

하이퍼텍스트 기반 디소러스의 실험적 설계와 운용

노 진 구*

<목 차>

- | | |
|-------------------------|-----------------------|
| I. 서 론 | Ⅲ. 하이퍼텍스트 기반 디소러스의 특성 |
| Ⅱ. 정보검색에서의 디소러스 | Ⅳ. 시스템의 설계 및 실행 |
| 1. 자연언어와 통제언어 | 1. 시스템 개요 |
| 2. 통제어휘집으로서의 디소러스 | 2. 실험시스템의 설계 |
| 3. 온라인 디소러스에 의한 탐색용어 선정 | 3. 시스템 실행 |
| | V. 결 론 |

I. 서 론

색인은 개개의 정보자료의 특성을 표현하는 데이터요소를 추출하여 각 정보자료의 내용을 대표하게 하는 것으로 그 역할은 특정한 내용의 정보가 어떠한 정보자료에 포함되어 있는가를 알려주며 방대한 정보원으로부터 가장 유사한 내용의 정보자료만을 선별하게 해준다. 또한 이용자가 관심을 가지고 있는 주제에 속하는 그러한 문헌들을 효율적으로 검색할 수 있도록 해 준다. 이와같은 역할을 고려해 볼때 색인은 정보검색시스템에서 이용자의 정보요구를 충족시켜줄 수 있는 이상적인 형태로 작성되어야 하는데, 이를 위한 하나의 수단으로 색인작성 보조도구인 디소러스의 사용을 들 수 있다. 디소러스는 색인작성자가 문헌의 주제를 표현하기 위하여 사용될 뿐만 아니라 탐색자가 문헌들에 대한 질의어를 표현하기 위해서 사용될 수 있다.

* 동래여전 문헌정보과 겸임교수

특히 훌륭한 디소러스의 사용은 유용한 단어를 찾을 수 있는 것 이상의 도움을 주는데, 이용자가 잘 알지 못하는 분야의 영역에서 탐색을 할 때 디소러스는 그 주제를 보다 잘 이해 할 수 있도록 도와주며, 이용자가 사용하려고 생각하고 있는 용어들간의 관계를 보여줌으로써 온라인 탐색에서 흔히 일어나는 폴스 드롭(false drop)의 발생을 감소시킬 수 있다. 한마디로 정보검색의 성공여부는 이용자의 질의어 용어와 문헌을 표현하고 있는 탐색용어간의 부합 문제에 달려있다고 볼 때 훌륭한 디소러스의 사용은 그만큼 중요한 의의를 갖게 되는 것이다.

기존의 디소러스는 대부분 인쇄형태로 이루어져 있으며 이런 형태는 시간과 공간의 제약으로 인해 어떤 개념을 찾았을때 그 개념에 관련된 모든 정보를 한꺼번에 찾아볼 수 없으며,¹⁾ 관련 계층을 찾아보기 위해서는 한번에 하나씩 모든 디스크립터를 찾아보아야 하는 수고가 따른다.

이 문제를 완화하기 위해서 일부 데이터베이스에서는 기계가독형 디소러스화일이 내장되어 있어서 특정한 탐색어와 관련된 다른 용어들을 찾아 볼 수 있으며, 자연어 색인화일(도치색인 화일)로 부터는 자모순으로된 키워드들을 인접 연산자를 사용하여 어느 정도의 개념관계를 표현하여 정보를 찾아볼 수 있다. 그러나 인접 연산자들에 의한 자연어 탐색이나 온라인 디소러스에 의한 통제어 탐색은 둘 다 이용자가 인접 연산자 사용방법과, 온라인 디소러스의 탐색 방법을 알아야 할 뿐만 아니라 용어선정이 순차적으로 이루어질 뿐, 역방향으로는 탐색하지 못하는 등, 용어 탐색을 위한 이용자와 시스템간의 상호작용 성능이 미약하다. 그런데, 오늘날 정보탐색 전략기법으로 그 중요성이 점차 인식되고 있는 브라우징을 지원하는 하이퍼텍스트 기반의 디소러스는 용어들간의 의미적 관계를 탐색할때 반구조적(semi-structured)이고 반방향적(semi-directed)인 네비게이션을 보조할 수 있고, 컴퓨터 통제 인터페이스를 위하여 직접조작 기법을 사용하기 때문에 사용이 용이하고

1) 최석두, "디소러스의 표시형식에 관한 연구," 제1회 한국정보관리학회 전국논문대회 논문집(서울, 1994). p. 105.

시스템과 이용자간의 상호작용을 원활하게 해준다.

따라서 본 논문은 지금까지 국내에서 연구된 바가 없는 하이퍼텍스트 기반의 한글 디소러스를 실험적으로 설계하여 운용해 봄으로써 서지데이터베이스의 주제영역에 대한 용어탐색을 위하여 브라우징 혹은 네비게이션을 보조할 수 있게 함으로써 정보검색의 효율을 향상시키려는데 그 의의가 있다.

II. 정보검색에서의 디소러스

1. 자연언어와 통제언어

문헌을 색인하고 탐색하기 위해서는 두 가지 기본적인 방법이 사용되는데 즉 통제어휘와 자연어에 의한 방법이다.

색인어 및 검색어 선택시 용어에 대한 통제여부에 따라 자연어색인(natural language index)과 통제어색인(controlled language index)으로 구분한다. 자연어색인은 정보자료 속에 있는 형태 그대로의 용어를 색인어로 채택하거나 색인자가 임의로 색인어를 주게되므로 같은 개념이 여러개의 다른 용어에 의해 표현될 가능성이 있으며²⁾ 또한 많은 단어들이 하나 이상의 의미를 가지고 있기도 하다.

비통제어휘 혹은 자연언어는 문헌의 실제 단어들을 색인 용어로 사용하는 어휘라는 점에서 자연언어를 사용하는 색인작성은 용어를 추출한다고 말하여진다. 왜냐하면 색인에 사용되는 용어들이 문헌의 본문이나 어떤 다른 정보 단위부터 추출되기 때문이다. 비통제언어를 사용하는 색인작성 시스템에서는 색인작성 용어들 사이의 의미적관계들을 통제해야 할 노력이 필요없다. 이는 하나의 특정의 개념이 단위개념으로 색인되며 탐색시에 이용자가 필요한 용어들을 조합할 수 있기 때문이다.

2) 정영미, 정보검색론, 서울, 구미무역, 1993, p.55.

자연언어는 특별한 훈련을 받지 않은 대부분의 이용자들이 이해하기 쉽고 의미에 대한 낱양스가 표현될 수 있다는 커뮤니케이션의 수단으로서의 장점을 가지고 있다.” 그리고 자연언어를 사용한 색인 시스템은 동의어, 유사동의어, 변형된 단어 형태 및 철자를 포함하므로 특정성이 매우 높은 검색이 가능하며 높은 정도율을 제공하는 잇점이 있다. 또한 새로운 용어들을 즉시 사용할 수가 있어 입력비용이 최소화 된다는 장점이 있다.⁴⁾

이러한 장점들에 반해 자연언어를 사용한 색인의 단점으로는 첫째 한 주제에 대한 적합문헌의 완전한 리스트를 얻기 위해서는 키워드의 모든 가능한 변형들을 알아야 하고 탐색해야 한다. 이 문제를 해결하기 위해 절단(truncation)이라는 방법을 사용하지만 이것은 동일한 개념에 대한 접근을 허용하지 않기 때문에 완전한 해결책이 되지 못하고 있다.⁵⁾ 둘째로는 구문론적 문제들로 부정확한 용어관계로 인한 폴스 드롭에 대한 위험성이며,⁶⁾ 셋째로는 검색시 특정한 개념을 표현하는 모든 용어를 탐색어로 사용하지 않을 경우 재현율의 저하를 가져오며, 동형의어를 조절하지 않음으로써 정도율의 저하를 가져올 수 있다.⁷⁾

이와 같은 자연언어 색인의 특징들은 통제언어 색인이 가지는 특징들과 잘 대비된다. 통제언어는 동의어를 통제하고, 관련어를 제시함으로써 재현성을 향상시키며 또한 복합어와 동형의어를 통제하고, 한 개념의 계층적, 그리고 연관적 관계들을 명료히 기록하고 각 주제의 크기 혹은 범위를 확립한다. 그밖에도 단어 기반의 시스템을 위하여 통제된 어휘는 동의어 용어들을 확인하고, 그들 중에 한 개의 용어를 표준으로 선정한다.

