

전문데이터베이스의 탐색특성에 관한 연구*

- 주제전문가와 탐색전문가 -

A Study of the Behaviours in Searching Full-Text Databases

- Subject Specialists vs. Professional Searchers -

이 응 봉(Eung-Bong Lee)**

목 차

- | | |
|----------------------------|----------------------------|
| 1. 서 론 | 3. 연구의 방법 |
| 1.1 연구의 필요성 | 3.1 가 설 |
| 1.2 연구의 목적 | 3.2 연구의 계획 |
| 1.3 용어의 정의 | 3.3 자료수집 |
| 2. 선행연구 | 3.4 연구의 제한점 |
| 2.1 전문데이터베이스의 특성에 관한 연구 | 4. 주제전문가와 탐색전문가의 탐색특성 비교분석 |
| 2.2 전문데이터베이스의 이용자에 관한 연구 | 4.1 탐색질문의 분석에 관한 비교분석 |
| 2.3 전문데이터베이스의 탐색시스템에 관한 연구 | 4.2 탐색과정에 관한 비교분석 |
| 2.4 전문데이터베이스의 탐색효율에 관한 연구 | 4.3 탐색결과에 관한 비교분석 |
| | 5. 결론 및 제언 |

초 록

본 연구는 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가는 어떠한 특징적인 차이가 있는지를 밝히려 하였다. 이 연구를 통하여 얻어진 결론을 요약하면, 주제전문가와 탐색전문가는 전문데이터베이스의 탐색에 있어서 탐색질문의 분석(탐색질문의 이해도, 탐색어 선정의 난이도, 탐색결과에 대한 기대수준), 탐색과정(탐색어 수, 연산자 수, 브라우징 문헌 수, 접속시간, 출력문헌 건당 소요시간, 적합문헌 건당 소요시간), 탐색결과(적합문헌 수, 재현율, 정확율, 총 탐색비용, 출력문헌 건당 소요비용, 적합문헌 건당 소요비용, 탐색결과에 대한 만족도)에 대하여 뚜렷한 차이를 나타내는 것으로 밝혀졌다.

* 본 논문은 1996년도 한국문헌정보학회 춘계발표회 발표논문을 수정 보완한 것임.

** 국토개발연구원 책임연구원

접수일자 : 96. 5. 2

ABSTRACT

The primary purpose of this study is to verify the difference of behavioural characteristics between the subject specialists and professional searchers in searching full-text databases. The major findings and conclusions from this study are summarized as follows. Analyses of Search questions(the degree of understanding with search questions, the degree of difficulty in selecting search terms, and the degree of expectation of search results), search processes(the number of search terms used, the number of Boolean operators and qualifiers used, the number of documents browsed and the search time(the connecting time, time to spend per one output document, time to spend per one relevant output document)) and search results(the searching efficiency(the number of relevant documents, the recall ratio and the precision ratio), the search cost(the total search cost, the search cost per one output document and the search cost per one relevant output document) and the degree of satisfaction with search results) are significantly different between the subject specialists and professional searchers in searching full-text databases.

1. 서 론

최근의 정보처리기술과 통신기술의 발달에 따른 컴퓨터의 도입과 네트워크화의 진전에 의하여 정보화는 산업활동, 경제활동, 연구활동 뿐만 아니라 일반가정에서의 개인생활 수준까지 파급되고 있다. 이러한 정보화의 진전과 정보의 가치에 대한 인식의 향상에 의하여 폭발적으로 증가하는 정보를 효율적으로 이용하기 위한 지적인 인프라스트럭처(infrastructure)로서의 데이터베이스의 중요성은 더욱 더 높아지고 있다.

오늘날의 정보 수집방법은 데이터베이스의 보급과 더불어 급속도로 변화하고 있다. 기존의 책자형태나 인쇄매체에 의한 정보의 수집은 컴퓨터와 통신회선을 이용한 수집으로 변하고 있으며, 온라인 방식으로 수행하는 데이터베이스 검색은 데이터의 수집에 대한 시간적, 공간적인 제약을 해결함과 동시에 대량의 데이터를 효율적으로 검색하는 것을 가능하게 하였다.

1960년대 중반에 미국에서 개발된 데이터베이스 검색시스템은 1970년대에 들어 통신기술의 발달과 함께 온라인화되었고, 1980년대 이후에는 세계적으로 널리 이용되고 있다. 데이터베이스의 형태도 1980년대 초반까지는 주로 서지데이터베이스였지만, 이후 다양한 발전을 거듭하여 전문데이터베이스 및 CD-ROM 데이터베이스로 발전하였고, 현재에는 하이퍼텍스트데이터베이스의 이용이 일반화되기에 이르렀다.

전문데이터베이스는 기존의 서지데이터베이스가 지닌 모든 특성 외에 정보탐색의 궁극적인 목적이라 할 수 있는 원문에 대한 즉시 접근 및 이용이 가능하며, 규모면은 물론 질적인 면에서도 급속적인 발전을 하고 있고, 이용율 또한 급증하고 있다. 이와 같은 데이터베이스의 다양화와 발전은 정보이용자들에게 정보검색에 대한 새로운 도전 뿐만 아니라 새로운 기회를 제공해 주고 있다.

1.1 연구의 필요성

전문데이터베이스는 최근 들어 컴퓨터 기록용량의 증대와 전자출판의 발달로 각광을 받고 있으며, 초기에는 신문기사, 법령 및 판례 등의 전문을 수록하는 것이 주종을 이루었지만, 최근에는 학술논문, 단행본의 전문 및 학술잡지의 본문기사도 수록되고 있다.

상용의 온라인 데이터베이스 서비스 분야에서도 데이터베이스의 전문화는 꾸준히 증가하고 있는데 예를 들면, 세계적인 데이터뱅크 DIALOG에는 1995년 9월 현재, 보유하고 있는 전체 503종의 데이터베이스 중에서 전문데이터베이스의 수는 92종에 이르고 있다(Dialog Information Service 1995). 그리고 최근에는 CD-ROM판 전문 데이터베이스 즉, 오프라인 방식의 전문데이터베이스의 보급도 급속도로 증가되고 있는 상황이다.

이와 같은 전문데이터베이스 수의 급속한 증가와 종류의 다양화에도 불구하고, 전문데이터베이스에 관한 연구는 서지데이터베이스에 관한 연구에 비하여 극히 미약하며, 광범위

한 연구가 이루어지지 않고 있는 실정이다.

전문데이터베이스에 관한 연구는 대부분 다음의 세가지 분야 첫째, 전문데이터베이스의 타당성과 가능성에 대한 이용자의 견해 둘째, 전문데이터베이스의 검색효율과 검색효과에 대한 평가 셋째, 검색효율 또는 검색효과를 개선하기 위한 데이터베이스와 탐색시스템의 수정 등을 중심으로 이루어지고 있다(Tenopir and Jung-Soon Ro 1991, 69).

특히, 전문데이터베이스의 탐색이나 이용자에 관한 연구는 아직까지는 초보적인 단계에 머물러 있으며, 여전히 일련의 경험에 의존한 전문직들에 의해 수행되고 있고, 대부분의 연구가 동일한 환경하에서의 최종이용자를 대상으로 하여 사회과학, 의학 및 화학분야의 전문데이터베이스에 치중하여 수행되고 있다.

본 연구는 이와 같은 연구경향에 착안, 기존의 연구와는 접근방법을 달리하여, 광범위한 주제분야를 포괄하고 있는 전문데이터베이스를 탐색대상으로 선정하고, 피험자 집단도 과학기술분야의 주제전문가(최종이용자)와 탐색전문가로 구분하여, 실제의 전문데이터베이스 탐색에 있어서 두 집단 사이의 탐색특성은 어떻게 다른지를 밝혀보려는 것이다.

본 연구의 결과는 전문데이터베이스 시스템의 설계자들에게 주제전문가와 탐색전문가의 탐색특성의 차이를 활용할 수 있는 기법이나 그 대응물을 개발하는데 도움을 주고, 전문데이터베이스의 이용자들에게 효율적인 탐색을 위한 하나의 지침을 제공하게 될 것이다.

1.2 연구의 목적

본 연구는 특정 분야의 주제에 대하여 전문적인 지식을 소유하고 있는 주제전문가와 전반적인 검색시스템 및 다양한 데이터베이스에 관한 지식과 경험이 풍부한 탐색전문가(정보전문가)의 두 집단 사이에는 전문데이터베이스의 탐색에 있어서 탐색질문의 분석, 탐색과정 및 탐색결과에 어떠한 차이가 있는지를 조사하여 비교분석하는 것이다.

본 연구의 목적은 전문데이터베이스의 탐색에 대한 시스템과 이용자의 상호작용인 주제전문가와 탐색전문가의 두 집단 사이에서 나타나는 탐색특성의 차이를 밝히는 것이다.

1.3 용어의 정의

본 연구에서 사용된 용어의 개념을 정의하면 다음과 같다.

- 1) 주제전문가 : 특정분야의 주제에 대하여 전문적인 지식을 소유하고 있으며, 자신의 정보요구를 충족시키기 위해서는 스스로 탐색을 수행할 수도 있지만, 타인을 위해서는 일상적으로 탐색을 수행하지 않는 최종이용자를 말한다.
- 2) 탐색전문가 : 전반적인 정보검색시스템 및 다양한 데이터베이스에 관한 지식과 경험이 풍부한 정보전문가로서 최종이용자를 대신하여 일상적으로 탐색을 수행하는 정보검색중개인을 말한다.
- 3) 내용지향적 탐색 : 탐색시 탐색질문에 대하여 용어의 선택이 다양하고, 여러가

지의 탐색기능을 사용하기 보다는 한 두 개의 질문공식을 사용하며, 탐색식의 작성이나 수정보다는 최종 탐색결과와 출력에 앞서 본문의 일부를 훑어보거나 읽어보는 브라우징에 치중하고, 탐색어도 통제어휘나 디스크립터를 사용하기 보다는 자연어 또는 특정한 언어를 사용하며, 정확한 답을 요구하는 종류의 탐색을 말한다.

- 4) 과정지향적 탐색 : 탐색시 탐색질문에 대하여 용어의 선택에 초점을 맞추고, 불연산, 위치 연산 및 제한 연산과 같은 다양한 탐색기능을 사용하며, 탐색식의 작성이나 수정에 치중하고, 탐색어는 자연어 보다는 통제어휘나 디스크립터를 주로 사용하며, 정확한 답 보다는 타당한 답을 요구하는 종류의 탐색을 말한다.

2. 선행연구

선행연구의 검토는 전문데이터베이스의 전반에 걸친 포괄적인 연구보다는 주로 본 연구의 주제와 관련이 있는 전문데이터베이스의 특성, 이용자, 탐색시스템 및 탐색효율에 관한 사례를 중심으로 살펴본다.

2.1 전문데이터베이스의 특성에 관한 연구

터런트(Terrant 1980, 191-219)는 전문데이터베이스는 서지사항이나 초록 등 2차정보만을 수록하지 않고 하나 이상의 본문 전체를

수록하고 있는 데이터베이스라고 하였다.

테노피(Tenopir 1985, 150)는 전문데이터베이스는 기존의 서지데이터베이스를 근거로 하여 개발된 것이기 때문에 서지데이터베이스의 특성을 포함하고 있다. 전문데이터베이스는 서지데이터베이스의 기본 출력형태와 같이 서명, 저자명, 기관명, 언어 및 출판사항 등에 대한 정보를 제공하며, 추가정보로서 전문이나 요약문 및 인용문까지 수록하고 있다고 하였다.

또한, 테노피(1985, 224)는 전문데이터베이스는 원정보에 대한 접근이 가능하고, 특정의 정보에 대한 접근이 용이하다는 장점을 가지고 있다. 서지데이터베이스의 탐색은 적합자료에 관한 정보를 입수하여도 원자료의 입수까지는 상당한 비용과 시간이 소요된다. 반면에, 전문데이터베이스의 탐색은 적합자료로 판단된 자료에 대한 원자료를 입수하는 데는 별도의 비용과 시간이 소요되지 않는다. 그리고 통제어를 이용한 서지데이터베이스 탐색 방식에서는 원문에 실려있지 않은 용어가 디스크립터어로 사용될 수 있으며, 특정의 용어도 일반적인 디스크립터어로 전환되어 입력되거나, 원문에서 사용된 원래의 용어자체가 누락될 수 있기 때문에 탐색어 선정에 상당한 주의를 기울여야 하는 단점을 지니고 있지만, 전문데이터베이스는 원문의 모든 용어가 대부분 디스크립터어가 될 수 있다는 장점을 가지고 있다. 그러나 접근점이 많으면 많을수록 그만큼 출력전수도 또한 많아지기 때문에 전문데이터베이스의 탐색은 재현율이 높은 반면에 정확률은 떨어지게 되며, 또한 온라인 방식의 경우에는 부적합 정보의 출력전수가

많으면 많을수록 이용자의 비용부담은 늘어날 뿐만 아니라, 적합정보의 선택에 있어서도 혼란을 일으키게 된다는 단점을 가지고 있다고 하였다.

러브(Love 1985, 274)는 서지데이터베이스가 인용문헌에 대한 정보를 그 갯수로만 제한하는데 비하여, 전문데이터베이스는 인용문의 서지사항 전부를 원문 그대로 제공할 뿐만 아니라, 색상표, 삽도 및 사진과 같은 화상정보를 파일에 입력시켜 이에 대한 탐색도 가능하도록 되어 있다고 하였다.

