.10	0.5603	0.5563	0.4726	0.4261	0.3215	0.2753	0.4517	0.3721
0.20	0.5273	0.5236	0.4464	0.3948	0.3050	0.2628	0.4362	0.3573
0.30	0.4790	0.4575	0.4143	0.3426	0.2891	0.2381	0.3963	0.3155
0.40	0.4533	0.4331	0.3714	0.2978	0.2737	0.2264	0.3652	0.3014
0.50	0.4408	0.4139	0.3522	0.2856	0.2696	0.2218	0.3590	0.2908
0.60	0.3562	0.3300	0.2842	0.2223	0.2204	0.1862	0.3000	0.2405
0.70	0.3147	0.2876	0.2546	0.1962	0.1973	0.1823	0.2823	0.2112
0.80	0.2888	0.2718	0.2300	0.1823	0.1892	0.1761	0.2646	0.1988
0.90	0.2630	0.2528	0.2139	0.1683	0.1784	0.1657	0.2507	0.1864
1.00	0.2521	0.2453	0.2040	0.1662	0.1738	0.1650	0.2412	0.1722
A.P	0.4175	0.3996	0.3434	0.2811	0.2535	0.2179	0.3525	0.2810
P_10	0.1771	0.1687	0.1675	0.1361	0.1120	0.0904	0.1386	0.1217
P_30	0.0928	0.0847	0.0867	0.0751	0.0815	0.0651	0.0884	0.0779
P_50	0.0651	0.0578	0.0612	0.0530	0.0622	0.0511	0.0627	0.0549
P_70	0.0492	0.0461	0.0480	0.0417	0.0501	0.0417	0.0494	0.0427
P_100	0.0364	0.0352	0.0375	0.0336	0.0389	0.0316	0.0375	0.0333
R_10	0.5021	0.5116	0.4784	0.3814	0.3838	0.3382	0.4705	0.3623
R_30	0.6857	0.6830	0.6639	0.6004	0.6463	0.5870	0.6882	0.6213
R_50	0.7496	0.7356	0.7229	0.6804	0.7435	0.6873	0.7677	0.6738
R_70	0.7971	0.7762	0.7702	0.7158	0.8293	0.7482	0.8227	0.7339
R_100	0.8365	0.8152	0.8361	0.7923	0.8783	0.7677	0.8863	0.7886

111: 초록-표제/BI/피이드백 유 112: 초록-표제/BI/피이드백 무 121: 초록-표제/TPI/피이드백 유
 122: 초록-표제/TPI/피이드백 무 211: 전문/BI/피이드백 유 212: 전문/TPI/피이드백 무
 221: 전문/TPI/피이드백 유 222: 전문/TPI/피이드백 무
 A.P: 평균정확률 (Average Precision)
 P_10, P_30, P_50, P_70, P_100: 10, 30, 50, 70, 100개 문헌검색후의 정확률
 R_10, R_30, R_50, R_70, R_100: 10, 30, 50, 70, 100개 문헌검색후의 재현율



〈그림 4〉 표준재현율에서의 정확률(83개 질문서에 대한 평균치)

다. B(용어분산모형)와 C(적합성 피이드백) 사이에 상호작용효과도 통계학적으로 유의한 것으로 보여진다. B와 C의 상호작용유형은 그림 3에서 보여 주고 있으며 그것은 적합성 피이드백이 BI모형 보다 TPI모형으로부터 더 많은 혜택을 받고 있다.

6.2 검색효율성의 평가

하나 혹은 둘 이상의 적합문헌을 가지고 있는 83개의 질문서가 각 조건하에서 보여주는 전반적인 검색효율성은 <표 7>(10개 지점의 표준재현율에서의 정확률, 평균정확률, 그리고 10, 30, 50, 70, 100개 문헌검색 후의 정확률과 재현율)과 <그림 4>

에 나타나 있다.

초록-표제의 검색효율성이 다른 것에 비해 높은 것으로 나타나고 있으나 <그림 4>에서 나타난 대로 적합성 피이드백후에 전문/TPI는 비교적 높은 재현율에서의 정확률이 초록-표제의 정확률에 접근하고 있다. 다음에는 적합성 피이드백이 검색효율성에 미치는 영향과 t-검증의 결과를 알아 보았다.