3) C. T. Meadow, *Text Information Retrieval Systems*, San Diego, Academic Press, 1992, p.37.

4) Jean Aitchison, Alan Gilchrist, *Thesaurus Construction: A Practical Manual*, London, Aslib, 1987, p. 4.

5) Rao Aluri, D. Alasdoir Kemp and John J. Boll, *Subject Analysis in Online Catalogs*, Englewood, Colo., Libraries Unlimited, 1990, p.37.

6) Aitchison and Gilchrist, *op. cit.*, p.4.

7) Aluri, Kemp and Boll, *op. cit.*, p.33.

그러나 비록 통제어휘 탐색이 효율적 정보검색을 위한 매우 강력한 도구이고 매우 훌륭한 질적인 탐색결과를 생산할 수 있지만 통제어휘를 사용할 수 없을 경우가 있고 또 그것이 잘 적용하지 않을 때가 있다. 즉 모든 데이터베이스들이 다 통제어휘집을 가지고 있지 않다는 점과 서지데이터베이스가 아닌 수치 데이터베이스나 참고자료 데이터베이스는 통제어휘는 의미가 없기 때문에 통제어휘집을 갖고 있지 않은 점과 많은 데이터베이스들은 <미국 국회도서관 주제명 표목: LCSH>을 사용하고 있어 데이터베이스를 위한 색인언어와 색인어 표준들이 데이터베이스마다 다를 수 있다는 점에서 그러하다.⁸⁾

또한 디소러스의 사용이 정보탐색에 적합하지 않는 경우가 있다.

첫째, 비록 데이터베이스가 통제 어휘로 색인이 되어있다 할지라도 이용자가 그 색인방법을 모르기 때문에 자연어로 탐색을 시작할 수 밖에 없는 경우가 있고 둘째, 질의어 용어가 새로운 용어이고 아직 그 용어가 디소러스에 포함되어 있지 않을 경우이며 셋째, 질의어의 주제영역을 포함하는 하나의 표준용어가 없을 경우에는 자연어 탐색을 하게 된다.⁹⁾

랑카스터(F. W. Lancaster)는 일반적으로 여러학자들이 자연언어와 통제언어에 대해 제시하고 있는 비슷한 결론들을 다음과 같이 요약하고 있다.¹⁰⁾ 즉 “자연언어 탐색은 일반적 개념보다 개별적 개념의 경우에 훌륭한 결과를 나타내며 새로운 용어의 편입이 신속한 반면에 통제언어는 정도율을 떨어 뜨리는 경우가 있으나 용어간의 개념을 알 수 있고 이용자로 하여금 정확한 탐색용어를 발견할 수 있게 해주는 도구이다.”¹¹⁾

이상에서 살펴본대로 자연언어와 통제언어가 가지고 있는 서로서로의 장단점들은

8) Geraldene Walker, *Online Retrieval: A Dialogue of Theory and Practice*, Englewood, Colo., Libraries Unlimited, 1993, p.63.

9) *Ibid.*, p.86.

10) F. W. Lancaster, *Indexing and Abstracting in Theory and Practice*, London, The Library Association, 1991, p.201.

11) *Ibid.*, p.214.

자연언어뿐만 아니라 일부 통제된 용어들을 포함하는 혼합(hybrid) 시스템의 구축이 필요하다는 것을 대변해 주고 있기도 하다. 혼합접근의 가치는 통제언어탐색에 의해 발견되지 못한 일부 적합문헌들이 자연어 탐색에서 검색된 대부분의 수행된 연구에 의해 지지되고 있다.

2. 통제어휘집으로서의 디소러스

정보검색 효율을 높일 수 있는 방안의 하나인 용어통제 방식에는 분류표, 주제명 표목표, 코드집과 같은 코드통제 그리고 디소러스와 같은 도구들이 있다.

디소러스는 색인작성 단계에서는 색인표목을 선정할 수 있게 하고, 정보검색의 단계에서는 시스템에서의 검색효율의 제고를 위한 탐색용어의 선정을 도와주고 용어 상호간의 관계성을 밝혀줌으로써 검색의 정확성을 높일 수 있게 한다. 즉 디소러스는 용어들의 동의성, 다의성을 통제하여 색인자와 검색자의 용어선택이 일치될 이루도록 하며 용어개념의 관계를 표현함으로써 주제검색의 효율을 향상시키는 기능을 수행한다.¹²⁾

3. 온라인 디소러스에 의한 탐색용어 선정

1970년대 이후부터 점차 성행하게된 온라인 정보 검색 환경에서는 후조합 색인기법에 의한 온라인 탐색이 질의어 용어와 문헌용어가 일치되는 탐색을 가능하게는 하지만, 용어간의 관계가 잘못 이루어져서 부적합한 문헌을 출력할 수 있는 여지가 많다. 이러한 자연언어 탐색기법의 결함은 인접 연산자들의 사용에 의하여 어느 정도 완화될 수 있으나, 컴퓨터내에 온라인 디소러스, 예를들면 <The Thesaurus

12) F. W. Lancaster, *Vocabulary Control for Information Retrieval*, Alington, Information Resources Press, 1986.

of ERIC Descriptors)와 <The Thesaurus of Psychological Index Terms>등과 같은 온라인 디소러스가 내장되어 있다면 검색용어와 문헌의 색인용어를 표준화할 수 있음으로써 더욱 유용한 용어통제를 할 수 있게 된다.

기존의 온라인 검색 시스템들은 아직 명료하게 인식되지 않은 정보요구를 가지고 있거나 명료하게 인식은 하고 있으나 적절히 그 정보요구를 표현할 용어를 발견하지 못한 상태에 있는 이용자들에게 최상의 탐색용어들을 선택하는 것을 도와주기 위한 브라우저장치를 지원하고 있는 경우가 많지만 그 대부분이 구조적인 알파벳순 배열에 의한 용어 디스플레이를 제공하고 있다.

디소러스는 도치화일의 구조와 비슷하게 일련의 엔트리 용어로 구성되어 있고, 각 엔트리에 관련된 다른 용어들과 그 용어들의 관련을 보여주도록 축적되어 있다. 모든 데이터베이스에 반드시 디소러스가 있어야 하는 것은 아니며 만약 있다면 디소러스는 도치화일과 같은 방법으로 탐색된다.¹³⁾ 도치화일의 탐색 명령어로는 DIALOG 시스템의 EXPAND, ORBIT시스템의 NBR (neighbor의 약어), BRS의 ROOT가 있다. 세 가지 시스템 모두 각 용어아래 색인된 문헌수가 함께 디스플레이된다.¹⁴⁾

대부분의 용어 브라우징은 알파벳순 용어 배열을 디스플레이하여 이용자로 하여금 적절한 용어를 선정할 수 있게 한다. 서지데이터베이스를 구성하고 있는 문헌 레코드의 디스크립터 필드를 발견하기 위하여 온라인 디소러스내에 들어가야 한다.

탐색 용어를 선정한 후에 서지데이터베이스에서 디스크립터 필드를 검색할 수 있는 방법은 세 가지가 있다.¹⁵⁾

첫째, 이미 구로 묶여 있는 디스크립터를 탐색할 수 있으며,

둘째, 단일어로 된 디스크립터를 탐색할 수 있으며,

13) Meadow, *op. cit.*, p. 139.