마드링(Madelung 1981, 45-48)은 전문데이터베이스는 불용어와 알파벳의 석자 이상의 모든 명사들을 전부 키워드화하기 때문에 파일에 신규정보를 입력할 경우 색인전문가가 필요없게 된다. 이와 같은 전문데이터베이스의 특성 때문에 색인전문가에 대한 인건비를 줄일 수 있다는 점에서 전문데이터베이스의 개발에 더욱 더 박차를 가하게 되었다는 주장을 하였다.

페이지(Pagell 1988, 255-262)은 전문데이터베이스는 이용자에게 있어서 정보자원으로서 뿐만 아니라, 정보검색효율을 향상시키고 문헌전송을 용이하게 하는 수단으로서 매우 중요하다고 언급하면서 전문데이터베이스의 제한점에 대하여 다음의 일곱가지를 들고 있다. 첫째, 원 잡지에 포함되어 있는 모든 기사 중 단지 일부분만이 전문데이터베이스에서 이용이 가능하다는 점 둘째, 전문 온라인파일은 원문헌과 동일하지 않다는 점 셋째, 주요 기사나 컬럼을 처음부터 끝까지 수록하는 것이 아니라 발췌하여 수록하고 있으며 사실, 표 및 도

형 데이터가 수록되지 않는다는 점 넷째, 정기간행물의 경우에는 출판 후 파일화되기까지 시간이 오래 걸린다는 점 다섯째, 동일 잡지에 대한 정보검색시 서지화일과 차이가 난다는 점 여섯째, 인접연산자나 필드제한자와 같은 명령어들의 사용방법이 복잡하여 탐색 전문가에게는 효과적일지 모르지만, 초보자에게는 어렵다는 점 일곱째, 데이터베이스 시스템이 자주 변환된다는 점 등이다. 그러나 페이지가 지적한 일곱가지의 제한점들은 최근의 정보기술의 발달에 따라 수정·보완이 이루어지고 있다."

테노퍼(1988, 263-270)는 전문데이터베이스가 지니고 있는 문제점에도 불구하고, 현재 이용가능한 전문데이터베이스의 분야는 매우 다양할 뿐만 아니라 전문데이터베이스에 수록되어 있는 문헌의 형태 또한 다양하기 때문에 전문데이터베이스에 대한 이용은 계속 확대될 것이고 탐색전략의 개발 또한 늘어날 것이라는 주장을 하였다.

2.2 전문데이터베이스의 이용자에 관한 연구

미국화학회(American Chemical Society)에서는 화학자들을 대상으로 전문데이터베이스인 ACS Journals Online의 이용에 관한 조사

1) 페이지가 지적한 제한점들 중 "넷째, 정기간행물의 경우에는 출판 후 파일화 되기까지 시간이 오래 걸린다는 점"은 최근의 전자출판의 발달로 인하여 거의 해결이 되었고 나머지 여섯가지의 제한점들도 수정·보완이 이루어지고 있지만, 여전히 전문데이터베이스가 해결하여야 할 문제점으로 남아 있다.

연구를 수행하였다. 조사결과를 보면, 화학자들은 다양한 목적을 위하여 전문데이터베이스를 이용하는 것으로 나타났는데 특히, 본문에 포함되어 있는 사실정보(fact information)와 기사의 주제에 대한 주변정보를 찾거나 또는 서지검색의 보조도구로 이용하는 것으로 나타났다. 그러나 대부분의 화학자들은 전문데이터베이스에서 브라우징을 하거나 읽기를 하는 것이 아니라, 온라인 방식으로 관련기사의 위치를 확인하여 해당 기사를 복사하는 것으로 나타났는데, 이러한 이유는 전문데이터베이스에서의 그래픽의 지원부족이나 관련장비의 질적인 문제 때문인 것으로 밝혀졌다. 하지만, 이러한 화학자들의 ACS Journals Online에 대한 적극적인 반응은 미국화학회에서 발간하는 잡지의 전문데이터베이스화를 촉진하는 계기가 되었다(Garson and Cohen 1983, 40 : Winkler 1983, 23-24).

콜린과 플러글(Collen and Flagle 1985, 2768-2774)은 1982년 Mead Data Central사가 구축한 의학 관련분야의 전문데이터베이스인 MEDIS(Medical Information System)를 탐색대상으로 선정하여 500명 이상의 의학분야 종사자들이 3개월간 수행한 9,377건의 탐색활동을 조사 검토하였다. 이 조사는 이용자 반응을 중요시한 것으로 전문데이터베이스의 출력정보에 대한 적합성 여부는 51.3%가 긍정적이었고, 23.4%가 부정적이었으며, 25.3%가 무응답으로 나타났다. 그리고 무응답자 중 41.0%는 출력정보의 적합성 유무에 관계없이 탐색방법의 편리성에 대해서 긍정적인 것으로 나타났다.

2.3 전문데이터베이스의 탐색시스템에 관한 연구

페이젤(1987, 33-47)은 Magazine ASAP 전문데이터베이스를 대상으로 인쇄물과 데이터베이스 버전(version)에 대한 비교연구를 수행하였는데, 연구결과를 보면 데이터베이스 버전에는 잡지의 전체내용이 모두 포함된 것이 아니었으며, 수록기간도 다르고, 데이터베이스의 파일에 따라서는 잡지명도 차이가 나는 것으로 밝혀졌다. 이 연구의 결론은 Magazine ASAP의 인쇄본은 주요 기사의 전문을 이용자에게 제공하는 유용한 도구가 될 수 있지만, Magazine ASAP 전문데이터베이스는 인쇄본에 대한 완전한 대응물은 될 수 없다는 것이었다.

블레어와 메이런(Blair and Maron 1985, 289-299)은 1985년 STAIRS(STorage and Information Retrieval System)를 사용하여 전문 문헌검색시스템의 평가에 관한 연구를 수행하였는데, 이 연구에서는 두명의 탐색자가 최소한 75%의 적합문헌을 검색하려고 시도하였지만, 탐색결과는 20%에도 못미치는 적합문헌을 검색하였다. 이와 같은 이유는 규모가 큰 데이터베이스에 적용되는 전문 검색시스템은 어떠한 탐색환경하에서도 검색을 제대로 수행할 수 없다는 것이었다. 이와 같은 점에 대해서는 테노퍼(1988, 263-270)도 전문데이터베이스의 탐색은 전통적인 서지데이터베이스에 비하여 더 많은 부적합 문헌을 생산해낸다는 것을 지적한 바 있다.

2.4 전문데이터베이스의 탐색효율에 관한 연구

러브(1985, 273-282)는 미국화학협회(American Chemical Association)에서 작성한 전문데이터베이스인 ACS Journals Online을 대상으로 전문데이터베이스의 검색에서 출력되는 부적합 정보의 양과 얼마만큼 부적합 정보를 감소시킬 수 있는가에 관한 조사연구를 수행하였다. 이 연구에서는 탐색전략과 탐색 질문이 적절하게 계획되었다는 가정하에, 40회의 탐색실험에서 23%가 100%의 정확률을 나타냈고, 35%가 90%의 정확률을, 88%가 50% 이상의 정확률을 나타내는 결과를 얻었는데, 이러한 수치는 불 연산자 외에 위치 연산자인 "ADJ", "WITH" 및 "SAME"을 적절하게 사용하였기 때문인 것으로 밝혀졌다. 이 연구에서 얻은 결론은 특정의 용어를 탐색어로 사용할 때 최상의 탐색결과를 구할 수 있으며, 광범위하게 사용되는 일상의 용어를 탐색어로 사용할 때는 정확률이 급격하게 떨어진다는 것이다. 그리고 위치 연산자가 탐색어와 탐색어 사이에서 정확률을 높여 주는 주요 연결어라는 가정하에, 전문데이터베이스의 탐색에서는 "AND" 연산자가 반드시 탐색문에 포함되어야만 양호한 결과를 얻을 수 있는 것으로 나타났다.

프랭클린(Franklin 1983, 6-8) 등은 의학분야의 전문데이터베이스인 MEDIS를 대상으로 기술적인 정보내용을 필요로 하는 연구자들에게 전문데이터베이스의 효율성에 관한 피드백(feedback)을 받았는데, 이 연구에서는 전

문데이터베이스를 통한 검색자료의 약 50-70%가 부적합한 것으로 나타났다. 그러나 시간, 거리상의 이유로 도서관 이용에 애로를 느끼고 있는 과학자나 연구자들의 경우에는 비록 경제적인 비용지출을 수반하더라도 필요한 원정보를 검색과 동시에 입수할 수 있는 전문데이터베이스가 서지정보만을 제공하는 서지데이터베이스에 비하여 효율도가 큰 것으로 나타났다.

버틀러(Butler 1985, 79-84)는 온라인 전문데이터베이스의 경제적인 측면을 언급하면서 특히, 온라인 전문데이터베이스의 가장 큰 약점은 한 페이지의 출력에 약 US1.50-US5.00이나 소요되는 비용문제를 제기하고 있다.

테노퍼(1984)는 탐색효율에 관한 다음의 두 가지 가설을 설정하여 이를 검증하였다. 첫째 가설은 "정기간행물의 전문에서 추출한 주제어나 자연어를 사용한 온라인 탐색은 높은 재현율과 낮은 정확률을 나타낼 것이며", 둘째 가설은 "이와 같은 방법을 사용한 탐색은 표제, 초록, 통제어 및 이들의 조합을 이용한 탐색보다 출력자료의 수가 월등하게 많을 것이다"라는 것이었다. 그녀는 이 연구를 위하여 BRS 시스템의 전문데이터베이스인 HBRO(Harvard Business Review Online)를 대상으로 하여, 31개의 탐색주제를 작성하여, 3명의 교수에게 피드백을 받았으며, 피드백 결과는 전문, 초록, 통제어, 서명, 서지사항의 조합 및 서지사항과 전문의 조합으로 구성되어 조사하였는데, 두가지의 가설이 모두 입증된 것으로 나타났다. 그러나 테노퍼는 이 연구에서 전문탐색은 비교적 정확한 탐색결과를

제공하지만, 대부분의 경우에 있어서 전문탐색 한가지만으로는 정확률이 높은 정보를 제공하지 못하며, 이에 반하여 통제어나 초록을 이용한 탐색은 좀 더 효용성이 있는 정보를 제공할 수 있으므로, 전문과 통제어 뿐만 아니라 초록을 함께 조합하는 탐색방법이 필요하다는 것을 언급하고 있다.

노정순(1985)은 위에서 언급한 테노퍼의 실험에 대한 적합성 판단에 관하여 검증을 하고, 재현율의 손상 없이도 전문데이터베이스의 낮은 정확률을 높이는 검색시스템을 설계하기 위하여 탐색어의 가중치 알고리즘(algorithm)에 관한 연구를 수행하였다. 노정순은 이를 증명하기 위하여 다음과 같은 두가지의 탐색실험을 수행하였다. 첫번째 실험에서는 테노퍼와 마찬가지로 Harvard Business Review지를 대상으로 하여, 테노퍼의 탐색주제 31개 중 9개를 임의로 추출한 후 탐색질문을 완성시켜 이 완성된 탐색질문을 BRS의 전문데이터베이스인 HBRO에서 탐색을 수행하였는데 탐색 결과를 보면, 전문데이터베이스에서의 탐색효율은 재현율이 높고 정확률은 낮은 것으로 나타났다지만, 그러나 포괄적인 탐색을 위해서는 전문데이터베이스가 유리하다고 하는 테노퍼와 동일한 결론을 제시하였다. 두번째 실험에서는 다음의 두가지 가설 "첫째, 자동색인시스템에서 사용하는 발생빈도수에 의한 가중치 부여방법은 재현율의 손상없이도 전문데이터베이스의 낮은 정확률을 향상시킬 수 있을 것이다. 둘째, 29개의 알고리즘은 정확률을 높이는 성능면에서는 차이가 없을 것이다."를 설정하여 탐색실험을 수행하였는데, 결론은 본문

가중치법을 실제로 사용한 전문 탐색방식은 서지 탐색방식에서와 마찬가지로 재현율을 같게 할 수 있을 뿐만 아니라, 정확률도 높일 수 있으므로, 전문탐색시 발생하는 높은 재현율과 낮은 정확률이라는 약점을 해결할 수 있다는 것이다.

남영준(1987)은 HBRO를 대상으로 전문데이터베이스 탐색방식과 서지데이터베이스 탐색방식을 데이터 항목으로 선정하여 양방식에서의 재현율, 정확률 및 비용대 효과를 비교하였다. 연구결과를 보면, 표제는 탐색의 접근점이 되지 못하였으며, 탐색범위를 전문까지 확장할 경우에는 표제, 초록 및 통제어로 범위를 정하였을 경우에 비하여 출력건수가 약 3.6배 증가하는 것으로 나타났으며, 적합정보의 수도 또한 약 3.4배가 많았다. 그리고 전문데이터베이스 탐색방식이 서지데이터베이스 탐색방식에 비하여 재현율은 높고 정확률은 낮게 나타났으나, 비용적인 측면에서는 전문데이터베이스의 탐색방식이 서지데이터베이스의 탐색방식 보다 경제성이 높은 것으로 나타났다.