6.3 적합성 피이드백의 효과

적합성 피이드백의 주효과에 대한 검증결과는 유의한 것으로 <표 5>에 나타났으며, 용어분산모

<표 8> 70, 100개 문헌검색이후의 적합성 피드백의 효과

	111	121	211	221
Precision after 70	0.0492	0.0480	0.0501	0.0494
Recall after 70	0.7971	0.7702	0.8293	0.8227
Precision after 100	0.0364	0.0375	0.0389	0.0375
Recall after 100	0.8365	0.8361	0.8783	0.8863

111: 초록-표제/BI/피드백 유 121: 초록-표제/TPI/피드백 유
 211: 전문/BI/피드백 유 221: 전문/TPI/피드백 유

형과 상호작용효과가 있는 것으로 판명되었다. 전문의 평균 정확률은 TPI모형과 함께 25.44%(0.2810에서 0.3525) 향상되었고 BI모형과 함께 16.34%(0.2179에서 0.2535) 향상되었다. 반면에 초록-표제의 평균정확률은 TPI모형과 함께 22.16%(0.2811에서 0.3434) 향상되었고, BI모형과 함께 4.48%(0.3996에서 0.4175) 향상되었다. 상호작용유형은 <그림 3>에서 가리키듯이 TPI모형이 적합성 피드백으로부터 혜택을 더 받고 있다고 보여 주고 있다. 그러나 초록-표제/BI/피드백의

평균정확률(0.4175)이 여전히 가장 높은 것으로 나타나고 있다. 이는 검색에 있어 초록/표제의 유용성을 확인하여 주고 있다

본 연구의 주목적은 전문검색시스템이 갖고 있는 높은 재현율을 유지하면서 낮은 정확률을 적합성 피드백을 이용하여 향상시키는 것이다. 표 8은 그 목적이 일부 성취되고 있음을 보여주고 있다. 예를들면, 적합성 피드백 후에 전문의 P_70(BI:0.050, TPI:0.0494), P_100(BI:0.0389, TPI:0.0375)은 초록-표제의 P_70(BI:0.0492,

<표 9> 초록-표제/BI모형에서 적합성 피드백의 효과

	NOFDBK	FDBK	Change%	Signif
Average Precision	0.3996	0.4175	4.48%	YES
Precision after 10	0.1687	0.1771	4.98%	NO
Precision after 30	0.0847	0.0928	9.56%	YES
Precision after 50	0.0578	0.0651	12.63%	YES
Precision after 70	0.0461	0.0492	6.72%	YES
Precision after 100	0.0352	0.0364	3.41%	YES
Recall after 10	0.5116	0.5021	-1.86%	NO
Recall after 30	0.6830	0.6857	0.40%	NO
Recall after 50	0.7356	0.7496	1.90%	NO
Recall after 70	0.7762	0.7971	2.69%	NO
Recall after 100	0.8152	0.8365	2.61%	NO

〈표 10〉 초록-표제/TPI모형에서 적합성 피이드백의 효과

	NOFDBK	FDBK	Change%	Signif
Average Precision	0.2811	0.3434	22.16%	High
Precision after 10	0.1361	0.1675	23.07%	High
Precision after 30	0.0751	0.0867	15.45%	High
Precision after 50	0.0530	0.0612	15.47%	High
Precision after 70	0.0417	0.0480	15.11%	High
Precision after 100	0.0336	0.0375	11.61%	High
Recall after 10	0.3814	0.4784	25.43%	High
Recall after 30	0.6004	0.6639	10.58%	High
Recall after 50	0.6804	0.7229	6.25%	High
Recall after 70	0.7158	0.7702	7.60%	High
Recall after 100	0.7923	0.8361	5.53%	High

TPI:0.0480), P_100(BI:0.0364, TPI:0.0375)을 증가하고 있으며 다른 것들에 비해 비교적 높은 재현율을 유지하고 있다.

이러한 적합성 피이드백 전후, 검색효율성에 있어 차이점에 대한 유의도검증(t검증)의 결과는 표

의 마지막 칸에 표시되어 있는데 만약 $\alpha=0.05$ 에서 유의하다면 "Yes"이고, 만약 $\alpha=0.01$ 에서 유의하다면 "High" 그밖에는 "No"라고 표시되어 있다. 〈표 9〉와 〈표 10〉에서 30, 50, 70, 100 문헌검색후에 정확률은 두 용어분산모형과 함께 적합성