14) 정영미, *삼계서*, p. 286.

15) Walker, *op. cit.*, pp. 63-64.

셋째, 단일어나 구로된 단어들을 포함하고 있는 모든 디스크립터를 탐색할 수 있다.

완전에 가까운 온라인 디소러스가 존재할 경우 디소러스를 이용한 용어선정은 매우 훌륭한, 질적인 탐색 결과를 생산할 수 있게 하는 강력한 도구가 될 수 있다. 그러나 탐색 중개자의 도움이 없거나 검색 경험이 풍부하지 않은 최종 이용자의 경우, 데이터베이스마다 상이한 디소러스의 구조에 대한 지식을 가지고 있거나, 특정 용어에 대한 정확한 디스크립터를 알고 있다는 것은 거의 불가능할 뿐만 아니라 EXPAND 명령을 사용해서 디스크립터를 발견해 나가는 과정을 터득하기 위해서도 상당한 훈련을 필요로 할 것이다.

또한 디스크립터에 의한 탐색방법은 적절한 교육이 필요하며 알파벳순의 브라우징에 의한 탐색은 속도가 느릴 뿐만 아니라 잘못 선택한 용어의 경우, 바로 이전으로 되돌아 갈 수 있는 역기능이 없음으로 탐색을 처음부터 다시 시작해야 한다는 제약점을 가지고 있다.

Ⅲ. 하이퍼텍스트 기반 디소러스의 특성

자연어 탐색 방법의 많은 이용에도 불구하고 디소러스는 여전히 대규모 온라인 서지데이터베이스를 색인하는데 사용되고 있다.¹⁶⁾ 지금까지는 대부분의 디소러스는 인쇄형태로나 온라인 디소러스 형태로 사용이 가능하지만 하이퍼텍스트 기법들을 적용하여 디소러스를 개발하려는 연구들이 외국에서는 수 년전 부터 이루어져 오고 있다. 하이퍼텍스트 기반의 디소러스는 기존의 인쇄형 디소러스나 온라인 디소러스가 구현할 수 없는 몇 가지 특성들과 하이퍼텍스트로 표현하기에 적합한 디소러스

16) L. M. Chan, and R. Pollard, *Thesauri Used in Online Database: An analytical guide*, Westport, Greenwood, 1988.

자체의 구조를 가지고 있다.

첫째로, 슈나이더맨(Ben Shneiderman)이 제시한 하이퍼텍스트의 응용 분야에 적합한 조건들,¹⁷⁾ 즉

- ① 대규모 정보가 수많은 조각들(fragment)로 조직되어 있어야 한다.
- ② 그 정보조각은 서로서로 관련되어 있어야 한다.
- ③ 이용자는 어느 시점에서 단지 작은 정보분량 만을 필요로 한다

는 것에 비추어 볼때 디소러스는 이 조건들에 아주 적절하다. 디소러스는 잘 정의된 방식으로 상호간에 명백히 관련되어 있는 수 천개의 개개 용어들을 포함하며 어느 특정한 순간에 디소러스의 이용자는 사용 가능한 용어들 중에 작은 부집합만을 찾아 보려는데 관심을 가진다.

다음으로, 하이퍼텍스트의 특성중의 하나는 그것이 이용자에게 비선형적 방법으로 정보공간을 네비게이트할 수 있게 하는 것이다. 알파벳순 디소러스의 이용자들은 알파벳순 인접에 근거한 선형적 배열이 동등, 계층적 혹은 연관적 관계에 의하여 연결된 용어들을 한 곳에 모으는데 특히 효과적이지 못하다는 것¹⁸⁾을 심각하게 깨닫고 있다. 그런데 디소러스의 구조가 본래 비선형적이고 하이퍼텍스트는 비선형적으로 정보를 조직하고 그것을 네비게이트할 수 있는 특성을 가지고 있다. 즉 하이퍼텍스트 기반의 디소러스에서는 이용자가 디소러스내를 '돌아다닐 수'있고 상세한 부분들에 '초점을 맞출 수'가 있다. 네비게이션 행위의 방향상실 문제를 해결하기 위하여 하이퍼텍스트 기반 디소러스의 이용자 인터페이스는 서포 및 히스토리 리스트 기능을 수행하는 버튼을 만들 수가 있다.

셋째, 콘클린은 하이퍼텍스트의 장점으로서 '새로운 참조정보의 생성이 용이하다'는 것을 지적하였으며 이것은 하이퍼텍스트의 노드들이 독립성을 갖고 있어 새로

17) B. Shneiderman, Reflections on Authoring, Editing, and Managing Hypertext, In (E. Barrett(Ed), *The Society of Text*, Cambridge, Mass., MIT Press, 1989, pp.115-131.

18) Pollard, "A Hypertext-Based Thesaurus as a Subject Browsing Aid for Bibliographic Databases," p.348.

은 디스크리터틀 추가하고 삭제하는등의 작업이 용이하게 이루어짐을 말해준다. 다시 말해서 파일의 갱신이 용이하며 새 용어의 입력과 용어의 수정이나 삭제가 자유롭게 때문에 훌륭한 디소러스를 유지할 수 있게 할 수 있다.

넷째, 하이퍼텍스트에 통합된 브라우징 기능에 의하여 이용자의 탐색용어 선정이 매우 효율적으로 이루어질 수 있다. 기존의 데이터베이스 접근방법이 매개변수에 의존하는 탐색(parameterised search : PS)에 초점을 두어온데 비하여 그 대안으로 개발된 브라우저(browser) 시스템에서 사용되어온 브라우징 기능은 잘 연결된 레코드 공간내에서의 휴리스틱 탐색(heuristic search)을 할 수 있게 한다.¹⁹⁾

용어탐색을 위한 브라우징 기능에 대한 학자들의 논의는 긍정적으로 나타나고 있다.

코브(J. F. Cove)는 "인식(recognition)은 환기(recall)보다 더 쉽기 때문에 브라우징은 탐색의 시각적 행위이며 잘못 정의된 정보요구를 만족시킬 수 있는 항목의 위치를 찾는 행위이며 정확한 탐색 질의어를 정의할 수 있게 하는 행위이다"²⁰⁾라고 지적하고 있다.

그는 또한 브라우징 전략에는 탐색 브라우징, 범용적인 브라우징, 우연적인 발견을 위한 브라우징등이 있으며 이중에서 탐색 브라우징은 이용자가 마음속에 특정의 목적을 가지고 있으나 반드시 명료하게 정의된 질의어 형성이 가능하지 않을때 디스플레이된 용어들을 눈으로 보면서 자신이 정확히 무엇을 바라는지를 인식하고, 탐색 도중에 발견하는 정보로부터 정확한 질의어를 구축해 갈 수 있다²¹⁾고 피력한다.

미콜라쥬크(Ibigniew Mikolajuk and Robert Chafetz) 등은 브라우저는 전문

19) Andrew J. Palay and Mark S. Fox, "Browsing Through Databases," in ed. by R.N. Oddy, et al., *Information Retrieval Research*, London, Butter-worths, 1981, p.310.

20) J. F. Cove, "Online Support for the Intellectual Activity of Browsing," *Unpublished Ph. D. Dissertation*, University of Liverpool, 1987, p.49.

21) *Ibid.* pp.50-51.

영역 용어들을 학습하는데 사용될 수 있는데 즉 개념들의 존재에 대하여 점검하고 개념들의 구조에 대한 상세한 표현을 보는데 사용될 수 있다²²⁾고 언급한다.

이상의 몇몇 학자들의 논의는 ‘하이퍼텍스트에 통합된 브라우징과 네비게이션 기법이 하이퍼텍스트 기반의 디소러스’를 이용하여 탐색자가 비선형적으로 그리고 자유로운 이동을 행하면서 인식을 바탕으로 하는 적합용어 선정의 과정을 매우 원활하게 이행할 수 있게 한다는 것을 주장하고 있다고 할 수 있다.