문성빈(1993)은 확률모형에 근거하여 적합성 피드백(relevance feedback)을 사용함으로써, 전문데이터베이스 시스템에서의 탐색효율은 높은 재현율과 낮은 정확률을 나타낸다는 문제점을 해결하기 위한 연구를 수행하였다. 연구결과를 보면, 전문데이터베이스의 탐색은 적합성 피드백을 혼성하여 사용한다 하더라도 재현율과 같은 높은 수준의 정확률에는 도달할 수 없는 것으로 나타났지만, 그러나 적합성 피드백의 효과는 매우 크게 작용하였다는 것이다.

3. 연구의 방법

3.1 가 설

본 연구의 이론가설은 다음과 같다.

전문데이터베이스의 탐색에 있어서 주제전문가는 내용지향적 탐색을 수행할 것이고, 탐색전문가는 과정지향적 탐색을 수행할 것이다. 따라서, 전문데이터베이스의 탐색에 있어서 탐색질문의 분석, 탐색과정 및 탐색결과는 주제전문가와 탐색전문가 사이에 유의한 차이를 나타낼 것이다.

이러한 추가설을 더 구체적으로 분석하여 검증가능한 작업가설화하면 다음과 같다.

가설 1 : 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가는 탐색질문의 분석(① 탐색질문에 대한 이해도, ② 탐색어의 선정에 대한 난이도, ③ 탐색결과에 대한 기대수준)에 유의한 차이를 나타낼 것이다.

가설 2 : 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가는 탐색과정(① 탐색어 수, ② 연산자 수, ③ 브라우징 문헌 수, ④ 탐색 시간)에 유의한 차이를 나타낼 것이다.

가설 3 : 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가는 탐색결과(① 탐색효율, ② 탐색비용, ③ 탐색결과에 대한 만족도)에 유의한 차이를 나타낼 것이다.

3.2 연구의 계획

3.2.1 변인의 설정 및 측정요소

본 연구의 가설을 검증하기 위하여 설정한 변인은 주제전문가와 탐색전문가의 탐색질문에 대한 분석, 탐색과정 및 탐색결과에 관한 것이다.

본 연구는 주제전문가와 탐색전문가의 두 집단을 피험자로 선정하여 독립변인으로 하고, 주제전문가와 탐색전문가의 탐색질문에 대한 분석과 탐색과정 및 탐색결과를 종속변인으로 하여 주제전문가와 탐색전문가에게 제공된 공통적인 탐색환경과 관련된 여러 요인들을 통제변인으로 하여 실험한 결과를 수집하여 처리한다.

3.2.1.1 탐색질문의 분석

탐색질문의 분석이란 실제의 탐색을 위한 하나의 준비과정으로 주어진 탐색질문의 주제에 대하여 이해를 구하며, 관련자료를 활용하여 탐색에 필요한 탐색어를 선정하고, 탐색결과에 대한 기대수준을 파악하는 과정이라 할 수 있다.

탐색질문의 분석은 질문지를 사용하며, 탐색질문의 분석에 관한 변인으로는 다음의 세가지 1) 탐색질문의 주제에 대한 이해도, 2) 탐색

어 선정에 대한 난이도, 3) 탐색결과에 대한 기대수준 등을 설정하기로 한다.

3.2.1.2 탐색과정

탐색과정(search process)이란 탐색절차(search procedure) 또는 탐색노력(search effort)이라는 용어로도 사용되며, 탐색자가 데이터베이스에서 탐색을 시작하여 최종의 탐색 결과를 얻을 때 까지의 전 과정을 말하는 것으로 탐색에 투입된 노력의 양이 직접 반영된 것이라 할 수 있다.

피델과 서겔(Fidel and Soergel 1983, 163-170)은 온라인 데이터베이스의 탐색에 영향을 미치는 주요 변인을 제시하면서 탐색과정과 관련된 일곱가지의 주요변인으로 1) 접속시간, 2) 속도(단위시간당 명령어 수), 3) 탐색계획 수립에 소요된 시간, 4) 탐색문(탐색명령문, 탐색용어 및 논리조합의 수), 5) 시스템상의 특수기능의 활용정도, 6) 다양한 탐색전략, 7) 환류(feedback) 등을 언급하고 있다.

페니첼(Fenichel 1981, 23-32)은 탐색과정에 영향을 미치는 주요 변인으로 1) 명령어 수, 2) 디스크립터 수, 3) 세트(set) 검토 및 횟수, 4) 사이클(cycle) 수, 5) 탐색어의 변경 횟수, 6) 접속시간, 7) 속도, 8) 실수 등의 여덟가지를 설정하고 있다.

본 연구에서는 위에서 언급한 변인들을 참고하여 탐색에 영향이 클 탐색과정의 주요 변인으로 1) 탐색어 수, 2) 연산자 수, 3) 브라우징 문헌 수, 4) 접속시간, 5) 출력문헌 건당 소요시간, 6) 적합문헌 건당 소요시간 등의 여섯가지를 설정하며 다음과 같이 측정한다.

① 탐색어 수

정보탐색에 사용된 탐색어의 수로, 인접연산자 기호(w)가 사용된 용어는 하나의 탐색어로 취급하였으며, 용어절단기법의 기호(?)를 사용하거나, 따로 입력된 동일어근의 단어는 각각 다른 탐색어로 취급하고, 잘못 입력된 단어는 제외한다.

② 연산자 수

연산자는 탐색어를 논리적으로 결합하기 위한 불 연산자와 레코드 상의 특정부분을 제한하는 제한 연산자(어구검색시 용어사이의 특정순서와 인접성을 표시하는 위치 연산자는 제한 연산자의 범위에 포함시킴) 등의 두가지로 구분한다.

③ 브라우징 문헌 수

탐색의 최종결과를 출력하기 전 검색결과 중의 일부분을 온라인상에서 화면으로 읽거나 또는 훑어 본 문헌 수를 계산한다.

④ 접속시간

단말기상에서 소요된 Logon부터 Logoff까지의 시간으로 DIALOG에서 제시된 시간을 초단위로 환산한다.

⑤ 출력문헌 건당 소요시간

접속시간을 최종 출력문헌 건수로 나누어 초단위로 환산한다.

⑥ 적합문헌 건당 소요시간

단위시간이라고도 말하며, 다음과 같은 식

으로 구한다.

$$T = \frac{S}{A}$$

T : 적합문헌 건당 소요시간(초)

S : 접속시간(초)

A : 탐색자에 의하여 출력된 검색문헌 중에서 탐색질문에 대하여 적합하다고 평가된 검색문헌의 수

3.2.1.3 탐색결과

본 연구에서는 탐색결과를 나타내는 변인으로 1) 적합문헌 수, 2) 재현율, 3) 정확률, 4) 총 탐색비용, 5) 출력문헌 건당 소요비용, 6) 적합문헌 건당 소요비용, 7) 탐색결과에 대한 만족도 등의 일곱가지를 설정하며 다음과 같이 측정한다.

① 적합문헌의 수

적합문헌의 수를 구하는 것은 탐색효율의 측정에 있어서 가장 중요한 요소이다. 본 연구에서는 적합문헌의 수를 구하기 위하여 다음의 두가지 단계를 사용하기로 한다. 첫번째 단계는 본 연구자와 탐색질문의 작성자가 각각 개별적으로 탐색을 수행한 후 두가지의 탐색결과를 합쳐 중복된 검색문헌을 제외하여 잠정적인 적합문헌집단을 만든다. 두번째 단계는 피험자들의 검색문헌 중에서 잠정적인 적합문헌집단에 포함되어 있지 않은 문헌들을 조사하여 그 중에서 적합문헌이 있으면 이것을 첫번째 단계에서 만든 잠정적인 적합문헌집단에 추가시켜 계산한다.

② 재현율

재현율을 측정하는 것은 상당한 시간과 노력을 필요로 한다. 실제로 어떤 검색시스템의 재현율을 표시하는 수치 자체는 아무런 의미도 없으며, 재현율이나 정확률의 어느 쪽도 그 자체로써 색인이나 정보검색시스템의 효율을 완전하게 표시하지는 못하지만, 두 측정치를 같이 사용하면 가장 합리적인 시스템 효율의 측정방법이 될 수 있다(최성진 1988, 216-217). 본 연구에서는 재현율을 측정하기 위하여 랑카스터(Lancaster 1979, 129-130)가 제시한 概算(estimation) 재현율을 적용하기로 한다. 적합문헌에 대한 재현율은 다음과 같은 식으로 구한다.

$$R = \frac{A}{B} \times 100$$

R : 재현율(%)

A : 탐색자 각자가 출력한 검색문헌 중에서 탐색질문에 대하여 적합하다고 평가된 검색문헌의 수

B : 탐색질문에 대하여 적합한 전체 문헌의 수

③ 정확률

적합문헌에 대한 정확률은 다음과 같은 식으로 구한다.

$$P = \frac{A}{C} \times 100$$

P : 정확률(%)

A : 탐색자 각자가 출력한 검색문헌 중에서 탐색질문에 대하여 적합하다고 평가된 검색문헌의 수

C : 탐색자 각자가 출력한 전체 검색문헌의 수

- ④ 총 탐색비용
단말기상에서 소요된 Logon부터 Logoff까지의 비용으로 DIALOG에서 제시된 비용을 원화 및 US로 계산한다.
- ⑤ 출력문헌 건당 소요비용
총 탐색비용을 출력문헌 건수로 나눈 값으로 원화 및 US로 계산한다.
- ⑥ 적합문헌 건당 소요비용
총 탐색비용을 적합문헌 건수로 나눈 값으로 원화 및 US로 계산한다.

- ⑦ 탐색결과에 대한 만족도
질문지를 사용하여 측정하며, 리커트(R. Likert)의 5점척도법을 적용하여 계산한다.

본 연구에서 설정된 변인과 측정요소를 정리하면 <표 3-1>과 같다

3.2.2 탐색질문의 선정

탐색질문을 결정하는 방법으로는 첫째, 피험자인 탐색자 스스로 원하는 탐색질문을 선택하게 하는 방법 둘째, 탐색에 참여하지 않는 이용자들로부터 탐색질문에 대하여 신청을 받

<표 3-1> 본 연구에서 설정한 변인과 측정요소

구 분	변 인	측 정 요 소
독립 변인	탐색집단	주제전문가 탐색전문가
종속 변인	탐색질문의 분석	탐색질문의 주제에 대한 이해도 탐색어 선정에 대한 난이도 핵심 탐색어 선정에 대한 난이도 탐색결과에 대한 기대수준
	탐색과정	탐색시 사용한 탐색어 수 탐색시 사용한 불 연산자 수 탐색시 사용한 제한 연산자 수 탐색시 브라우징한 문헌 수 접속시간 출력문헌 건당 소요시간 적합문헌 건당 소요시간
	탐색결과	적합문헌 수 재현률 정확률 총 탐색비용 출력문헌 건당 소요비용 적합문헌 건당 소요비용 탐색결과에 대한 만족도
탐색 질문	탐색질문 세 가지의 난이도를 서로 비슷하게 조절	

는 방법 셋째, 탐색자에게 다수의 탐색질문을 제공하는 방법 등의 세가지 방법이 있다 (Walker 1988, 403-410 : Sullivan 1990, 27-42). 그러나 대부분의 실험연구에서는 피험자들의 탐색질문에 대한 사전통제를 위하여 탐색자에게 동일한 탐색질문을 제시하는 것이 일반적이다(Penhale and Taylor 1986, 212-220 : Fenichel 1981, 23-42).

본 연구를 위한 탐색질문은 실험에 참가하지 않는 해당분야의 주제전문가와 탐색전문가가 함께 다수의 예상 질문을 작성하고, 실제의 탐색을 수행하여 모범답안을 작성한 후 그 중에서 본 연구를 수행하기에 적절하다고 판단된 세가지의 탐색질문을 선정하였으며, 신뢰도의 향상을 위하여 탐색질문에 포함된 개념의 성격과 수를 기준으로 하여 난이도를 서로 비슷하게 조절하였다²⁾

그리고 탐색질문은 피험자의 전공 및 연구분야와 일치시키기 위하여 주제분야가 명확하게 구분되는 것은 아니지만, 다음과 같은 세가지로 구분하였다. "탐색질문 1"은 전기·전자분야와 관련이 있고, "탐색질문 2"는 물리·화학·재료분야와 관련이 있으며, "탐색질문 3"은 컴퓨터·통신분야와 관련이 있다. 세가지의 탐색질문은 다음과 같다.

탐색질문 1 : 미래기술의 핵심이라 할 수 있는 광기술 또는 광전자학 분야의 레이저에 대한 최근의 연구동향에 관하여 탐색하시오

탐색질문 2 : 금속 수소화물 축전지의 최근의 연구동향에 관하여 탐색하시오

탐색질문 3 : 정보사회의 총아로서 각광을 받고 있는 개인 휴대용 통신서비스에 대한 최근의 연구동향에 대하여 탐색하시오

3. 2. 3 탐색집단의 선정

피험자는 독립적인 연구(원)소(이하 연구소라 한다.)에서 활동하는 과학기술분야의 연구원(주제전문가)들과 동일분야에서 활동하는 탐색전문가들을 대상으로 하며, 대학에 설치된 과학기술분야 연구소의 연구원들과 학자들 및 탐색전문가들은 본 연구에서 제외한다.

피험자 집단은 주제전문가와 탐색전문가의 두 집단을 선정하였는데, 주제전문가는 탐색질문과 관련된 주제분야의 연구를 수행하였거나 현재 연구를 수행하고 있는 연구원들 중에서 각 질문별로 15명씩 총 45명을 선정하였고 탐색전문가는 15명을 선정하였다. 주제전문가와 탐색전문가의 전공 및 학력분포는 <표 3-2>, <표 3-3>과 같다.

