

의미네트웍 모델을 이용한 탐색용어선택시스템의 설계 및 구현에 관한 연구

李孝淑*

초 록

본 연구에서는 컴퓨터를 이용하는 탐색시스템에서 탐색용어와 관련된 제한점을 해결하고자 의미네트웍 모델에 의한 지식베이스를 설계하였다. 이것은 탐색전략에서 주제분야의 용어 및 용어간의 어의적 관계에 대한 지식을 확장이용하였을 때 검색의 효율을 높이는 데에 초점을 두었다. 이를 위해서 정보의 탐색과 지식표현에 관한 문헌연구를 하였으며, 실제로 한글문헌의 화학공학분야에서 색인어 실험을 통하여 '유체의 흐름'과 '연소'에 관한 지식베이스를 구성하였다. 탐색방법은 두가지로 구분되었는데 용어에 관한 탐색과 질문식과 관련된 문헌검색으로 실시되었다.

ABSTRACT

It is purposed in this paper to improve the retrieval effectiveness through the use of the semantic knowledge of search terms in a computerbased search system. This study is developed in three stages include the experimentation of index terms on the probabilistic model, indexing with relational operators, and knowledgebase design. The subject experimented is the specific fields of Chemical Engineering, 'Fluid Flow' and 'Combustion'. As for the system implementation, two kinds of search method are done. One is to search terms related to one specialty word, the other is to retrieve the articles based on the queries.

I. 서 론

1. 연구의 목적

정보의 분석과 검색의 문제는 이와 관련된 연구자 및 실무자들에게 있어 지속적인 관심의 대상이 되어 있어 왔다. 최근에 와서는 정보의 소재규명의 단계에서 지식자체에 대한 분석과 활용에 대한 연구가 여러 각도에서 활발히 진행되고 있다.

이러한 추세 속에서 전문사서는 정보서비스

에 있어서 광범위한 사실적 정보를 이해하고 이를 활용할 수 있는 능력이 기대된다. 또한 생산된 지식이 이용자의 요구와 관련하여 최대한의 이용이 되도록 하는 것은 사서의 가장 중요한 임무이기도 하다.

그러나 지식의 양은 방대하고 지식 상호간에서 복잡하게 연관되어 있으므로 이를 체계적으로 조직하여 이용하는데 어려움이 있다. 더우기 가장 적합한 정보를 선정할 때 따르는 여러 가지 불확실성을 배제해야 하므로 이와 관련된

보다 깊이 있는 연구가 필요하다.

본 연구에서는 지식표현기법의 하나인 의미네트웍으로 한글문헌의 화학공학분야에 관한 지식을 표현하였다. 의미네트웍에 의한 지식표현은 사물이나 사실에 대한 여러가지 개념관계를 명확하고 간결하게 표현할 수 있다. 그리고 용어간의 어의지식을 탐색하는 경우와 새로운 관계에 있는 용어를 추가하기에 적합하므로 본 연구를 위해서 선택하였다.

설계된 시스템은 다음과 같은 두가지 목적을 위한 것이다.

첫째, 탐색자가 탐색과정에서 특정 용어에 관련된 개념이나 어휘간의 관계에 대한 지식이 불충분할 때, 이에 관련된 모든 탐색어와 용어간의 관계를 제공하도록 한다.

둘째, 탐색어가 주제분야내에서 여러가지 유형의 관계를 갖게 될 때 그 관계를 구별하여 줌으로써 탐색자가 원하는 내용의 문헌을 검색할 수 있도록 하여 검색효율을 증가시킨다.

2. 연구의 방법 및 범위

컴퓨터를 활용하는 환경에서 정보의 탐색과 지식표현에 관한 제반 문제들을 이론적으로 고찰하였으며, 실제로 학술논문 30건을 선정하여 실험연구를 하였다. 이 연구는 크게 세단계로 구별되었는데, 그것은 실험문헌에서의 주제어 추출과정, 색인작성, 지식베이스 설계 등으로 구성되었다. 주제어의 적합성 판정을 위해서 확률분포를 이용한 색인기법을 사용하였고, 색인어간의 관계는 어의적 관계에 초점을 두고 연관색인의 관계기호로서 표현하였다. 그리고 주제어 및 관련어가 갖는 의미지식을 하나의 네트웍으로 구성하고자 컴퓨터내에 지식베이스를 설계

하였다.