마지막으로 하이퍼텍스트 기반의 디소러스는 그 구축과 사용에 있어 컴퓨터 통제 인터페이스 방법들 명령어 방식, 메뉴방식, 직접 조작 방식-을 포함하고 있음으로써 단순히 메뉴나 명령어만을 사용하는 시스템보다 더 우수한 이용자와 컴퓨터간의 상호작용 성능을 가질 수 있다.

직접조작(direct manipulation)이란, WIMP(window, icons, mouse, popup menu) 인터페이스와 관련지어지는 특징들을 언급하기 위해 슈나이더만이 처음으로 만들어낸 용어로서, 명령어나 메뉴에 의한 중간코드를 통하여 물체를 불러내기 보다는 시스템에서 물체의 표현을 이용자가 보고 직접 조작하는 것을 말한다.²³⁾

IV. 시스템의 설계 및 실험

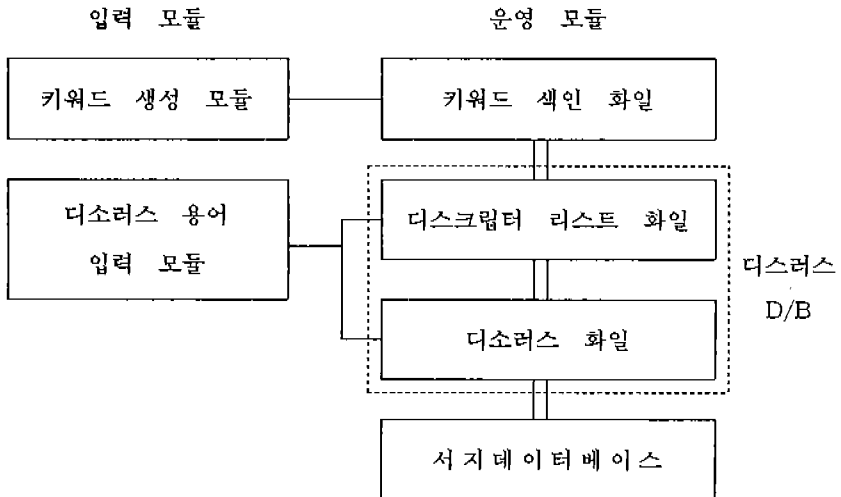
1. 시스템 개요

이 실험 시스템에서의 최초의 접근은 키워드 색인에 의해 시작되며 이 색인은

22) Zbigniew Mikolajuk and Robert Chajetz, "A Domain Knowledge-Based Natural Language Interface for Bibliographic Information Retrieval," in ed. by Martin Dillon, *Interfaces for Information Retrieval and Online Systems*, New York, Greenwood Press, 1991, p.86.

23) Ben Shneiderman, "Direct Manipulation: A Step Beyond Programming Languages," *IEEE Computer*, Vol.16, No.8(Aug. 1983), p.57.

디스크립터이건 아니건 상관없이 디소러스 용어들을 형성하는데 사용되는 모든 단어를 포함한다. 이 색인에 기재된 모든 용어들은 하이퍼텍스트 링크를 사용하여 디소러스 데이터베이스에 연결되며 이렇게 연결된 디소러스내에서의 디스크립터들은 서지적 데이터베이스에 연계되어 검색결과에 대한 건수를 제공한다. 실험 시스템의 구성을 그림으로 나타내면 <그림 1>과 같다.



<그림 1> 시스템 구성도

1) 실험시스템 환경

본 연구에서 사용한 하드웨어는 IBM 486 DX급 개인용 컴퓨터를 사용하고 사용자 인터페이스는 한글윈도우 3.1을 이용하였으며 엔진 및 사용자 인터페이스는 C++를 이용하여 객체지향적으로 설계하였다.

2) 실험시스템의 데이터베이스

실험시스템의 데이터는 크게 하이퍼텍스트 디소러스 데이터베이스부분과 시스템

에 연결된 서지 데이터베이스 두 가지로 구분할 수 있다.

먼저 본 연구를 위한 하이퍼텍스트 디소러스의 준비는 디소러스의 주제영역을 정보관리분야로 정하고 기존의 인쇄판 디소러스 중 《유네스코 디소러스: Unesco Thesaurus》²⁴⁾의 Information Libraries and Archives(Z Section)부분의 용어들 중에서 378개의 관련용어를 추출하여 한글로 번역하였다. 추출용어들의 한글화를 위해서는 《해로드(Harrod) 도서관학 용어집》²⁵⁾과 정보학 관련문헌들²⁶⁾을 참조했으며 대부분의 용어번역은 도서관학 정보학 용어사전²⁷⁾을 근거로 하였다.

다음으로 하이퍼텍스트 기반의 디소러스와 연결되는 서지데이터베이스의 구축은 실험시스템의 가동을 위해서 기존의 검색 시스템중의 하나를 이용하지 않고 실험용 서지데이터베이스를 구축하였다. 서지데이터베이스의 입력자료는 국내에서 이미 발표된 자료를 대상으로 《한국문헌정보학 색인: 1975-1992》²⁸⁾에서 수록된 일부 주제분야(색인·색인법, 색인·색인법·자동색인, 정보검색, 정보검색·검색기법, 정보검색·언어, 정보검색·온라인검색, 정보서비스)에서 310개 서지데이터를 입력하였으며, 입력항목은 제목, 저자, 서지사항과 디스크립터로 하였다. 검색을 위한 키워드는 제목에서 추출한 용어들과 디소러스에서 사용된 용어들로 구성하였다. 일반적인 연구방법을 나타내는 명사나 관사, 조사등은 키워드에서 제외시켰다.

2. 실험시스템의 설계

1) 데이터 모델

하이퍼텍스트 기반의 디소러스의 개발에 있어서 디소러스를 하이퍼텍스트로 사상

24) Unesco, *UNESCO Thesaurus*, Paris, UNESCO, 1977.

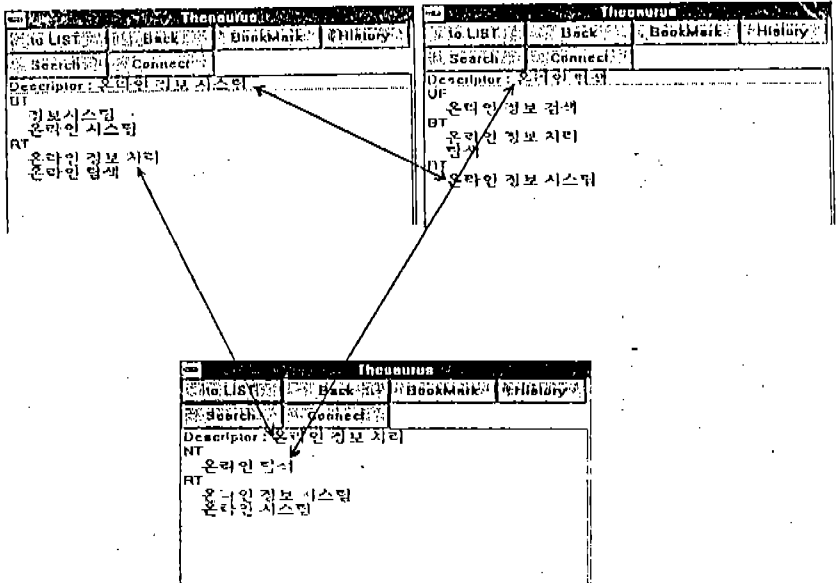
25) L. M. Harrod, *Harrod's Librarians' Glossary*, Brookfield, Gower, 1990.

26) 사공철, *정보검색론*, 서울, 아세아출판사, 1977; 정영미, *정보검색론*.

27) 한국도서관협회, *도서관학정보학 용어사전*, 서울, 한국도서관협회, 1986.

28) 박준식, *이애란, 한국문헌정보학 색인: 1975-1982*, 서울, 한국도서관협회, 1994.