주제전문가는 각 질문별로 15명씩 총 45명을 선정한 반면, 탐색전문가는 15명만 선정하였는데 그 이유는 탐색전문가는 세가지의 탐색질문 모두를 답하기 때문이다. 따라서 실제로 탐색에 참여하는 피험자 수는 주제전문가

2) 본 연구를 위한 탐색질문의 작성 및 선정에는 탐색질문의 주제분야와 관련이 있는 대학교수 3인과 피험자로 선정되지 않은 과학기술분야의 연구원 3인, 그리고 본 연구자를 포함하여 총 7명이 참여하였고, 최종 탐색결과와 출력문헌에 대한 적합성의 평가는 교수 3인이 수행하였다.

〈표 3-2〉 주제전문가와 탐색전문가의 전공별 분포

구 분	주제전문가			탐색전문가
	탐색질문1	탐색질문2	탐색질문3	탐색질문 1, 2, 3
전기·전자공학	13	-	5	-
물리·화학	-	9	-	1
재료·금속공학	-	6	-	-
기계·산업공학	2	-	2	1
컴퓨터공학	-	-	8	2
문헌정보학	-	-	-	10
경영정보학	-	-	-	1
합 계	15	15	15	15

〈표 3-3〉 주제전문가와 탐색전문가의 학력별 분포

구 분	주제전문가			탐색전문가
	탐색질문1	탐색질문2	탐색질문3	탐색질문 1, 2, 3
학 사	1	1	1	10
석 사	10	11	11	5
박 사	4	3	3	-
합 계	15	15	15	15

45명과 탐색전문가 15명을 합하여 총 60명이 된다. 그러나 세가지의 탐색질문에 답하는 기준으로 환산하면, 탐색질문 한가지당 주제전문가와 탐색전문가가 각각 15명씩 합계 30명이 되므로, 총 90명(탐색질문 3종 X 30명)이 된다.

3.2.4 탐색대상 전문데이터베이스

모든 주제분야의 문헌을 포괄적으로 수록하고 있으며, 인쇄본에 수록되어 있는 모든 기사를 전부 데이터베이스에 수록하고 있는 전문 데이터베이스 McGraw-Hill Publications Online(DIALOG File 624)을 탐색대상으로

선정하고(BiblioData 1995), DIALOG 시스템을 이용한다. McGraw-Hill Publications Online은 McGraw-Hill사에서 발행하는 34종의 정기간행물 및 뉴스레터의 전문을 수록하고 있는 전문데이터베이스로서 1995년 9월 현재 500,000여건의 레코드가 수록되어 있으며, 매주 약 1,800여건의 신규 레코드가 추가되고 있다.³⁾

3) 본 연구자가 1995년 9월 15일에 DIALOG 시스템의 DIALOG Bluesheet(File 415)를 온라인으로 연결하여 검색한 자료를 참고하였다.

3.2.5 탐색시스템 환경

본 연구에서는 시간 및 비용요인을 측정하여야 하기 때문에 다음과 같이 동일한 환경의 탐색시스템을 사용하기로 한다.

피험자 개개인의 탐색은 개인용 컴퓨터 단말기를 사용하며, 공중전화교환망을 통한 다이얼업(dial-up) 방식을 채택하고, 통신속도는 9,600bps로 한다. 통신소프트웨어는 이야기 6.1 버전을 사용하며, 탐색시간이나 탐색비용에 대해서는 아무런 제약을 두지 않는다. 그리고 탐색결과는 시스템에 자동적으로 기록되도록 하여 탐색이 종료된 후 디스켓에 다운로드(download)를 받아 나중에 프린터로 출력한다.

피험자 개개인의 탐색은 각자가 근무하고 있는 기관에서 수행하며, 탐색질문의 분석이나 실제의 탐색과정에 도움이 되도록 탐색메뉴얼, 영한사전, 한영사전, 영영사전 등의 사전류, 탐색질문의 주제와 관련된 시소러스 등을 제공한다.⁴⁾ 피험자들은 본 연구자의 입회하에 탐색을 수행하는데, 본 연구자는 통신 및 시스템상의 장애문제에 대한 해결을 제외한 그 밖의 부분에는 관여하지 않는다.

3.2.6 정보탐색의 실시방법

- ① 피험자 개개인별로 간단한 탐색교육을 실시한다.
- ② 피험자 개개인에게 탐색메뉴얼, 사전류 및 시소러스 등의 자료를 제공한다.
- ③ 탐색은 시간과 비용에 전혀 제약을 두지 않으며, 격리된 장소에서 각자 수행하도록 한다.
- ④ 탐색시 발생할 수도 있는 통신 및 시스

템상의 장애를 해결하기 위하여 연구자가 다소 떨어진 장소에서 대기한다.

- ⑤ 탐색을 위하여 DIALOG 시스템에 접속되는 시점(Logon)부터 탐색이 종료되어 시스템을 종결하는 시점(Logoff)까지의 모든 데이터는 자동적으로 시스템에 기록되도록 한다.
- ⑥ 탐색질문을 제공한 다음 실제의 탐색에 앞서 인적사항 및 정보탐색경험에 관한 질문, 탐색질문의 분석에 관한 질문에 응답케 한다.
- ⑦ 탐색종료 후 탐색과정 및 탐색결과를 해석하는데 참고하기 위하여 탐색결과와 만족도 평가에 관한 질문에 응답케 하고, 본 탐색질문을 포함한 전문데이터베이스에서의 탐색에 대한 피험자의 견해를 청취한다.

3.3 자료수집

본 연구의 수행에 필요한 자료는 다음과 같은 방법으로 수집하였다. 피험자 집단의 선정⁵⁾

4) McGraw-Hill Publication Online은 자체의 시소러스를 구축하거나 발행하지 않고 있기 때문에 과학기술분야에서 가장 널리 이용되고 있는 INSPEC Thesaurus를 제공한다.

5) 본 연구에 참가한 피험자의 소속기관은 아래의 10개 기관이다.

금성정보통신연구소, 삼성전자연구소, 삼성종합기술원, 통신개발연구원, 한국과학기술원, 한국에너지기술연구소, 한국원자력연구소, 한국전자통신연구소, 한국표준과학원, 한국화학연구소 (가나다순)

은 1995년 8월 12일에 완료되었고 자료의 수집 기간은 1995년 8월 16일부터 9월 2일까지였다. 관련 자료는 본 연구자가 해당 피험자 개개인을 직접 방문하여 수집하였다.

자료의 수집을 위한 질문지는 주제전문가용(부록 1)과 탐색전문가용(부록 2)의 두가지로 구분하였다. 질문지의 구성은 크게 다섯가지 1) 인구·통계적인 특성, 2) 탐색경험, 3) 탐색 질문의 주제에 대한 분석, 4) 탐색질문, 5) 탐색결과에 대한 만족도의 측정부분 등으로 이루어져 있다.

3.4 연구의 제한점

본 연구는 다음과 같은 제한점을 가지고 있다.

- ① 본 연구의 탐색시스템을 온라인 정보검색시스템인 DIALOG 시스템으로 제한하였다. DIALOG 시스템으로 제한한 이유는 국내에서 가장 이용율이 높은 점을 고려하였기 때문이며, 그리고 CD-ROM 시스템을 제외한 것은 전문데이터베이스의 특성상 CD-ROM의 매수가 많아 시간요인을 측정하기가 불가능하며, 탐색비용도 측정할 수 없기 때문이다.
- ② 본 연구의 탐색질문의 수가 적다. 탐색질문의 수를 적게 한 이유는 온라인으로 탐색을 수행하는 경우에 소요되는 탐색 시간 및 그에 따르는 탐색비용의 문제를 고려하였기 때문이다.

4. 주제전문가와 탐색전문가의 탐색특성 비교분석

전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가의 두 피험자 집단사이에서 나타나는 탐색특성에 관한 비교분석은 첫째, 탐색질문의 분석 둘째, 탐색과정 셋째, 탐색결과 등의 세가지 요인으로 구분하여 앞에서 제시한 제가설을 검증하기로 한다.

본 연구에서는 가설의 검증을 위하여 독립된 두개의 표본 집단의 평균 차이를 검증하는 분석방법인 T 검증(T-Test)을 사용하였으며, 자료의 분석에 필요한 관련 기본 데이터는 Microsoft Excel Version 5.0을 사용하여 입력하였고, 통계처리는 The SAS System for Windows를 사용하였다.

여기서 유의수준(p)은 <.05로 하기로 한다.

4.1 탐색질문의 분석에 관한 비교분석

탐색질문의 분석과 관련하여 주제전문가와 탐색전문가의 탐색특성의 차이에 관한 데이터의 수집은 질문지를 사용하였으며, 수집된 데이터는 리커트의 5점척도법을 적용하였다.

탐색질문의 분석에 관한 비교분석은 1) 탐색질문의 주제에 대한 이해도, 2) 탐색어의 선정에 대한 난이도, 3) 탐색결과에 대한 기대수준 등 세가지의 변인을 설정하여 수행하였다.

〈표 4-1〉 탐색질문의 분석에 관한 비교분석

탐색질의	주제전문가 (N=45)				탐색전문가 (N=45)				검 증 통계치
	평 균	표준편차	최대치	최소치	평 균	표준편차	최대치	최소치	
변인 1	3.89	0.61	5.00	3.00	2.82	0.83	4.00	1.00	t=6.92
변인 2a	3.33	0.48	4.00	3.00	2.64	0.88	4.00	1.00	t=4.61
변인 2b	3.60	0.65	5.00	2.00	2.82	0.89	5.00	1.00	t=4.74
변인 3	3.58	0.58	5.00	3.00	2.91	0.79	4.00	1.00	t=4.54

양방향 t 검증, $t \geq \pm 1.991$, 유의수준(p) < .05, 자유도(df)=88

변인 1 : 탐색질문의 주제에 대한 이해도 변인 2b : 핵심 탐색어의 선정에 대한 난이도
 변인 2a : 탐색어의 선정에 대한 난이도 변인 3 : 탐색결과에 대한 기대수준

4.1.1 탐색질문의 주제에 대한 이해도

탐색질문의 주제에 대한 이해도 수준은 표 4-1(변인1)과 같이 주제전문가는 5점척도를 기준으로 평균 3.89의 값을 기록하여 보통 이상의 이해도 수준을 보였고, 탐색전문가는 평균 2.82의 값을 기록하여 보통 미만의 이해도 수준을 보였다.

주제전문가가 탐색전문가 보다 탐색질문의 주제에 대한 이해도 수준이 상대적으로 높은 것은 "3.2.3 탐색집단의 선정"에서 언급한 바와 같이 주제전문가의 피험자 집단 대부분이 해당 탐색질문과 관련된 주제분야의 학력배경을 가지고 있고, 관련 분야의 연구를 수행하였거나 또는 진행하고 있으며, 학력수준이 탐색전문가 보다 높기 때문이라고 생각된다. 반면에, 탐색전문가의 피험자 집단은 66.7%(10명)가 전공이 문헌정보학이고 학력 수준도 주제전문가에 비하여 상대적으로 낮은 결과에 기인하는 것이라고 생각된다(〈표 3-2〉, 〈표 3-3〉 참조).

4.1.2 탐색어의 선정에 대한 난이도

탐색어의 선정에 대한 난이도는 첫째, 탐색질문에 대하여 실제 탐색시 사용할 모든 탐색어를 선정하는데 있어서의 어려움의 정도 둘째, 선정된 탐색어들 중에서 가장 핵심이 되는 한개의 탐색어를 선정하는데 있어서의 어려움의 정도를 파악하는 두가지로 구분하여 조사하였다.

첫째, 탐색질문에 대하여 실제 탐색시 사용할 모든 탐색어의 선정에 대한 난이도의 조사 결과는 〈표 4-1〉(변인2a)과 같이 주제전문가는 5점척도를 기준으로 평균 3.33의 값을 기록하여 보통 이상의 난이도를 보였고, 탐색전문가는 평균 2.64의 값을 기록하여 보통 미만의 난이도를 보였다.

둘째, 탐색질문의 주제와 관련하여 가장 핵심이 될 수 있는 한개의 탐색어 선정에 대한 난이도에 관한 조사결과는 〈표 4-1〉(변인2b)과 같이 주제전문가는 5점척도를 기준으로 평균 3.60의 값을 보였고, 탐색전문가는 평균

2.82의 값을 기록하였다. 핵심 탐색어 선정에 대한 난이도 역시 표 4-2의 탐색어 선정에 대한 난이도와 마찬가지로 세가지의 탐색질문 모두 탐색전문가가 주제전문가 보다 어려움의 정도가 약간 심하였다.

4.1.3 탐색결과에 대한 기대수준

탐색질문의 탐색결과에 대한 기대수준은 <표 4-1>(변인3)과 같이 주제전문가는 5점척도를 기준으로 평균 3.58의 값을 기록하여 보통 이상의 기대수준을 보였고, 탐색전문가는 평균 2.91의 값을 기록하여 보통 미만의 기대수준을 보였다. 이와 같이 주제전문가가 탐색전문가 보다 탐색질문의 탐색결과에 대한 기대수준이 약간 높은 것은 주제전문가가 탐색전문가 보다 탐색질문에 대한 이해도가 상대적으로 높고, 탐색어의 선정에 대해서도 전반적으로 어려움의 정도를 덜 느낀데서 기인할 것이다.