〈표 11〉 전문/BI모형에서 적합성 피이드백의 효과

	NOFDBK	FDBK	Change%	Signif
Average Precision	0.2179	0.2535	16.34%	High
Precision after 10	0.0904	0.1120	23.89%	Yes
Precision after 30	0.0651	0.0815	25.19%	High
Precision after 50	0.0511	0.0622	21.72%	High
Precision after 70	0.0417	0.0501	20.14%	High
Precision after 100	0.0316	0.0389	23.10%	High
Recall after 10	0.3382	0.3838	13.48%	No
Recall after 30	0.5870	0.6463	10.10%	Yes
Recall after 50	0.6873	0.7435	8.18%	Yes
Recall after 70	0.7482	0.8293	10.84%	High
Recall after 100	0.7677	0.8783	14.41%	High

〈표 12〉 전문/TPI모형에서 적합성 피이드백의 효과

	NOFDBK	FDBK	Change%	Signif
Average Precision	0.2810	0.3525	25.44%	High
Precision after 10	0.1217	0.1386	13.89%	No
Precision after 30	0.0779	0.0884	13.48%	High
Precision after 50	0.0549	0.0627	14.21%	High
Precision after 70	0.0427	0.0494	15.69%	High
Precision after 100	0.0333	0.0375	12.61%	High
Recall after 10	0.3623	0.4705	29.84%	No
Recall after 30	0.6213	0.6882	10.77%	Yes
Recall after 50	0.6738	0.7677	13.94%	Yes
Recall after 70	0.7339	0.8227	12.10%	High
Recall after 100	0.7886	0.8863	12.39%	High

피이드백후 통계학적으로 유의한 향상이 있었음이 밝혀졌다.

전문/BI모형의 평균정확률은 적합성 피이드백 없이는 〈표 11〉에서 보여진 대로 매우 낮다. 이는 전문에서 BI모형은 아마도 부적절하다는 것을 보여준다. 왜냐하면 특히 비슷한 주제를 다루고 있는 문헌들간에 질문서의 용어가 전문문헌에서 한번 이상 출현할 기회는 그 문헌의 적합성여부에 관련 없이 많기 때문이다. 이는 BI모형에서는 문헌의 주제를 표현하기 위해 빈번히 출현하는 용어와 그렇지 않은 용어를 구별할 수 없음을 의미한다. 하지만 전문/BI모형은 10, 30, 50, 70, 100 문헌검색후 정확률이 적합성 피이드백후에 향상하고 있고 70, 100 문헌검색후 가장 높은 정확률을 보여주고 있다.

〈표 12〉에서 전문/TPI모형의 검색효율성은 전문/BI모형의 것보다 높은 것으로 나타났고 적합성 피이드백후 검색효율성의 향상은 주목할 만하다. 두 모형 모두 가설 1에 대한 검증은 평균정확률과 30,

50, 70, 100 문헌검색후의 정확률에 있어 $\alpha=0.01$ 에서 유의한 것으로 나타난다.

7. 결 론

전문데이터베이스의 급격한 증가는 1980년을 통해 현재까지 계속되고 있고 이러한 경향에 대처하기 위해 많은 정보학자들이 전문데이터베이스를 이용한 검색시스템의 검색효율성 및 실행가능성에 대해 연구하였다. 하지만 이에 대한 결과는 일관성이 부족하여 만족스럽지 못하였다. 이전의 연구들은 대부분 상용데이터베이스를 이용하고, 불리안모형에 의존하였으며 실제상황(operational environment)에서 이루어졌고 전문검색시스템의 문제점을 해결할 수 있는 새로운 검색기법을 개발하기보다는 검색효율성의 보고에 그쳤다.

전문검색시스템의 문제점은 낮은 정확률에 있는데, 본 연구의 주목적은 전문검색시스템의 문제점인 낮은 정확율을 해결하는 검색기법을 개발하는

데 초점을 두었다. 실험결과는 다시 한번 적합성 피이드백이 검색효율성의 향상을 가져다 준다는 것을 확인하였고 특히 적합성 피이드백은 전문/TPI 모형의 검색효율성을 증진시키는 데에 크게 기여하고 있음을 발견하였다. 비록 적합성 피이드백을 적용한 후에도 여전히 전문검색은 초록-표제를 이용한 검색의 효율성을 능가하고 있지 못하지만, 높은 재현율을 제공하는 전문검색시스템을 선호하는 정보센터(법률도서관이나, 특허도서관 등)의 경우에 정확률의 손실을 최소화하기 위하여 적합성 피이드백의 적용이 필요하다는 것을 본 연구는 제시하고 있다.