하기 위한 데이터 모델은 용어의 상관관계가 표시된 하나의 디소러스 용어를 노드로 하고 각 용어간의 관계를 노드들 사이의 링크로 할당한다.



〈그림 2〉 디스크립터의 연관적 링크

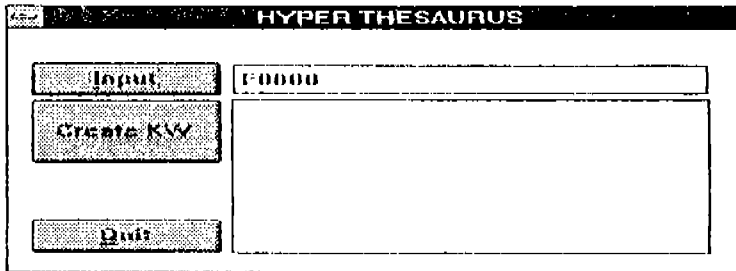
디소러스에 포함된 세 가지 유형의 관계는 〈그림 2〉와 같이 연관적 링크 (associative links)로 표현된다. 하이퍼텍스트 기반 디소러스에 대한 최초의 접근은 키워드 색인으로 시작되며 이 색인은 디스크립터이거나, 혹은 디스크립터로 사용할 수 없는 용어이건 관계없이 디소러스 용어들을 형성하는 모든 단어들로 구성된다. 이 색인이 디소러스내의 어떤 디스크립터와 일치하는 경우는 해당되는 디스크립터를 포함하고 있는 노드에 직접 연결된다. 그렇지 않을 경우에는 키워드가

포함된 여러 디소러스 용어리스트에 연결되어 적합용어를 선택할 수 있도록 한다.

키워드와 관련된 키워드색인(keyword index)은 알파벳순으로 배열되어 있으며, 이 색인에 나타난 모든 용어들은 디스크립터이전 디스크립터로 사용할 수 없는 용어들이전 간에 하이퍼텍스트 링크를 사용하여 디소러스 데이터베이스내의 각 용어에 연결된다.

2) 하이퍼텍스트 기반의 디소러스 구축

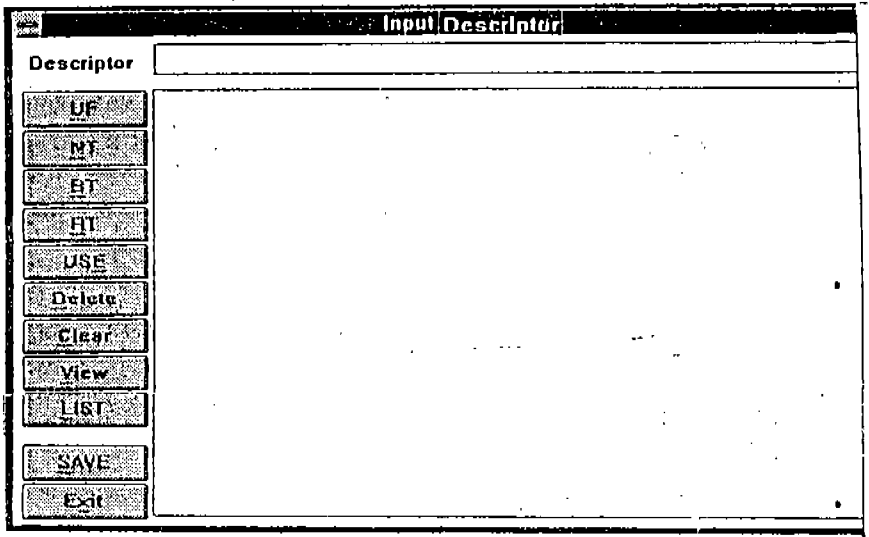
디소러스 데이터베이스의 구축을 위한 초기입력작업 선택화면은 <그림 3>과 같이 설계하였고, 용어들간의 관계가 표시된 디소러스의 기본 입력형태는 <그림 4>과 같이 설계하였다. 그리고 키워드 색인작성을 위한 입력화면은 <그림 5>와 같다.



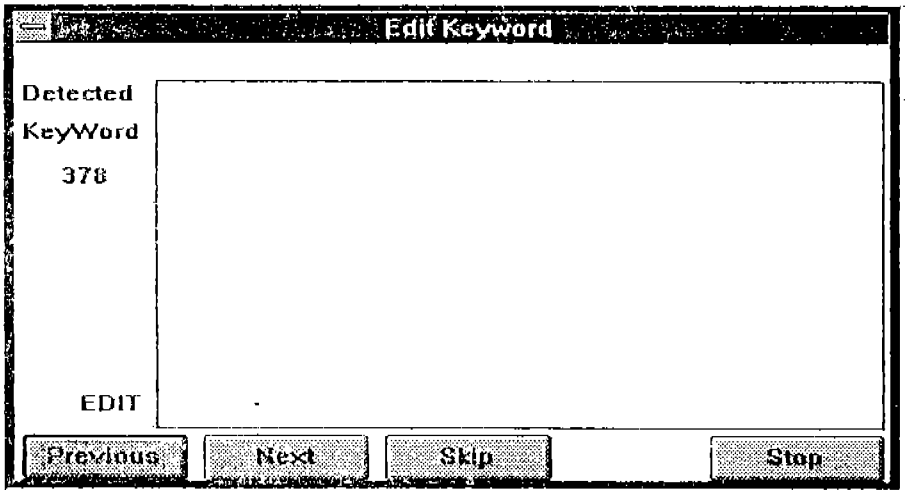
<그림 3> 초기 입력작업 선택화면

<그림 3>에서 <Input>버튼은 디소러스를 작성하기 위한 입력작업을 수행하기 위한 버튼이며, "F0000"는 입력된 디스크립터 데이터의 관리를 위한 내부 파일명이다. <Create KW>버튼은 입력된 디스크립터들의 관련있는 키워드파일의 작성을 위해 사용하는 버튼이다. <Quit>는 입력작업을 종료하기 위한 버튼이다.

다음 화면은 <그림 3>에서 <Input>버튼을 선택하였을때 디스플레이되는 디스크립터의 기본 입력화면이다.



<그림 4> 디스크립터의 기본 입력화면



<그림 5> 키워드 색인 작성화면

이 화면은 <그림 3>에서 <Create KW>버튼을 선택하였을때 디스플레이되는 화면이다. 키워드 색인 화일의 작성은 디스크립터를 이용자가 정확히 모를 경우에 대비하여 디스크립터를 구성하고 있는 용어들을 분리시켜 부분일치에 의한 디스크립터 탐색이 이루어지도록 하였다.

3) 이용자 인터페이스

하이퍼텍스트 기반의 디소러스에 대한 이용자 인터페이스는 다음과 같은 세 가지 단계로 나누어진다.

(1) 키워드 색인 탐색

이 단계는 디소러스 데이터베이스내의 디스크립터와 디스크립터로 선택되지 않은 용어를 구분하지 않고 모든 단어들을 탐색할 수 있게 한다. 이 단계에서는 시스템내에서 자기가 탐색하고자 하는 단어를 먼저 입력하여 디소러스 용어들을 형성하고 있는 단어들 중에 이용자가 원하는 단어가 있는지를 먼저 확인하는 과정으로 단어를 입력하거나 스크롤(scroll) 함으로써 하나의 단어를 선택하여 마우스로 클릭(click)하여 디스크립터 리스트로 연결시킬 수 있다. 이 탐색과정에서의 이용자 인터페이스는 두 개의 버튼(exit, descriptor)과 탐색어 입력박스(box)와 스크롤 기능을 가진 윈도우로 구성된다.