4.1.4 탐색질문의 분석에 관한 검증

주제전문가와 탐색전문가의 두 피험자 집단의 탐색질문의 주제에 대한 이해도 차이에 관한 비교분석은 <표 4-1>(변인1)과 같이 주제전문가는 평균 3.89의 값을 보였고, 탐색전문가는 평균 2.82의 값을 보였다. 두 피험자 집단의 탐색질문의 주제에 대한 이해도 값의 평균 차이에 대한 검증 통계치 t 값은 6.92를 보였다. 따라서 검증 통계치 t 값이 유의수준 <.05(이하 $p<.05$ 라 한다.), 자유도 88(이하 $df=88$ 이라 한다.)일 때의 t 값 1.991 보다 큰 것으로 나타나 통계적으로 유의한 차이를 보

이고 있다.

주제전문가와 탐색전문가의 두 피험자 집단의 실제 탐색시 사용할 탐색어 선정의 난이도 차이에 관한 비교분석은 두가지로 나누어 수행하였는데, 첫번째는 탐색질문에 대하여 실제 탐색에 사용할 모든 탐색어를 선정하여 모두 기재하게 한 다음, 이 과정에 대한 어려움의 정도를 파악하는 것이었고, 두번째는 선정된 모든 탐색어 중에서 핵심이 되는 한개의 탐색어를 선정하게 한 다음, 이 과정에 대한 어려움의 정도를 파악하는 것이었다.

전자의 차이에 대해서는 <표 4-1>(변인 2a)과 같이 주제전문가는 평균 3.33의 값을 보였고, 탐색전문가는 평균 2.64의 값을 보였다. 두 피험자 집단의 탐색질문에 대한 탐색어의 선정에 대한 난이도 값의 평균 차이에 대한 검증 통계치 t 값은 4.61로 나타났다. 따라서 $p<.05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

후자의 차이에 대해서는 <표 4-1>(변인2b)과 같이 주제전문가는 평균 3.60의 값을 보였고, 탐색전문가는 평균 2.82의 값을 나타냈다. 두 피험자 집단의 탐색질문에 대하여 선정된 탐색어 중에서 핵심이 되는 한개의 탐색어 선정에 대한 난이도 값의 평균 차이에 대한 검증 통계치 t 값은 4.74로 나타났다. 따라서 후자의 경우에도 $p<.05$, $df=88$ 일때 $t \geq \pm 1.991$ 의 결과를 보여 전자의 경우와 마찬가지로 통계적으로 유의한 차이를 보이고 있다.

주제전문가와 탐색전문가의 두 피험자 집단의 탐색질문의 탐색결과에 대한 기대수준의 차이에 관한 비교분석은 <표 4-1>(변인3)과

같이 주제전문가는 평균 3.58의 값을 보였고, 탐색전문가는 평균 2.91의 값을 나타냈다. 그리고 두 피험자 집단의 탐색질문의 탐색결과에 대한 기대수준 값의 평균 차이에 대한 검증 통계치 t값은 4.54로 나타났다. 따라서 $p < .05$, $df = 88$ 일때 $t \geq \pm 1.991$ 이므로 통계적으로 유의한 차이를 보이고 있다.

4.2 탐색과정에 관한 비교분석

탐색과정의 분석은 탐색질문에 대한 실제의 탐색이 완료된 후 시스템에 자동으로 기록된 탐색의 전과정에 대한 데이터를 근거로 하였다. 탐색과정의 분석은 1) 탐색어 수, 2) 연산

자 수(불 연산자, 제한 연산자), 3) 브라우징 문헌 수, 4) 탐색시간(접속시간, 출력문헌 건당 소요시간, 적합문헌 건당 소요시간) 등의 크게 네가지 요인으로 구분하여 각 탐색질문 별로 산정하여 종합하였다.

4.2.1 탐색어 수

탐색질문에 대하여 실제 탐색과정에서 사용한 탐색어의 수는 표 4-2(변인4)와 같이 주제전문가는 평균 6.80개를 사용하였고, 탐색전문가는 3.56개를 사용하였다. 따라서, 주제전문가가 탐색전문가 보다 평균 3.24개의 탐색어를 더 많이 사용한 것으로 나타났다.

〈표 4-2〉 탐색과정의 비교분석

탐색질의	주제전문가 (N=45)				탐색전문가 (N=45)				검 증 통계치
	평 균	표준편차	최대치	최소치	평 균	표준편차	최대치	최소치	
변인 4	6.80	1.77	12.00	3.00	3.56	1.27	7.00	1.00	t=10.00
변인 5a	0.60	0.13	0.80	0.20	1.11	0.36	2.20	0.50	t=-8.93
변인 5b	0.53	0.18	1.80	0.20	1.29	0.51	2.70	0.70	t=-9.36
변인 6	56.07	43.51	229.00	2.00	3.04	8.34	43.00	0.00	t= 8.03
변인 7a	1,341.40	477.72	2,426.00	645.00	909.91	391.54	2,191.00	282.00	t= 4.69
변인 7b	128.00	79.19	469.00	45.80	47.34	13.92	75.10	18.80	t= 6.73
변인 7c	221.08	138.27	684.00	74.80	115.78	74.51	497.50	38.40	t= 4.50

양방향 t 검증, $t \geq \pm 1.991$, 유의수준(p) < .05, 자유도(df) = 88

변인 4 : 탐색어 수

변인 5a : 탐색어 한개당 불 연산자 수

변인 5b : 탐색어 한개당 제한 연산자 수

변인 6 : 브라우징 문헌 수

변인 7a : 접속시간(초)

변인 7b : 출력문헌 건당 소요시간(초)

변인 7c : 적합문헌 건당 소요시간(초)

4.2.2 연산자 수

탐색질문에 대하여 실제 탐색과정에서 사용된 연산자는 불 연산자와 제한 연산자의 두가지로 구분하여 분석을 하였다. 한편, 피험자의 집단 구성이 단일 집단으로 되어 있는 경우, 탐색과정에서 사용된 탐색어의 수는 연산자의 사용 횟수에 영향을 끼쳐, 탐색어 수가 많으면 연산자의 사용 횟수가 증가된다는 것은 주지의 사실이며, 최근 우리나라의 연구자들도 밝히고 있다(장혜란 1991 : 사공미 1992).

본 연구의 주 목적이 두 피험자 집단의 탐색특성의 차이를 밝히는 것이기 때문에, 연산자 수는 탐색어 한개당 사용된 연산자의 수를 척도로 삼아 비교분석을 하기로 한다.

4.2.2.1 불 연산자 수

탐색과정에서 사용된 탐색어 한개당 불 연산자 수는 <표 4-2>(변인5a)와 같이 세가지의 탐색질문 모두 탐색전문가가 주제전문가 보다 높은 수치를 보이고 있다. 주제전문가는 평균 0.60개를 사용하였고, 탐색전문가는 평균 1.11개를 사용하였다. 따라서 탐색전문가가 주제전문가 보다 평균 0.51개를 더 사용한 것으로 나타났다.

4.2.2.2 제한 연산자 수

탐색과정에서 사용된 탐색어 한개당 제한 연산자 수는 <표 4-2>(변인5b)에 나타난 바와 같이 주제전문가는 평균 0.53개를 사용하였고, 탐색전문가는 1.29개를 사용하여 탐색전문가는 주제전문가 보다 평균 0.76개 더 많이 사용한 것으로 나타났다.

4.2.3 브라우징 문헌 수

탐색과정에서 사용된 브라우징 문헌 수는 표 4-2(변인6)와 같이 주제전문가와 탐색전문가 사이에 매우 큰 차이를 보이고 있다. 주제전문가는 평균 56.07건의 문헌을 브라우징하였으며, 이에 반하여 탐색전문가의 브라우징 문헌 수는 불과 3.04건에 지나지 않는 것으로 나타났다. 따라서 주제전문가가 탐색전문가에 비하여 월등히 높은 수치를 보이고 있다.

주제전문가는 세가지의 탐색질문 모두에 대하여 피험자 집단의 구성원 전부(45명/45명, 100.00%)가 브라우징을 하면서 탐색을 수행한 것에 반하여, 탐색전문가는 “탐색질문 1”에서 “탐색질문 2”에서 3명(3명/15명, 20.00%), 그리고 “탐색질문 3”에서 3명(3명/15명, 20.00%), 모두 8명(8명/45명, 17.78%)만이 브라우징을 하면서 탐색을 수행한 것으로 나타났다(부록 1, 2 참조).

이와 같이 탐색과정에서 나타난 브라우징 문헌 수에 관한 분석결과는 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가의 두 피험자 집단의 탐색특성의 차이를 극명하게 보여주고 있다.

4.2.4 탐색시간

탐색과정의 변인으로 설정한 탐색시간은 1) 접속시간, 2) 출력문헌 건당 소요시간, 3) 적합문헌 건당 소요시간 등의 세가지로 구분하여 분석을 행하였다.

4.2.4.1 접속시간

접속시간은 <표 4-2>(변인 7a)와 같이 주제

전문가는 평균 1,341.40초를 사용하였고, 탐색전문가는 평균 909.91초를 사용하였다. 따라서, 접속시간은 주제전문가가 탐색전문가 보다 평균 431.49초를 더 사용하여 두 집단 사이에 큰 차이를 보이고 있다.

4.2.4.2 출력문헌 건당 소요시간

탐색결과에 대한 최종 출력문헌 건당 소요시간은 <표 4-2>(변인 7b)와 같이 주제전문가는 평균 128.00초를 사용하였고, 탐색전문가는 평균 47.34초를 사용하여, 주제전문가가 탐색전문가 보다 평균 80.66초를 더 사용한 것으로 나타났다.

그러나 앞에서 언급한 바와 같이 탐색결과에 대한 출력문헌 수는 부록 1, 부록 2에 나타난 바와 같이 주제전문가가 평균 12.16건이고, 탐색전문가는 19.11건으로 나타났다. 탐색전문가는 주제전문가에 비하여 탐색결과에 대한 출력문헌 수는 6.95건이 더 많지만, 오히려 탐색전문가의 출력문헌 건당 소요시간은 주제전문가의 평균 128.00초의 36.98%에 불과한 평균 47.34초에 지나지 않는 것으로 나타났다. 이와 같은 결과는 주제전문가는 탐색전문가에 비하여 탐색과정에서 브라우징을 하면서 대부분의 시간을 소비하기 때문이라고 본다.

4.2.4.3 적합문헌 건당 소요시간

적합문헌 건당 소요시간은 <표 4-2>(변인 7c)에 나타난 바와 같이 주제전문가는 평균 221.08초를 사용하였고, 탐색전문가는 평균 115.78초를 사용하였다. 적합문헌 건당 소요시간은 주제전문가가 탐색전문가 보다 평균

105.30초를 더 많이 사용한 것으로 나타났다.

따라서, 탐색과정의 변인으로 설정한 탐색시간(접속시간, 출력문헌 건당 소요시간 및 적합문헌 건당 소요시간)은 주제전문가가 탐색전문가 보다 더 많이 사용한 것으로 나타났다.

4.2.5 탐색과정에 관한 검증

실제의 탐색과정에서 사용한 탐색어 수의 차이에 관한 비교분석은 <표 4-2>(변인 4)와 같이 주제전문가는 평균 6.80개의 탐색어를 사용하였고, 탐색전문가는 평균 3.56개의 탐색어를 사용한 것으로 나타났다. 그리고 두 피험자 집단의 탐색어 수의 평균 차이에 대한 검증 통계치 t 값은 10.00을 나타냈다. 따라서 $p < .05$, $df = 88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

실제의 탐색과정에서 사용된 연산자 수는 탐색어 한개당 사용된 연산자 수를 기준으로 하여 계산하였다. 그리고 연산자는 불 연산자와 제한 연산자로 구분하여 비교분석을 행하였다(4.2.2 참조).

먼저, 불 연산자 수의 차이에 관한 비교분석은 <표 4-2>(변인 5a)와 같이 주제전문가는 평균 0.60개를 사용하였으며, 탐색전문가는 평균 1.11개를 사용한 것으로 나타났다. 그리고 두 피험자 집단의 불 연산자 수의 평균 차이에 대한 검증 통계치 t 값은 -8.93을 나타냈다. 따라서 $p < .05$, $df = 88$ 일때 $t \geq \pm 1.991$ 이 성립되어 통계적으로 유의한 차이를 보이고 있다.

다음으로, 제한 연산자 수의 차이에 관한 비교분석은 <표 4-2>(변인 5b)와 같이 주제전문가는 평균 0.53개의 제한 연산자를 사용하였

으며, 탐색전문가는 평균 1.29개의 제한연산자를 사용하였다. 두 피험자 집단의 제한연산자 수의 평균 차이에 대한 검증 통계치 t 값은 -9.36으로 나타났다. 따라서 $p < .05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

실제의 탐색과정에서 사용한 브라우징 문헌 수의 차이에 관한 비교분석은 <표 4-2>(변인 6)와 같이 주제전문가는 평균 56.07건의 문헌을 브라우징하였고, 탐색전문가는 평균 3.04건의 문헌을 브라우징하였다. 두 피험자 집단의 브라우징 문헌 수의 평균 차이에 대한 검증 통계치 t 값은 8.03으로 나타났으므로, $p < .05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립하여 통계적으로 유의한 차이를 보이고 있다.

탐색과정의 변인으로 설정한 탐색시간의 차이에 관한 비교분석은 첫째, 접속시간 둘째, 출력문헌 건당 소요시간 셋째, 적합문헌 건당 소요시간 등의 세가지로 구분하여 비교분석을 하였다.