아직까지 전문검색시스템의 이용 및 효율성 연구는 초기단계에 있으며 지금까지 언급되어 온 전문검색시스템의 연구들도 전문검색시스템의 효율성에 대해 결론적인 해답을 제시하지는 못하고 있다. 향후 전문검색시스템에 대한 끊임없는 연구가 촉구된다. 특히 실험실내에서 적절한 독립변인의 통제와 함께 보다 큰 규모의 데이터베이스 및 다른 주제 분야의 문헌들을 이용하여 보다 정교한 적합성 피이드백 방법을 개발, 이를 이용한 실험연구가 이루어져야만 할 것이다. 이로부터 얻어지는 결과는 실제상황에도 일반화될 수 있는 결론을 얻는데 공헌하게 될 것이다.

참고문헌

- Barker, F.H., Veal, D.C., & Wyatt, B.K. (1972). Comparative Efficiency of Searching Titles, Abstracts and Index Terms in a Free-Text Data Base. *Journal of Documentation*, 28, 22-36.
- Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for Information Retrieval: Part I. Background and Theory. *Journal of Documentation*, 38, 61-71.
- Bhattacharyya, K. (1974). The Effectiveness of Natural Language in Science Indexing and Retrieval. *Journal of Documentation*, 30, 235-254.
- Blair, D.C., & Maron, M.E. (1990). Full-Text Information Retrieval: Further Analysis and Clarification. *Information Processing and Management*, 26, 437-447.
- Blair, D.C., & Maron, M.E. (1985). An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. *Communications of the ACM*, 28, 289-299.
- Bonzi, S. (1984). Terminological Consistency in Abstract and Concrete Disciplines. *Journal of Documentation*, 40, 247-263.
- Bookstein, A. (1983). Information Retrieval: A Sequential Learning Process. *Journal of the American Society for Information Science*, 34, 331-342.
- Bookstein, A., & Swanson, D.R. (1974). Probabilistic Models for Automatic Indexing. *Journal of the American Society for Information Science*, 25, 312-318.
- Cleveland, D. B., & Cleveland, A.D., & Wise, O.B. (1984). Less Than Full-Text Indexing Using a Non-Boolean Searching Model. *Journal of the American Society for Information*

- Science*, 35, 19-28.
- Cleverdon, C.W. (1962). Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. College of Aeronautics, Cranfield.
- Croft, W.B., & Harper, D. (1979). Using Probabilistic Models of Document Retrieval without Relevance Information. *Journal of Documentation*, 35, 285-295.
- Durkin, K., Egeland, J., Garson, L.R., & Terrant, S.W.(1980). An Experiment to Study the Online Use of a Full-Text Primary Journal Database. In: *Proceedings of the 4th International Online Information Meeting* (pp. 53-56). London: Learned Information Ltd.
- Fuhr, N. (1992). Probabilistic Models in Information Retrieval. *The Computer Journal*, 35, 243-255.
- Harper, D.J., & Van Rijsbergen, C.J. (1978). An Evaluation of Feedback in Document Retrieval Using Co-occurrence Data. *Journal of Documentation*, 34, 189-216.
- Harter, S.P. (1975a). A Probabilistic Approach to Automatic Keyword Indexing. Part I. On the Distribution of Specialty Words in a Technical Literature. *Journal of the American Society for Information Science*, 26, 197- 205.
- Harter, S.P. (1975b). A Probabilistic Approach to Automatic Keyword Indexing. Part II. An Algorithm for Probabilistic Indexing. *Journal of the American Society for Information Science*, 26, 280-289.
- Ide, E. (1971). New Experiments in Rlevance Feedback. In: *The SMART System - Experiments in Automatic Document Processing*, 337-354. Englewood Cliffs, NJ:Prentice Hall Inc.
- Keen, E.M. (1973). The Aberystwyth Index Languages Test. *Journal of Documentation*, 29, 1-35.
- Lesk, M.E., & Salton, G. (1968). Relevance Assessments and Retrieval System Evaluation. *Information Storage and Retrieval*, 4, 343-359.
- Losee, Jr., R.M. (1988). Parameter Estimation for Probabilistic Document Retrieval Models. *Journal of the American Society for Information Science*, 39, 8-16.
- Losee, Jr., R.M., Bookstein, A., & Yu, C. (1986). Probabilistic Models for Document Retrieval: A Comparison of Performance on Experimental and Synthetic Databases. In: *ACM Annual Conference on Research and Development in Information Retrieval* (pp. 258-264). New York: ACM Press.
- Maron, M.E. (1988). Probabilistic Design Principles for Conventional and Full-Text Retrieval Systems. *Information Processing and Management*, 24, 249-255.