(2) 디스크립터 리스트 제공

이 단계에서는 키워드 색인 탐색과정에서, 탐색된 색인어가 포함하고 있는 디스크립터들의 리스트를 보여준다. 디스크립터로 채택되지 않은 용어들도 함께 디스플레이된다. 이 단계에서의 이용자 인터페이스는 두 개의 버튼(back, thesaurus)만으로 이루어지고 있다. <back>버튼은 이전의 키워드 탐색단계로 다시 돌아가는 기능이며 <thesaurus>버튼은 이용자가 선택한 디소러스내의 디스크립터를 디스플레이한다.

(3) 디소러스 탐색

이 단계는 이용자가 선택한 하나의 디스크립터 노드를 디스플레이하여 준다. 디스크립터와 용어들간의 관계들을 표시하기 위해 USE/UF, BT/NT, 그리고 RT 를 사용한다.

하이퍼텍스트 디소러스에서 이 기호들 아래 열거된 용어들은 해당 노드에 대한 링크들을 위한 시작점(anchor)이다. 또한 디스플레이 된 노드가 시작노드(anchor node)가 되는 것이다.

이 단계에서의 이용자 인터페이스는 다음과 같은 버튼들이 있다.

- ① list - 디스크립터 리스트 제공단계로 다시 돌아가는 기능
- ② back - 현재의 디스플레이 되고있는 디스크립터 바로 전 디스크립터(직전 노드)로 돌아가는 기능
- ③ bookmark - 디스크립터 탐색중 필요하다고 생각되는 노드를 표시하여 나중에 다시 확인하는 책갈피 기능
- ④ history - 현재까지 탐색해온 모든 디스크립터들의 기록 축적 기능
- ⑤ search - 하이퍼텍스트 디소러스에서 선택한 특정 디스크립터를 입력하여 서지 데이터베이스의 관련 문헌전수를 검색하는 기능

3. 시스템 실행

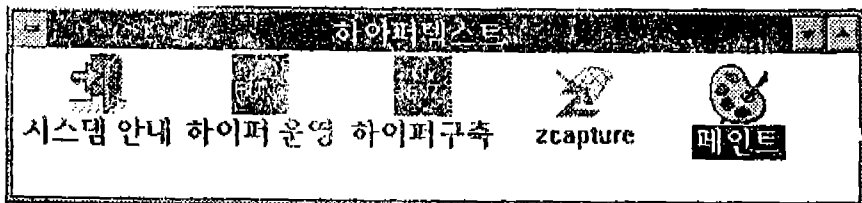
본 시스템 실행과정은 먼저 하이퍼텍스트 기반의 디소러스의 구축부터 시작하여 디소러스의 디스크립터를 탐색하기 위한 키워드 색인 작성단계를 거치면서 디소러스 데이터베이스가 구축된다. 이 단계에 대한 설명은 실험시스템의 설계부분에서 다루었다.

이 부분에서는 구체적인 사례를 들어 키워드 색인의 탐색단계, 디스크립터 리스트의 제공단계를 설명하고 디스크립터가 표시되는 하이퍼텍스트 디소러스내에서의

링크에 의한 디스크립터의 탐색단계를 설명한다. 마지막 단계로 이용자가 디소러스에서 선택한 디스크립터에 의해 검색된 서지 데이터베이스 검색 결과를 보여준다.

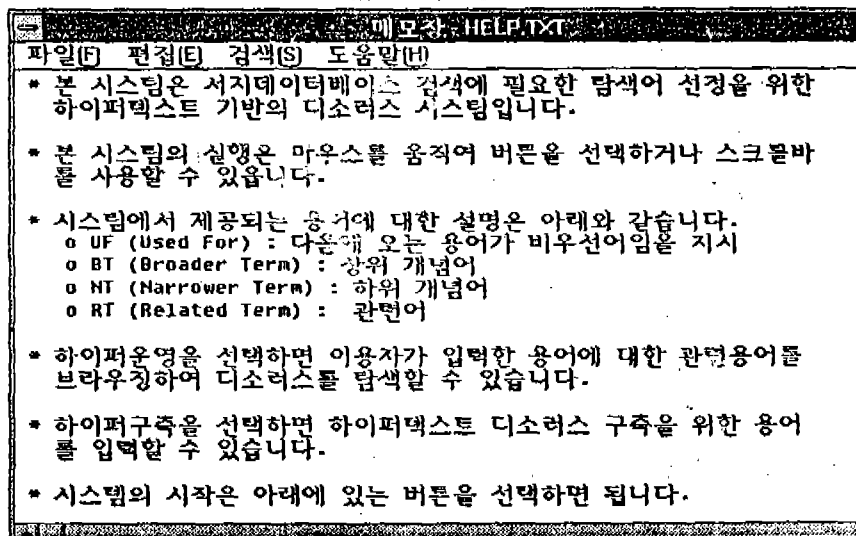
1) 초기화면과 디소러스 구축

Window 3.1 프로그램을 실행시키면 제일 먼저 <그림 6>과 같은 시스템 시작화면이 나타난다.



<그림 6> 시스템 시작화면

이 화면의 다섯개의 아이콘 중 <시스템 안내> 버튼을 클릭하게 되면 <그림 7>과 같은 시스템 안내가 제공된다.



<그림 7> 시스템 안내 화면

다음으로 디소러스의 디스크립터 군을 구축하는 기능을 가진 <하이퍼구축> 버튼을 마우스로 클릭하여 실행하게 되면 제일 먼저 앞에서 설명한 하이퍼텍스트 디소러스 구축을 위한 <HYPER THESAURUS>라는 이름의 초기입력작업 선택화면이 나타나고, 이 중에서 <Input>버튼을 클릭하게 되면 하이퍼텍스트 디소러스의 디스크립터 입력화면이 <그림 4>와 같이 <Input Descriptor>라는 제목으로 디스플레이 된다.

Input Descriptor	
Descriptor	컴퓨터 목록
UF	>> 컴퓨터 목록
UF	컴퓨터 장서 목록 장서 목록의 자동화
NT	COM 목록
BT	목록
USE	장서 목록
Delete	RT
Clear	컴퓨터색인 인쇄 목록
View	
LIST	
SAVE	
Exit	

<그림 8> 디스크립터 '컴퓨터 목록'의 관련용어들을 입력한 상태

한 예로써 '컴퓨터 목록'이라는 디스크립터를 입력하게 될 경우는 디스크립터 입력을 위한 박스에 '컴퓨터 목록'을 입력한 다음 화면 좌측에 있는 버튼 중 해당되는 용어간의 관계를 클릭한 다음 관련용어들을 입력한다. 그리고 나서 다시 화면

좌측의 버튼 중 <SAVE>버튼을 누르면 저장이 된다. 이와 같은 방법으로 '컴퓨터 목록'이 가지는 모든 용어간의 관계를 입력하고 <SAVE>시키게 된다. 이렇게 입력된 화면은 <그림 8>과 같다.

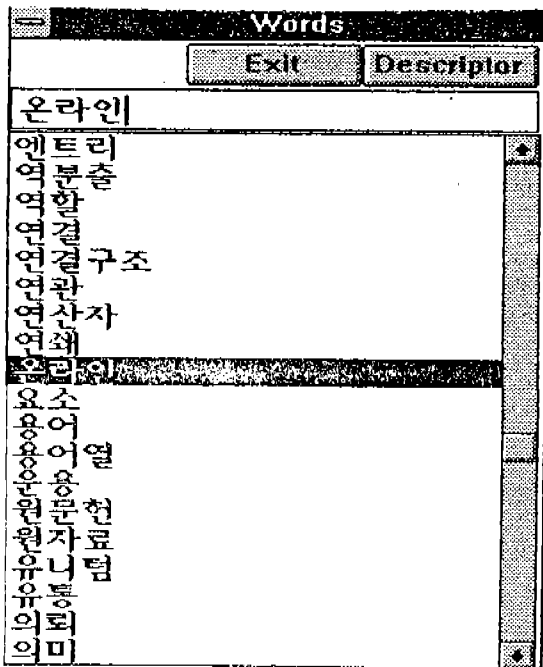
한 개의 디스크립터에 대한 용어관계를 모두 입력시키고 난 뒤 필요하다면 <View>버튼을 클릭하여 전체적으로 입력된 사항들을 <그림 8>과 같이 확인할 수 있다. 입력작업이 모두 끝난뒤에는 <LIST>버튼을 사용하여 입력된 전체 디스크립터들을 확인할 수 있다. 입력작업중에 수정은 <Clear>버튼을 사용하며, 모두 입력한 후 일부 디스크립터를 삭제하기 위해서는 <Delete>버튼을 사용한다. 작업이 끝난 뒤에는 <Exit>버튼으로 빠져 나오게 된다.