첫째, 접속시간의 차이에 관한 비교분석은 <표 4-2>(변인 7a)와 같이 주제전문가는 평균 1341.40초의 시간을 사용하였으며, 탐색전문가는 평균 909.91초의 시간을 사용한 것으로 나타났다. 두 피험자 집단 사이의 접속시간의 평균 차이에 대한 검증 통계치 t 값은 4.69를 나타냈다. 따라서 $p < .05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

둘째, 출력문헌 건당 소요시간의 차이에 관한 비교분석은 <표 4-2>(변인 7b)와 같이 주제전문가는 평균 128.00초, 탐색전문가는 평균

47.34초를 사용하였다. 그리고 두 피험자 집단 사이의 출력문헌 건당 소요시간의 평균 차이에 대한 검증 통계치 t 값은 6.73을 나타내어, $p < .05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

셋째, 적합문헌 건당 소요시간의 차이에 관한 비교분석은 <표 4-2>(변인 7c)와 같이 주제전문가는 평균 221.08초를 사용하였고, 탐색전문가는 평균 115.78초를 사용하였다. 그리고 두 피험자 집단 사이의 적합문헌 건당 소요시간의 평균 차이에 대한 검증 통계치 t 값은 4.50로 나타났다. 따라서, $p < .05$, $df=88$ 일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

4.3 탐색결과에 관한 비교분석

탐색결과에 관한 비교분석은 탐색과정에 관한 비교분석과 마찬가지로 탐색질문에 대한 실제의 탐색이 완료된 후 시스템에 자동으로 기록된 데이터를 다운로드하여 프린터로 출력한 자료를 근거로 하였다.

탐색결과와 관련된 변인은 크게 세가지 요인으로 구분하여 각 질문별로 산정하여 종합하였다. 세가지 요인은 첫째, 탐색효율(적합문헌 수, 재현율, 정확률) 둘째, 탐색비용(총 탐색비용, 출력문헌 건당 소요비용, 적합문헌 건당 소요비용) 셋째, 탐색결과에 대한 만족도 등이다.

4.3.1 탐색효율

탐색결과에 관한 변인으로 설정한 탐색효율은

〈표 4-3〉 탐색결과와 비교분석

탐색질의	주제전문가 (N=45)				탐색전문가 (N=45)				검 증 통계치
	평 균	표준편차	최대치	최소치	평 균	표준편차	최대치	최소치	
변인 8a	7.29	2.57	12.00	1.00	8.58	2.07	12.00	2.00	t=-2.62
변인 8b	64.26	20.93	92.30	11.10	75.43	13.90	92.30	22.20	t=-2.98
변인 8c	62.54	19.07	100.00	22.50	47.73	15.12	84.60	13.30	t= 4.08
변인 9a	34,402	12,761	85,297	13,572	28,558	7,496	64,814	19,132	t= 2.65
변인 9b	3,205	1,735	9,113	1,460	1,543	262	2,256	983	t= 6.35
변인 9c	5,574	3,553	21,326	2,537	3,602	1,686	12,369	1,913	t= 3.36
변인 10	3.47	0.76	5.00	2.00	2.93	0.69	4.00	2.00	t= 3.50

양방향 t 검증, $t \geq \pm 1.991$, 유의수준(p) < .05, 자유도(df) = 88

변인 8a : 적합문헌 수	변인 9b : 출력문헌 건당 소요비용(원)
변인 8b : 재현율(%)	변인 9c : 적합문헌 건당 소요비용(원)
변인 8c : 정확률(%)	변인 10 : 탐색결과에 대한 만족도
변인 9a : 총 탐색비용(원)	

1) 적합문헌 수, 2) 재현율, 3) 정확률 등의 세 가지로 구분하여 분석을 행하였다.

4.3.1.1 적합문헌 수

탐색결과에 대한 최종 출력문헌 수는 부록 1, 2에 나타난 바와 같이 주제전문가는 평균 12.16건을 출력하였고, 탐색전문가는 평균 19.11건을 출력한 것으로 나타나, 탐색전문가가 주제전문가 보다 평균 6.95건을 더 출력하였다.

탐색결과에 대한 최종 출력문헌 수 중 적합문헌 수는 앞에서 언급한 탐색결과에 대한 출력문헌 수의 경우와 마찬가지로 〈표 4-3〉(변인 8a)과 같이 탐색전문가가 주제전문가 보다 약간 높은 수치를 보이고 있다. 주제전문가는

평균 7.29건의 적합문헌을 검색하였고, 탐색전문가는 평균 8.58건의 적합문헌을 검색하여, 탐색전문가가 주제전문가 보다 평균 1.29건이 많은 적합문헌을 검색한 것으로 나타났다.

4.3.1.2 재현율

재현율은 〈표 4-3〉(변인 8b)과 같이 주제전문가는 평균 64.26%의 재현율을 보였고, 탐색전문가는 평균 75.43%의 재현율을 보였다. 재현율은 탐색전문가가 주제전문가 보다 평균 11.17%가 높은 것으로 나타났다.

4.3.1.3 정확률

정확률은 〈표 4-3〉(변인 8c)과 같이 재현율의 경우와는 반대의 현상을 보였는데, 주제전문

문가는 평균 62.54%의 정확률을 보였고, 탐색전문가는 47.73%의 정확률을 보였다. 정확률은 주제전문가가 탐색전문가 보다 평균 14.81%가 높은 것으로 나타냈다.

일반적으로 전문데이터베이스의 탐색효율에 있어서 재현율은 높은 반면에 정확률은 떨어지게 된다는 것이 정설로 되어 있다. 본 연구의 결과에서도 주제전문가와 탐색전문가의 두 집단 모두 재현율이 정확률 보다 높은 것으로 나타났다.

그러나 탐색전문가의 경우에는 재현율과 정확률의 비율이 평균 75.43% : 평균 47.73%로 평균 27.70%의 뚜렷한 차이가 나타나지만, 주제전문가의 경우에는 그 비율이 평균 64.26% : 평균 62.54%로 그 차이가 평균 1.72%에 불과한 것으로 나타났다.

이와 같은 분석결과는 피험자인 주제전문가의 집단 구성원이 대부분 탐색질문의 주제분야와 관련된 전공 배경을 가지고 있고 아울러 관련분야의 연구를 진행하고 있거나 수행한 결과에서 기인하기도 하겠지만, 더 큰 요인은 주제전문가의 피험자 집단 구성원 모두가 브라우징을 하면서 탐색을 수행하였기 때문에 주제전문가의 재현율과 정확률 사이에는 큰 차이가 나타나지 않는다고 볼 수 있다.

4.3.2 탐색비용

탐색결과의 변인으로 설정한 탐색비용⁶⁾은 1) 총 탐색비용, 2) 출력문헌 건당 소요비용, 3) 적합문헌 건당 소요비용 등의 세가지로 구분하여 분석을 행하였다. 탐색비용에 대한 분석은 이해의 편의를 위하여 원 단위까지만 표

시하고 소숫점 이하의 단위는 생략하였다.

4.3.2.1 총 탐색비용

총 탐색비용은 <표 4-3>(변인 9a)과 같이 주제전문가는 평균 34,402원을 사용하였고, 탐색전문가는 평균 28,558원을 사용하였다. 총 탐색비용은 주제전문가가 탐색전문가 보다 평균 5,844원을 더 사용한 것으로 나타났다.

4.3.2.2 출력문헌 건당 소요비용

탐색결과에 대한 최종 출력문헌 건당 소요비용은 <표 4-3>(변인 9b)과 같이 주제전문가와 탐색전문가의 두 집단 사이에 큰 차이를 보이고 있다. 주제전문가는 출력문헌 건당 평균 3,205원을 사용하였고, 탐색전문가는 평균 1,543원을 사용하였다. 탐색결과에 대한 최종 출력문헌 건당 소요비용은 주제전문가가 탐색전문가 보다 평균 1,662원을 더 많이 사용한 것으로 나타났다.

여기에서, 주제전문가와 탐색전문가의 두 피험자 집단 사이의 탐색특성의 차이는 출력문헌 건당 소요비용과 출력문헌 수를 대비하여 보면 더욱 더 명확하게 나타난다(<표 4-3>(변인 9b), [부록 1, 2] 참조).

탐색결과에 대한 최종 출력문헌 수는 주제전문가가 평균 12.16건, 탐색전문가는 19.11건

6) 탐색비용(총 탐색비용, 출력문헌 건당 소요비용, 적합문헌 건당 소요비용)은 실제의 탐색결과 데이터에는 US로 표시되어 있다. 본 연구소에서는 1995년 9월 19일자 전신환 매도율(US 100 : ₩78090)을 적용하여 원화로 환산하여 처리하였다.

으로 나타났다. 최종 출력문헌 수는 탐색전문가가 주제전문가 보다 평균 6.95건이 더 많다. 그러나 출력문헌 건당 소요비용은 주제전문가가 평균 3,205원, 탐색전문가는 평균 1,543원으로 주제전문가가 탐색전문가보다 평균 1,662원을 더 사용하였다. 탐색전문가는 주제전문가보다 최종 출력문헌 수가 많았음에도 불구하고, 출력문헌 건당 소요비용은 주제전문가보다 더 적은 것으로 나타났다. 이와 같은 분석 결과는 주제전문가와 탐색전문가의 두 피험자 집단 사이의 탐색특성의 차이를 뚜렷하게 보여주는 것이다.

4.3.2.3 적합문헌 건당 소요비용

탐색결과에 대한 최종 출력문헌 중 적합문헌 건당 소요비용은 <표 4-3>(변인 9c)과 같이 주제전문가는 평균 5,574원을 사용하였고, 탐색전문가는 평균 3,602원을 사용하여 주제전문가가 탐색전문가 보다 평균 1,972원을 더 사용한 것으로 나타났다. 따라서 비용대 효과 측면에서 보면 탐색전문가가 주제전문가 보다 경제적인 탐색을 수행한 것으로 나타났다.

여기에서 국내에서 소장하고 있지 않은 외국의 정기간행물 기사의 원문을 입수하는데 소요되는 시간과 비용을 살펴보면, 우리나라 대부분의 연구기관에서 가장 많이 이용하고 있는 것으로 나타난 DACOM을 통한 온라인 주문(Online Dialorder)을 이용하는 경우, 입수 소요기간은 평균 약 3주가 걸렸고(이용봉 1991, 23), 입수비용은 원문 기사 한 건당 평균 약 9,400원(US 1200)이 소요되는 것으로 나타났다(DACOM 1995). 물론, 온라인 전문

데이터베이스가 아직까지는 대부분 화상정보를 제외한 문자와 수치정보만을 제공하고 있다는 제한점은 있지만, 시간과 비용적인 측면을 고려해 볼 때 충분한 경제성을 가질 수 있는 것으로 생각된다.

4.3.3 탐색결과에 대한 만족도

탐색결과에 대한 만족도는 <표 4-3>(변인 10)과 같이 주제전문가는 5점척도를 기준으로 평균 3.47의 값을 나타내어 보통 이상의 만족도 수준을 보였고, 탐색전문가는 평균 2.93의 값을 나타내어 보통 미만의 만족도 수준을 보였다. 온라인 데이터베이스의 탐색에 있어서 탐색질문에 대한 이해도는 만족도에 영향을 미치는 요인임이 이미 밝혀져 있다(이용봉 1991). 앞에서 언급하였지만 탐색질문에 대한 이해도는 주제전문가가 탐색전문가 보다 높은 것으로 나타났다.

4.3.4 탐색결과에 관한 검증

탐색결과와 변인으로 설정한 탐색효율의 차이에 관한 비교분석은 첫째, 적합문헌 수 둘째, 재현율 셋째, 정확률의 세가지로 구분하여 비교분석을 하였다.

첫째, 주제전문가와 탐색전문가의 탐색결과에 대한 출력문헌 중 적합문헌 수의 차이에 관한 비교분석은 <표 4-3>(변인 8a)과 같이 주제전문가는 평균 7.29건의 적합문헌을 검색하였고, 탐색전문가는 8.58건의 적합문헌을 검색한 것으로 나타났으며, 두 피험자 집단 사이의 적합문헌 수의 평균 차이에 대한 검증 통계치 t 값은 -2.62를 나타냈다. 따라서 $p < .05$,

df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

둘째, 재현율의 차이에 대한 비교분석은 <표 4-3>(변인 8b)과 같이 주제전문가는 평균 64.26%의 재현율을 기록하였고, 탐색전문가는 평균 75.43%의 재현율을 기록하였다. 그리고 두 피험자 집단 사이의 재현율의 평균 차이에 대한 검증 통계치 t값은 -2.98을 나타냈다. 따라서 $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

셋째, 정확률의 차이에 관한 비교분석은 <표 4-3>(변인 8c)과 같이 주제전문가는 평균 62.54%의 정확률을 보였고, 탐색전문가는 평균 47.73%의 정확률을 보였다. 그리고 두 피험자 집단 사이의 정확률의 평균 차이에 대한 검증 통계치 t값은 4.08을 보였다. 따라서 $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

탐색결과와 변인으로 설정한 탐색비용의 차이에 관한 비교분석은 첫째, 총 탐색비용 둘째, 출력문헌 건당 소요비용 셋째, 적합문헌 건당 소요비용 등의 세가지로 구분하여 비교분석을 행하였다.