- Mckinin, E.J., Sievert, M., Johnson, E.D., & Mitchell, J.A. (1990). The Medline/Full-Text Research Project. *Journal of the American Society for Information Science*, 42, 297-307.
- Raghavan, V.V., Shi, H., & Yu, C.T. (1983). Evaluation of the 2-Poisson Model as a Basis for using Term Frequency Data in Searching, In: *ACM Annual Conference on Research and Development in Information Retrieval* (pp. 88-100). New York:ACM Press.
- Rath, G., Resnick, A., & Savage, T. (1961) Comparison of Four Types of Lexical Indications of content. *American Documentation*, 12, 126-130.
- Ro, J.S. (1988a). An Evaluation of the Applicability of Ranking Algorithms to Improve the Effectiveness of Full-Text Retrieval. I. On the Effectiveness of Full-Text Retrieval. *Journal of the American Society for Information Science*, 39, 73-78.
- Ro, J.S. (1988b). An Evaluation of the Applicability of Ranking Algorithms to Improve the Effectiveness of Full-Text Retrieval. II. On the Effectiveness of Ranking Algorithms on Full-Text Retrieval. *Journal of the American Society for Information Science*, 39, 147-160.
- Robertson, S.E. (1981). The Methodology of Information Retrieval Experiment. In Sparck Jones, K. (Ed.), *Information Retrieval Experiment* (pp. 9-31). London, Butterworths.
- Robertson, S.E. (1977). The Probability Ranking Principle in IR. *Journal of Documentation*, 33, 294-304.
- Robertson, S.E., Thompson, C.L., & Bovey, J.D. (1986). Weighting, Ranking and Relevance Feedback in a Front-End System. *Journal of Information Science*, 12, 71-75.
- Salton, G. (1986). Another Look at Automatic Text-Retrieval Systems. *Communications of the ACM*, 29, 648-656.
- Salton, G., & Zhang, Y. (1986). Enhancement of Text Representations Using Related Document Titles. *Information Processing and Management*, 22, 385-394.
- Salton, G., & Buckley, C. (1990). Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41, 288-297.
- Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A Re-Examination of Relevance: Toward a Dynamic, Situational Definition. *Information Processing and Management*, 26, 755-776.
- Shaw, W.M. Jr., Wood, R.E., & Tibbo, H.R. (1991). The Cystic Fibrosis Database: Content and Research Opportunities.

- LISR*, 13, 347-366.
- Smeaton, A.F. (1984). Relevance Feedback and a Fuzzy Set of Search Terms in an Information Retrieval System. *Information Technology: Research and Development*, 3, 15-23.
- Smeaton, A.F., & Van Rijsbergen, C.J. (1983). The Retrieval Effects of Query Expansion on a Feedback Document Retrieval System. *The Computer Journal*, 26, 239-246.
- Sparck Jones, K. (1981a). Retrieval System Tests 1958-1978. In: Sparck Jones, K. (Ed.), *Information Retrieval Experiment* (pp. 213-255). London: Butterworths.
- Sparck Jones, K., (Eds.). (1981b). *Information Retrieval Experiment*. London: Butterworths.
- Swanson, D.R. (1979). Information Retrieval as a Trial-and-Error Process. *Library Quaterly*, 47, 128-148.
- Tenopir, C. (1989). Why Don't Librarians Use Full-TextDatabases? *Library Journal*, 114, 76-79.
- Tenopir, C. (1988). Searching Full-Text Databases. *Library Journal*, 113, 60-61.
- Tenopir, C. (1985). Full Text Database Retrieval Performance. *Online Review*, 9, 149-164.
- Terrant, S.W., Garson, L.R., Cohen, S.M., & Meyers, B.E.(1983). ACS Primary Journal Online Database. In: Williams, M.E. (Ed.), *Proceedings of the Fourth National Online Meeting* (p. 551). Medford, New Jersey: Learned Information Inc.
- Terrant, S.W., Garson, L.R., Cohen, S.M., & Meyers, B.E.(1982). The American Chemical Society Online Primary Journal Database. In: *Proceedings of the American Society for Information Science 45th Annual Meeting* (p. 379). White Plains, New York: Knowledge Publications, Inc.