그 다음으로는 키워드색인 작성을 위해 <그림 3>의 <HYPER THESAURUS> 화면에서 <Create KW>버튼을 선택한 뒤 <그림 5>의 <Edit Keyword> 화면이 나오게 한 뒤 빈 공간에 앞서 입력한 디스크립터에 관련된 용어들을 입력하여 키워드 색인을 만드는 작업을 진행시킨다. 예를 들어 디스크립터 '컴퓨터 목록'에 관련된 용어들로 '컴퓨터'와 '목록'이라는 키워드를 입력시키고 난뒤 <NEXT>버튼을 선택하면 입력된 키워드가 저장된다. 그리고 작업을 끝내기 위해서는 <STOP>버튼을 선택한다.

2) 탐색용어 선정

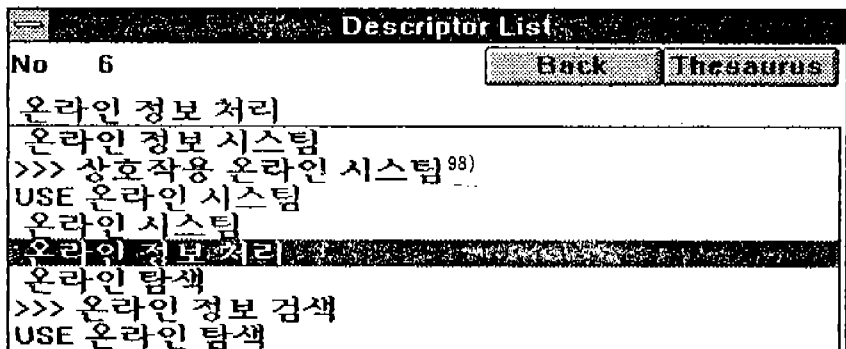
탐색어 선정은 초기화면의 4개의 아이콘 중 <하이퍼 운영>이라는 프로그램을 실행시킴으로써 시작된다. 프로그램을 실행시키면 제일 먼저 <그림 9>와 같은 <Words>란 제목을 가진 키워드 색인 화일이 나타난다.

여기에서 이용자는 제일 상단의 빈 공간에 자기가 탐색하고자 하는 용어를 입력하거나 오른쪽의 스크롤 버튼으로 자기가 원하는 용어를 찾는다. 직접 입력한 용어가 키워드로 선정되지 않았다면 가장 부분적으로 부합된 인접 키워드가 표시된다.



〈그림 9〉 키워드 '온라인'의 키워드 색인 화일

3) 디스크립터 리스트 확인



〈그림 10〉 '온라인'이 포함된 디스크립터 리스트

29) >>>기호는 디스크립터로 채택되지 않은 용어를 나타낸다.

이용자가 가장 적합하다고 생각하여 입력한 용어가 <그림 9>와 같이 키워드 색인 내에 존재하거나, 부분 부합이 되어 있으면 이용자가 <그림 9>에서 <Descriptor> 버튼을 클릭하게 되면 디스크립터 리스트가 <그림 10>과 같이 나타난다.

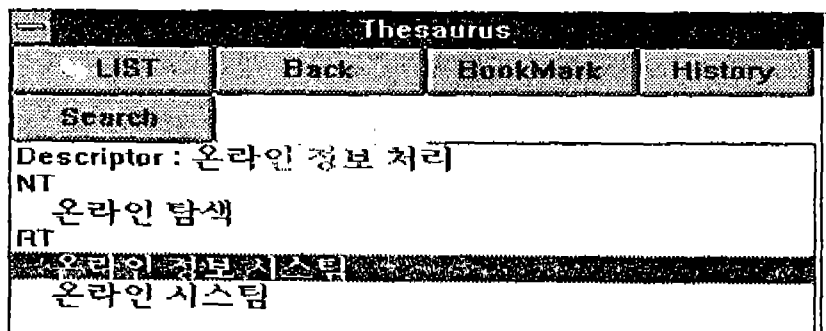
위 <Descriptor List>는 이용자가 원하는 '온라인'이라는 용어가 포함된 모든 디스크립터나 디스크립터로 채택되지 않은 용어들이 디스플레이 되어있다. 이 디스크립터 리스트에서 이용자가 가장 적합하다고 생각되는 디스크립터가 있으면 마우스의 화살표를 그 디스크립터에 맞추어 클릭하거나, 스크롤버튼을 이용하여 적합 디스크립터를 선택한다.

다음 단계로, 선택한 디스크립터가 수록된 디소러스를 찾아보고 싶을 때에는 <그림 10>의 우측 상단의 <Thesaurus>버튼을 클릭하면 된다.

이 화면에서 적절한 디스크립터나 용어를 찾을 수 없을 경우에 이용자 <Back>버튼을 클릭하여 다시 키워드 색인 화일로 돌아가서 다시 적정 용어를 선택하는 과정을 거친다.

4) 디소러스내에서의 질의어 탐색

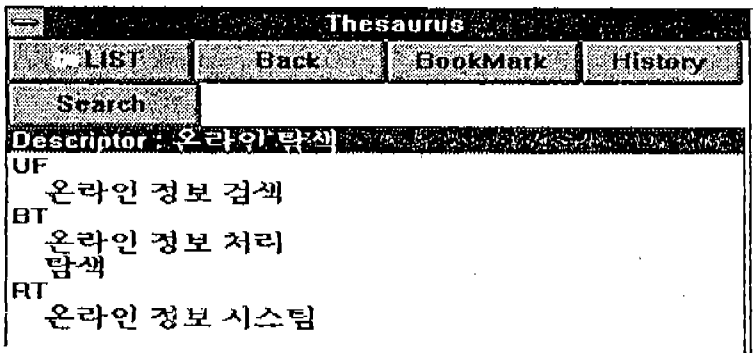
<그림 10>에서의 디스크립터 리스트에서 채택된 용어와 개념적으로 관련된 모든 용어들이 포함된 디스크립터 노드가 <그림 11>에 나타나 있다.



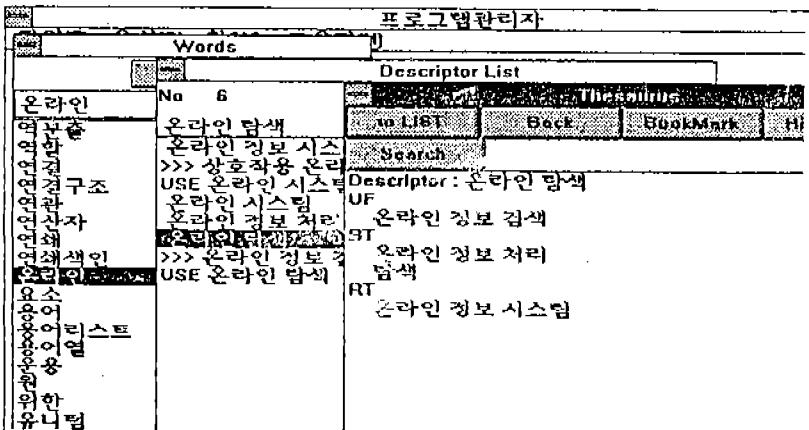
<그림 11> '온라인 정보 처리'의 디스크립터 노드 화면

이 화면내의 모든 용어들은 하나의 시작 노드(anchor node)가 된다. 이 시작 노드로부터 이용자는 이 노드에 연결된 링크들을 따라 관련용어들을 탐색할 수 있다. 또한 디스크립터내에 적절한 용어가 없을때 <List>버튼을 사용하여 다시 디스크립터 리스트로 되돌아갈 수도 있다.