첫째, 총 탐색비용의 차이에 관한 비교분석은 <표 4-3>(변인 9a)과 같이 주제전문가는 평균 34,402원을 사용하였고, 탐색전문가는 평균 28,558원을 사용하였다. 두 피험자 집단 사이의 총 탐색비용의 평균 차이에 대한 검증 통계치 t값은 2.65로 나타났다. 따라서, $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

둘째, 출력문헌 건당 소요비용의 차이에 관

한 비교분석은 <표 4-3>(변인 9b)과 같이 주제전문가는 평균 3,205원을 사용하였고, 탐색전문가는 평균 1,543원을 사용하였다. 두 피험자 집단 사이의 출력문헌 건당 소요비용의 평균 차이에 대한 검증 통계치 t값은 6.35로 나타났다. 따라서, $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

셋째, 적합문헌 건당 소요비용의 차이에 관한 비교분석은 <표 4-3>(변인 9c)과 같이 주제전문가는 평균 5,574원을 사용하였고, 탐색전문가는 평균 3,602원을 사용하였다. 두 피험자 집단 사이의 적합문헌 건당 소요비용의 평균 차이에 대한 검증 통계치 t값은 3.36으로 나타났다. 따라서, $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

탐색결과에 대한 만족도의 차이에 관한 비교분석은 <표 4-3>(변인 10)과 같이 주제전문가는 평균 3.47의 값을 나타냈고, 탐색전문가는 평균 2.93의 값을 나타냈다. 그리고 두 피험자 집단 사이의 탐색결과에 대한 만족도의 평균 차이에 관한 검증 통계치 t값은 3.50으로 나타났다. 따라서, $p < .05$, df=88일때 $t \geq \pm 1.991$ 이 성립되므로 통계적으로 유의한 차이를 보이고 있다.

5. 결론 및 제언

본 연구는 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가는 탐색질문의 분

석, 탐색과정 및 탐색결과에 어떠한 특징적인 차이가 있는지를 밝히려 하였다. 이 연구를 통하여 얻어진 결론은 다음과 같다.

- (1) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 탐색질문의 주제에 대한 이해도의 수준이 더 높다.
- (2) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 탐색질문의 주제에 대하여 실제 탐색시 사용할 모든 탐색어를 선정하는데 있어서는 물론 핵심 탐색어를 선정하는데 있어서도 어려움의 정도를 덜 느낀다.
- (3) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 탐색질문의 탐색결과에 대한 기대수준이 더 높다.
- (4) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 실제의 탐색과정에서 사용하는 탐색어의 수가 더 많다.
- (5) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 실제의 탐색과정에서 탐색어 한개당 불 연산자 및 제한 연산자를 더 적게 사용한다.
- (6) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 실제의 탐색과정에서 브라우징 문헌 수가 더 많다.
- (7) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 실제의 탐색과정에서 접속시간, 출력문헌 건당 소요시간 및 적합문헌 건당 소요시간을 더 많이 사용한다.

(8) 전문데이터베이스의 탐색에 있어서 탐색질문에 대한 적합문헌의 검색은 탐색전문가가 주제전문가 보다 더 많으며, 재현율은 탐색전문가가 주제전문가 보다 더 높지만, 정확률은 주제전문가가 탐색전문가 보다 더 높다.

(9) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 총 탐색비용, 출력문헌 건당 소요비용, 적합문헌 건당 소요비용을 더 많이 사용한다.

(10) 전문데이터베이스의 탐색에 있어서 주제전문가는 탐색전문가 보다 최종 탐색결과에 대한 만족도가 더 높다.

이상과 같이 전문데이터베이스의 탐색에 있어서 주제전문가와 탐색전문가의 두 피험자 집단 사이에는 탐색질문의 분석, 탐색과정 및 탐색결과에 대하여 뚜렷한 탐색특성의 차이가 있다는 것이 밝혀졌다. 따라서 전문데이터베이스 시스템의 설계자들은 시스템의 설계나 수정시 이러한 탐색특성의 차이를 최대한 반영함으로써 전문데이터베이스 시스템의 효율을 높일 수 있을 것이며, 전문데이터베이스의 이용자들 역시 기존의 정형화된 탐색방식을 탈피하여 본 연구에서 밝혀진 여러 요인들을 포함한 다양한 탐색기법을 활용한다면, 최대한의 효과를 이룰 수 있을 것이다.

아울러, 전문데이터베이스가 그 수에 있어서는 물론 이용율 또한 급증하고 있는 현실을 감안해 보면 다음의 두 가지 측면의 연구가 이루어질 필요가 있다고 본다. 첫째, 이용자 집단의 계층을 세분화하여 특정의 전문데이터베이스에 국한하지 않은 복수의 전문데이터베이스에

이스를 대상으로 하고, 다단계의 난이도로 구성된 다수의 탐색질문에 대한 이용자의 탐색 특성에 관한 광범위한 연구가 이루어지고 둘째, 본 연구에서 사용한 DIALOG 외에 다른 온라인 시스템 뿐만 아니라, CD-ROM판 전문데이터베이스의 탐색특성에 관한 연구 등이 이루어지면 전문데이터베이스의 탐색특성분석을 통한 정보이용의 효율성 제고에 도움이 될 것이다.

참고문헌

Lancaster, F. Wilfrid. 윤구호, 김태승 공역.
1985. 정보검색시스템. 서울: 구미무역.
남영준, 1987. 전문데이터베이스의 탐색효율성 분석. 석사학위논문, 중앙대학교
사공미, 1992. 이용자의 경험유무에 따른 정보 탐색행태의 비교연구. 석사학위논문, 숙명여자대학교
이응봉, 1991. 데이터베이스의 온라인 검색봉사에 대한 이용자의 만족도 연구. 석사학위논문, 성균관대학교
장혜란, 1991. 온라인 데이터베이스 탐색자의 탐색행태에 관한 연구. 박사학위논문, 연세대학교
Beard, Joseph J. 1971. "Information Systems Application in Law." *Annual Review of Information Science and Technology* 6: 369-396.
BiblioData, 1995. *Fulltext Sources Online for Periodicals, Newspapers, Newsletters,*

Newswires and TV/Radio Transcripts. Needham Heights, MA.

Bookstein, Abraham. 1992. "A Systematic Approach to Compressing a Full-Text Retrieval System." *Information Processing & Management* 28(6): 795-806.
Bull, Gillian. 1981. "A Brief Survey of Developments in Computerised Legal Information Retrieval." *Program* 15(3): 109-119.
Butler, Brett. 1985. "Economics of Full-Text Information". In: *National Online Meeting Proceedings, 1985.* Medford : Learned Information, Inc., pp. 79-84.
Collen, Morris F. and Flagle Charles D. 1985. "Full-Text Medical Literature Retrieval by Computer." *Journal of American Medical Association* 15: 2768-2774.
DIALOG Information Services, Inc. 1995. *McGraw-Hill Publications Online:* 624. Palo Alto.
DIALOG Information Services, Inc. 1994. *McGraw-Hill Publications Online: Documentation Chapter.* Palo Alto.
Fenichel, Carol H. 1981. "Online Searching: Measures that Discriminate among Users with Different Type of Experiences." *Journal of the American Society for Information Science*

- 32(1): 23-32.
- Fidel, R. and Soergel, D. 1983. "Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research." *Journal of the American Society for Information Science* 34(3): 163-180.
- Franklin, J. and Buckingham, M. C. S. et al. 1983. "Biomedical Journals in an Online Full Text Database: A Review of Reaction to ESPL". In: *7th International Online Meeting* pp. 6-8.
- Garson, L. R. and Cohen, S. M. 1983. *User's Manual: Primary Journal Database ACS Full-Text File*. Washington DC : American Chemical Society.
- Gauch, S. A. 1990. *An Expert System for Searching in Full-Text*. Doctoral Dissertation, The University of North Carolina at Chapel Hill.
- Hearst, M. A. 1994. *Context and Structure in Automated Full-Text Information Access*. Doctoral Dissertation, University of California, Berkeley.
- Hjerpe, Roland. 1986. "Electronic Publishing : Writing Machines and Machine Writing." *Annual Review of Information Science and Technology* 21: 123-166.
- Jung-Soon, Ro. 1985. *An Evaluation of the Capability of Ranking Algorithms to Improving the Effectiveness of Full Text Retrieval*. Doctoral Dissertation, Indiana University.
- Kainz, V. L. M. 1988. *A Case Study of Organizational Evaluation of the Effectiveness of Full Text Information Retrieval Systems : A Macro and Micro Perspective*. Doctoral Dissertation, The American University.
- Lancaster, F. Wilfrid. 1987. *Information Retrieval Systems: Characteristics, Testing and Evaluation*. New York : John Wiley & Sons.
- Larson, Signe E. and Williams, Martha E. 1980. "Computer Assisted Legal Research." *Annual Review of Information Science and Technology* 15: 251-286.
- Liebscher, Peter. 1993. *Information-Seeking in Hypertext : Multiple Access Methods in a Full-Text Hypertext Database*. Doctoral Dissertation, University of Maryland College Park.
- Love, Richard A. 1985. "Precision in Searching the Full Text Database / ACS Journal Online." In: *National Online Meeting Proceedings, 1985*, Medford : Learned Information, Inc., pp. 273-282.
- Madelung, Hans-ole. 1981. "Subject Indexing

- in the Social Science: a Comparison of PRECIS and KWIC Indexes to Newspaper Article." *Journal of Librarianship* 14(1): 45-48.
- McKinin, Emma Jean et al. 1991. "The Medline/Full-Text Research Project." *Journal of the American Society for Information Science* 42(4): 297-307.
- Marchionini, Gary et al. 1993. "Information Seeking in Full Text End-User-Oriented Search Systems: The Roles of Domain and Search Expertise." *LISR* 15: 35-69.
- Martin, T. Patrick. 1990. "A Case Study of Caching Strategies for a Distributed Full Text Retrieval System." *Information Processing & Management* 26(2): 227-247.
- Molto, Mavis. 1993. "Improving Full Text Search Performance through Textual Analysis." *Information Processing & Management* 29(5): 615-632.
- Sung Been, Moon. 1993. *Enhancing Performance of Full-Text Retrieval Systems Using Relevance Feedback*. Doctoral Dissertation, The University of North Carolina at Chapel Hill.
- Pagell, R. A. 1987. "Searching Full-Text Periodicals: How Full is Full?" *Database* 10: 33-38.
- _____. 1987. "Searching IAC's Full-Text Files: It's Awfully Confusing." *Database* 10: 39-47.
- _____. 1988. "Primary Full-Text Information: Databases for the End-User." *Online Information* 88. *Proceedings of the 12th International Online Information Meeting, 6-8 December 1988, Proceedings* 1: 255-262.
- Penhale, Sara. J. and Taylor, Nancy. 1986. "Integrating End-User Searching into a Bibliographic Instruction Program." *RQ* 27: 212-220.
- Sormunen, E. 1993. "Free-Text Searching in Full-Text Databases: Probing System Limits." *Proceedings of the 17th International Online Information Meeting, 93 pp.* 433-442.
- Sullivan, M. V., Borgman, C. L. and Wipperrn, D. 1990. "End-Users, Mediated Searchers, and Front-End Assistance Programs on Dialog: A Comparison of Learning, Performance, and Satisfaction." *Journal of the American Society for Information Science* 41(4): 27-42.
- Tenopir, Carol. & Jung Soon, Ro. 1990. *Full Text Databases*. New York : Greenwood Press.
- Tenopir, Carol. 1984. "Full-Text Databases." *Annual Review of Information Science and Technology* 19: 216-246.
- _____. 1985. "Full Text Database Retrieval Performance." *Online Review* 9(2):

149-164.

_____. 1984. *Retrieval Performance in a Full Text Journal Article Database*, Doctoral Dissertation, University of Illinois.

_____. 1988. "Users and Uses of Full Text Databases." *Online Information 88. Proceedings of the 12th International Online Information Meeting, 6-8 December 1988, Proceedings, 1: 263-270.*

Terrant, Seldon W. 1980. "Computers in Publishing." *Annual Review of Information Science and Technology* 15: 191-219.

Walker, G. 1988. "The Search Performance of End-User." In: *National Online Meeting Proceedings, 1988* Medford : Learned Information, Inc. pp. 403-410.

Winkler, K. J. 1983. "Chemical Society Offers Full Texts of Scholarly Journals by Computer." *The Chronicle of Higher Education* pp. 23-24.

[부 록 1] 주제전문가의 탐색질문의 분석, 탐색과정, 탐색결과의 기본 통계표

Q	탐 색 질 문 1															평균	소 계	탐 색 질									
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15			A16	A17	A18	A19	A20	A21	A22	A23	A24	
V1	4	3	3	4	3	3	5	3	4	3	4	4	4	4	3	3.6	54	4	4	3	4	5	4	4	3	4	
V2	3	3	3	3	3	3	4	3	3	3	3	4	4	4	3	3.3	49	3	3	3	3	4	4	4	3	3	
V3	3	5	5	5	3	4	3	4	4	3	4	4	4	4	3	3.9	58	3	4	3	4	4	4	4	3	3	
V4	3	3	4	3	3	3	4	3	3	4	3	4	4	3	4	3.4	51	3	3	3	5	4	4	4	4	4	
V5	9	8	7	8	5	6	5	7	7	7	9	7	7	8	7	7.1	107	5	5	7	12	7	6	5	9	6	
V6	4	5	4	5	3	4	3	4	4	4	5	4	4	5	4	4.1	62	2	3	4	2	3	5	3	6	4	
V7	0.4	0.6	0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	8.9	0.4	0.6	0.6	0.2	0.4	0.8	0.6	0.7	0.7	
V8	3	4	5	3	3	2	4	4	4	5	4	3	5	4	4	3.8	57	2	2	2	7	6	6	2	3	3	
V9	0.3	0.5	0.7	0.4	0.6	0.3	0.8	0.6	0.6	0.7	0.4	0.4	0.7	0.5	0.6	0.5	8.1	0.4	0.4	0.3	0.6	0.9	1.0	0.4	0.3	0.5	
V10	30	40	25	40	31	54	51	56	27	40	36	16	31	26	25	35.2	528	70	14	30	197	2	86	32	50	45	
V11	1747	1255	1354	1326	1786	1345	1876	2024	1931	2234	2426	1924	1780	1824	1620	1766.2	26493	701	684	1384	2252	938	1568	810	965	876	
V12	102.8	125.5	104.2	63.1	127.6	96.1	144.3	156.7	160.9	139.6	161.7	160.3	178.0	165.8	162.0	136.8	2051.7	100.1	342.0	115.3	204.7	469.0	104.5	90.0	80.4	67.4	
V13	194.1	216.0	169.3	147.3	223.3	168.1	469.0	253.0	214.6	446.8	303.3	274.9	222.5	182.4	180.0	244.3	3664.6	116.8	684.0	197.7	563.0	469.0	196.0	405.0	137.9	146	
V14	17	10	13	21	14	14	13	13	12	16	15	12	10	11	10	13.4	201	7	2	12	11	2	15	9	12	13	
V15	9	6	8	9	8	8	4	8	9	5	8	7	8	10	9	7.7	116	6	1	7	4	2	8	2	7	6	
V16	75.0	50.0	66.7	75.0	66.7	66.7	33.3	66.7	75.0	41.7	66.7	58.3	66.7	83.3	75.0	64.5	966.8	66.7	11.1	77.8	44.4	22.2	88.9	22.2	77.8	66.7	
V17	52.9	60.0	61.5	42.9	57.1	57.1	30.8	61.5	75.0	31.3	53.3	58.3	80.0	90.9	90.0	60.2	902.6	85.7	50.0	58.3	36.4	100.0	53.3	22.2	58.3	46.2	
V18	37.93	28.23	34.50	39.20	48.91	36.96	45.66	53.96	48.96	50.82	59.76	40.96	35.65	44.68	42.10	43.22	648.34	21.98	17.38	30.56	109.23	19.96	56.43	25.79	48.54	48.75	
V19	2.23	2.82	2.65	1.87	3.49	2.64	3.51	4.15	4.06	3.18	3.98	3.41	3.57	4.06	4.21	3.32	49.85	3.14	8.69	2.55	9.93	9.98	3.76	2.87	4.06	3.75	
V20	4.21	4.7	4.31	4.36	6.11	4.62	11.42	6.75	5.44	10.16	7.47	5.85	4.46	4.47	4.68	5.93	89.02	3.66	17.38	4.37	27.31	9.98	7.06	12.9	6.93	8.13	
V21	4	3	4	3	4	3	5	3	3	4	3	4	4	4	3	3.6	54	4	3	3	4	4	4	4	3	4	

주) Q - 탐색질문
 T - 피험자(주제전문가)
 V - 변인
 V1 - 탐색질문에 대한 이해도
 V2 - 탐색어 선정의 난이도
 V3 - 핵심 탐색어 선정의 난이도

V4 - 탐색결과에 대한 기대수준
 V5 - 탐색어 수
 V6 - 불 연산자 수
 V7 - 탐색어 한개당 불 연산자 수
 V8 - 제한 연산자 수
 V9 - 탐색어 한개당 제한 연산자 수

[부록 1] continued

문 2							탐 색 질 문 3															평균 소 계			평균 소 계			총평균	합 계
A25	A26	A27	A28	A29	A30		A31	A32	A33	A34	A35	A36	A37	A38	A39	A40	A41	A42	A43	A44	A45								
4	4	4	4	4	5	4.0	60	4	4	4	5	4	3	3	4	4	4	4	4	5	4	5	4.1	61	3.9	175			
3	3	4	4	3	4	3.4	51	3	3	3	3	4	3	3	3	3	4	4	3	3	4	4	3.3	50	3.3	150			
3	4	3	3	3	4	3.5	52	3	3	4	4	4	3	3	2	3	4	4	3	4	4	4	3.5	52	3.6	162			
3	3	4	3	4	4	3.7	55	4	5	3	3	4	3	4	4	3	4	4	4	3	4	3	3.7	55	3.6	161			
7	6	6	7	5	8	6.7	101	6	11	5	4	3	10	4	6	7	6	7	8	6	8	7	6.5	98	6.8	306			
3	4	4	5	4	6	3.9	58	3	6	2	2	2	6	3	4	3	4	4	5	4	5	4	3.8	57	3.9	177			
0.4	0.7	0.7	0.7	0.8	0.8	0.6	9.1	0.5	0.5	0.4	0.5	0.7	0.6	0.8	0.7	0.4	0.7	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.6	8.9	0.6	26.9	
3	4	4	5	2	4	3.7	55	2	5	1	1	1	4	2	5	4	3	4	5	5	4	4	3.3	50	3.6	162			
0.4	0.7	0.7	0.7	0.4	0.5	0.6	8.2	0.3	0.5	0.2	0.3	0.3	0.4	0.5	0.8	0.6	0.5	0.6	0.6	0.8	0.5	0.6	0.5	7.5	0.5	23.8			
28	76	23	73	75	73	58.3	874	66	97	25	72	12	229	20	123	41	49	52	99	78	67	91	74.7	1121	56.1	2523			
645	877	656	1342	778	1354	1055.3	15830	1765	1888	980	1021	1127	1743	964	1215	1011	989	897	1021	1047	1202	1170	1202.7	18040	1341.4	63363			
64.5	62.6	109.3	122.0	45.8	84.6	137.5	2062.2	84.0	118.0	98.0	85.1	75.1	193.7	321.3	93.5	67.4	82.4	64.1	92.8	95.2	85.9	90.0	109.8	1546.5	128.0	5750.4			
107.5	109.6	109.3	167.8	111.1	169.3	246.0	3690.0	160.5	314.7	163.3	102.1	93.9	217.9	482.0	121.5	91.9	98.9	74.8	127.6	174.5	240.4	130.0	172.9	2594.0	221.1	9948.6			
10	14	6	11	17	16	10.5	157	21	16	10	12	15	9	3	13	15	12	14	11	11	14	13	12.6	189	12.2	547			
6	8	6	8	7	8	5.7	86	11	6	6	10	12	8	2	10	11	10	12	8	6	5	9	8.4	126	7.3	328			
65.7	88.9	65.7	88.9	77.8	88.9	63.7	955.7	84.6	46.2	46.2	76.9	92.3	61.5	15.4	76.9	84.6	76.9	92.3	61.5	46.2	38.5	69.2	64.6	969.2	64.3	2891.7			
60.0	57.1	100.0	72.7	41.2	50.0	59.4	891.4	52.4	37.5	60.0	83.3	80.0	88.9	66.7	76.9	73.3	83.3	85.7	72.7	54.5	35.7	69.2	68.0	1020.1	62.5	2814.1			
31.44	37.89	23.21	39.57	39.80	50.21	40.05	600.74	53.20	44.96	36.27	38.76	38.95	82.88	35.00	77.08	35.87	43.52	41.58	46.54	46.90	62.14	49.76	48.89	733.41	44.06	1982.49			
3.14	2.71	3.87	3.60	2.34	3.14	67.52	4.50	2.53	2.81	3.63	3.23	2.60	9.21	11.67	5.93	2.39	3.63	2.97	4.23	4.26	4.44	3.83	4.49	67.36	4.11	184.73			
5.24	4.74	3.87	4.95	5.69	6.28	8.57	128.48	4.84	7.49	6.05	3.88	3.25	10.36	17.50	7.71	3.26	4.35	3.47	5.82	7.82	12.43	5.53	6.92	103.76	7.14	321.26			
2	3	4	3	3	5	3.5	53	3	2	2	3	5	4	4	2	3	3	4	3	3	4	4	3.3	49	3.5	156			

- V10 - 브라우징 문헌 수
- V11 - 접속시간(초)
- V12 - 출력문헌 건당 소요시간(초)
- V13 - 적합문헌 건당 소요시간(초)
- V14 - 출력문헌 수
- V15 - 적합문헌 수
- V16 - 재현율(%)
- V17 - 정확률(%)
- V18 - 총 탐색비용(US)
- V19 - 출력문헌 건당 소요시간(US)
- V20 - 적합문헌 건당 소요비용(US)
- V21 - 탐색결과에 대한 만족도

[부 록 2] 탐색전문가의 탐색질문의 분석, 탐색과정, 탐색결과의 기본 통계표

Q	탐색질문 1															탐색질										
	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	평균	스	계	B1	B2	B3	B4	B5	B6	B7	B8
V1	3	2	2	3	3	4	4	3	3	3	4	1	2	3	1	2.7	41	3	1	1	2	3	3	4	2	3
V2	4	1	4	2	3	3	3	2	2	2	4	1	3	3	2	2.6	39	3	2	1	2	3	4	3	2	3
V3	3	2	5	3	3	3	4	3	3	4	4	1	3	2	3	3.1	46	3	2	2	3	3	4	3	2	3
V4	3	3	3	3	2	4	4	3	3	3	4	1	3	4	3	3.1	46	4	3	3	3	2	3	3	3	3
V5	5	4	4	2	3	3	4	3	3	3	5	3	3	3	4	3.5	52	4	3	3	3	2	3	4	2	3
V6	4	5	4	1	2	3	3	3	3	5	4	4	3	3	3	3.4	51	4	3	3	2	2	3	5	2	4
V7	0.5	0.5	1.0	0.5	0.7	1.0	0.8	1.0	1.0	1.7	0.8	1.3	1.0	1.0	0.8	1.0	14.9	1.0	1.0	1.0	0.7	1.0	1.0	1.3	1.0	1.3
V8	5	4	4	2	3	3	3	4	7	4	8	3	3	3	4	4.5	67	5	2	3	2	3	3	7	5	5
V9	2.2	1.3	1.0	1.0	1.0	1.0	0.8	1.3	2.3	1.3	1.6	1.0	1.0	1.0	1.0	1.3	18.8	1.3	0.7	1.0	0.7	1.5	1.0	1.8	2.5	1.7
V10	0	0	0	0	15	0	0	0	0	0	0	0	0	0	43	3.9	58	0	0	0	4	15	0	0	0	0
V11	1664	566	1243	535	970	1666	1176	1320	1090	1134	998	1290	491	530	793	1018	15268	695	770	404	1256	511	1147	995	1011	720
V12	48.4	48.3	51.8	41.2	53.9	64.2	51.1	47.1	47.4	45.4	43.4	43.0	32.7	31.2	39.7	45.9	688.8	57.9	64.2	25.3	59.8	31.9	54.6	66.3	56.2	55.4
V13	166.4	197.3	124.3	48.6	107.8	166.6	117.6	146.7	121.1	113.4	110.9	258.0	49.1	48.2	88.1	114.3	1714.3	86.9	128.3	80.8	157.0	63.9	143.4	165.8	144.4	120.0
V14	22	22	24	13	18	26	23	26	23	25	23	30	15	17	20	21.8	327	12	12	16	21	16	21	15	18	13
V15	10	9	10	11	9	10	10	9	9	10	9	5	10	11	9	9.4	141	8	6	5	6	8	8	6	7	6
V16	53.3	75.9	83.3	91.7	75.0	83.3	83.3	75.0	75.0	83.3	75.0	41.7	83.3	91.7	75.0	78.3	1174.9	88.9	66.7	55.6	88.9	88.9	88.9	66.7	77.8	66.7
V17	45.5	45.0	41.7	84.6	50.0	38.5	43.5	32.1	39.1	40.0	39.1	16.7	66.7	64.7	45.0	46.2	692.2	66.7	50.0	31.3	38.1	50.0	38.1	40.0	38.9	46.2
V18	34.95	42.34	36.52	27.75	38.76	44.76	39.00	45.75	39.24	43.12	40.35	50.25	27.48	31.96	37.60	38.88	583.23	27.62	26.50	26.00	42.06	31.24	36.65	33.55	36.98	27.50
V19	1.55	2.12	1.52	2.13	2.15	1.72	1.70	1.67	1.71	1.72	1.75	1.68	1.63	1.88	1.88	1.81	27.15	2.30	2.38	1.63	2.00	1.95	1.75	2.24	2.05	2.12
V20	3.90	4.70	3.89	2.52	4.31	4.48	3.90	5.19	4.36	4.31	4.48	10.05	2.75	2.91	4.18	4.37	65.53	3.45	4.75	5.20	5.26	3.91	4.58	5.59	5.28	4.58
V21	3			3	3	3	3	4	3	3	2	1	3	3	3	2.6	39	4	3	2	3	2	3	3	2	3

주) Q - 탐색질문
 T - 피험자(탐색전문가)
 V - 변인
 V1 - 탐색질문에 대한 이해도
 V2 - 탐색어 선정의 난이도
 V3 - 핵심 탐색어 선정의 난이도

V4 - 탐색결과에 대한 기대수준
 V5 - 탐색어 수
 V6 - 불 연산자 수
 V7 - 탐색어 한개당 불 연산자 수
 V8 - 제한 연산자 수
 V9 - 탐색어 한개당 제한 연산자 수