<그림 12>는 <그림 11>의 디소러스에서 선택한 '온라인 탐색'의 디스크립터 노드를 보여준다. 이렇게 하여 디스크립터로 된 여러 노드들을 탐색해 나가면서 이용하는 가장 적합한 탐색용어를 선택할 수 있다.



<그림 12> '온라인 탐색'의 디스크립터 노드 화면

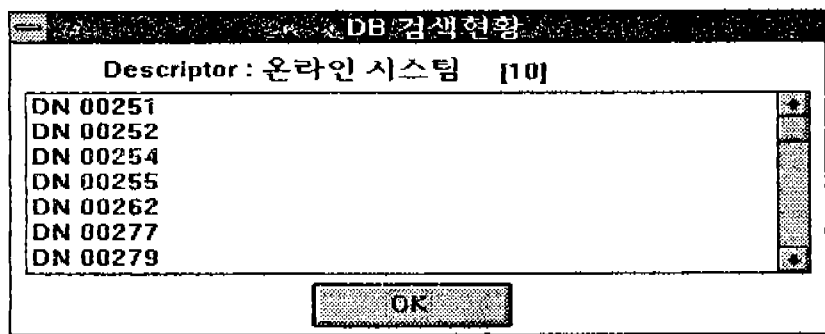


<그림 13> 전체 탐색 과정

이상과 같은 단계들은 <그림 13>과 같이 하나의 연결되는 탐색과정으로 표시된다.

3.5 디소러스와 서지데이터베이스의 연결

이상의 전체 탐색과정을 거쳐 선정된 디소러스내의 탐색어는 곧바로 연결된 서지데이터베이스에서 질의어로 등록되어 문헌검색에 이용 된다. <그림 12>에서 디스크립터를 선택한 후 <Search>버튼을 클릭하면 <그림 14>와 같은 검색화면이 나타난다. 문헌검색은 질의어가 포함된 문헌의 건수를 나타내며, 검색된 문헌번호에 해당되는 서지 사항은 <그림 15>와 같다.



<그림 14> 서지데이터베이스에서 검색된 문헌의 건수와 문헌번호 결과

00251. 이수경

온라인 정보검색시스템의 비교, 분석
 도서관학보(이화여대) 4(1986):9-49
 온라인탐색, 온라인 시스템

00254. 이진희

온라인 정보검색 시스템의 설계에 관한 연구
 연세대산업대학원(석사) 1980
 온라인 시스템, 온라인 탐색

00255. 임인의
온라인 정보검색시스템에 관한 고찰
국회도서관보 20, 5(1983. 07):27-35
온라인 시스템, 온라인 탐색, 정보검색
00262. 정혜순
STN International 온라인 정보검색 시스템
정보관리연구 23, 3(1992. 09):45-73
STN, 온라인 시스템, 온라인 탐색, 정보검색
00277. 한성희
온라인 검색시대의 도래
정보관리연구 17, 4(1984. 12):27-36
온라인 시스템, 정보검색, 온라인탐색
00279. 동래여전학생회 편
우리나라의 온라인 정보봉사;KIETLINE과 DACOMNET를 중심으로
1988:37-66
한국, 온라인 시스템, KIETLINE, DACOMNET, 정보서비스
00280. 설문원
온라인 정보서비스를 위한 교육 및 훈련
연구논집(이화여대대학원) 12(1984. 08):60-107
온라인 시스템, 정보 서비스, 교육
00289. 이수영
온라인 정보 봉사
연구논집(이화여대 대학원) 17(1989):82-113
온라인 시스템, 정보서비스
00303. 최순희
기업체의 정보이용자에 대한 온라인서비스의 만족도에 관한 조사분석
숙명여대 교육대학원(석사) 1991. 08
정보 서비스, 온라인 시스템
00306. 하태중
데이터베이스를 이용한 온라인 정보서비스에 관한 고찰
국립대학도서관보 9(1991. 07):59-71
데이터베이스, 온라인 시스템, 정보서비스

〈그림 15〉 검색된 문헌번호의 서지사항

V. 결 론

정보검색 있어서 자연어 색인과 통제어 색인의 채택 문제는여러가지 장단점을 가지고 있으며 검색결과와 특정성 향상을 위한 자연어 탐색에 대한 연구와 아울러 검색결과와 재현을 향상을 위한 통제어 색인에 대한 연구도 지속적으로 진행되고 있다.

아울러 정보 검색에 있어서 망라성의 향상을 위해서 디소러스의 이용에 대한 연구가 계속 이루어지고 있는 실정이다. 이러한 상황에서 본 연구는 전통적 서지 정보검색 시스템들이 제공하지 못하는 정보탐색 전략의 하나인 브라우징을 지원하기 위해 하이퍼텍스트를 기반으로 하는 디소러스를 탐색 보조도구로서의 역할을 검토하였다. 이에 대하여 정보검색시스템의 인터페이스로서의 실험적 하이퍼텍스트 기반의 디소러스를 개발하여 탐색에 선정의 도구로 사용할 수 있도록 하였다. 시스템의 설계는 디소러스의 디스크립터를 찾기 위한 키워드 색인 파일을 생성시켜 여기에 링크된 디스크립터 리스트에 접근할 수 있도록 하였으며, 그 리스트에서 특정 디스크립터를 선정하여 온라인 디소러스에 접근하도록 하였다. 그 다음 최초의 디스크립터를 시각 노드(anchor node)로 하여 용어간의 관계에 따라 이용자가 원하는 목표 노드로 찾아갈 수 있도록 하였다. 또한 이용자가 이와같은 과정을 거친후 최종적으로 선택한 용어들을 사용하여, 서지 데이터베이스를 검색할 수 있도록 하였다.

본 연구를 통해 얻어진 결론은 다음과 같다.

- ① 본 실험적 하이퍼텍스트 기반의 디소러스는 서지 데이터베이스 검색에서 탐색어 선정을 위한 보조도구로서 사용될 수 있었다.
- ② 디스크립터를 구성하고 있는 용어들에 대한 키워드 색인파일을 생성시켜 브라우징에 의한 적합용어 선택이 효율적으로 이행될 수 있었다.

- ③ 하이퍼텍스트가 브라우징시에 가질 수 있는것 중의 하나인 방향상실 문제를 해결하기 위해 책갈피 버튼과 히스토리(history) 버튼을 제공하였다.
- ④ 하이퍼텍스트 기반의 디소러스를 구축함으로써 인쇄형 디소러스에서는 실행이 어려운 비선형적인 이동이 이루어질 수 있어서 디스크컴퓨터내의 용어간의 관계 확인이 신속하게 이루어질 수 있었다.

<참고문헌은 각주로 대신함>

Experimental Development of Hypertext-based Thesaurus

Jin-Goo Noh *

This study aims to improve subject retrieval by constructing the hypertext-based thesaurus to provide a browsing interface to a thesaurus.

The experimental system used IBM 486 DXII as a hardware, C++ language as a programming language and Hanguk Window 3.1 version as a user interface.

The results of this study are summarized as follows :

① The experimental hypertext-based thesaurus can be used as an efficient search aid for query formulation for the retrieval of bibliographic information.

② The initial access to the hypertext-based thesaurus is via a keyword index. This index is consist of all the words used to form thesaurus terms, whether descriptors or nondescriptors.

③ Hypertext-based thesaurus allow bookmark button and history button to alleviating the problem of disorientation.

④ This system allow an end-user to view a rich variety of inter-term relationships and a complete conspectus of associations through the information space in a nonsequential manner.

* Extra Professor, Dept. of Library and Information Science, Tong-Nae Women's Junior College