II. 정보의 탐색과 전문가시스템의 응용

1. 지식표현방법의 이용

정보의 탐색과 관련하여 전문가시스템은 실제 도서관업무의 여러 분야에서 응용되고 있으며 그 접근방법은 크게 두가지로 구별된다.¹⁾

첫째, 이용자가 요구하는 문제에 직접 관련이 있는 지식축적의 방안으로 사실적 지식, 경험적 지식을 시스템내에 조직화하여 이용자에게 제공한다.

둘째, 외부의 데이터 및 정보에 대한 중개역할로서 지식의 수집 및 축적은 원거리에 있는 검색시스템에서 수행된다.

전문가시스템의 가장 중요한 특징은 시스템이 지식베이스를 갖는다는 점이다. 수집된 많은 양의 지식이 전문가시스템의 성공적 역할과 밀접한 관련이 있기 때문에 지식의 표현방법은 시스템 설계에 중요한 부분이다.

지식표현기법이란 지적인 추론을 하기 위해서 특정한 데이터 구조를 조작하기 위한 루틴들을 포함한다. 전문가시스템에서 지식을 표현하기 위해서 여러가지 방법들이 사용되고 있는데 가장 일반적인 것으로 생성규칙, 의미네트웍, 프레임 구조 등이 있다.

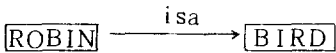
본 연구에서는 탐색용어간에 여러가지 의미관계를 갖는 지식구조를 표현하기 위해서 의미네트웍을 사용하였다. 다음에서 의미네트웍에 의

1) A. S. Pollitt (1984), "A 'front-end' System: an Expert System as Online Search Intermediary," *Aslib Proceedings*, 36, No. 5, p. 230.

한 지식표현기법, 이용방법 (Retrieval and Reasoning) 및 특징에 대해서 간략히 살펴본다.

1.1. 의미네트웍의 기본적 표현방법

의미네트웍은 원래 영어단어의 의미를 표현하기 위해 고안된 것이다.²⁾ 의미네트는 사물, 행위, 상황 등을 표현하는 노드(node:의미점)와 이 노드간의 관계를 나타내는 아크(arc:연결호)로 이루어진다. 만일 영어로 된 문장 All robins are birds 를 의미네트로 표현한다고 가정할 때, robins, birds 를 나타내는 두 개의 노드와 이들간의 관계를 나타내는 아크가만 들어질 수 있다.

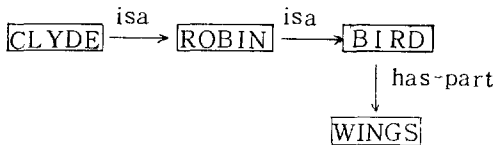


그런데 이때 Clyde를 Robin의 한 종류로서 이 네트에 더 추가하여 표현하고자 한다면



표현될 수 있다.³⁾

또한 의미네트에 의해 계층적 분류를 표현하는 것 외에 사물의 특성에 대한 지식을 표현해야 할 경우가 종종 있는데, 예를 들면 Birds have wings 의 의미를 앞의 예에 더 추가시켜야 할 경우, 그 표현은 다음과 같이 될 수 있다.



이와같은 예에서 Robins have wings and Clyde has wings 이라는 사실을 추론하는 것은 용이하다. 이때 필요한 것은 isa 아크를 추적해 나가는 것이며, 상위 계층의 노드에 대해서도 적용될 수 있어서 따로 이같은 사실을 네트웍

에 표현하지 않아도 된다. 인공지능분야의 용어로서 이런 종류의 추론은 특성계승성(Property Inheritance)이라고 하며 isa 아크는 특성계승링크(Property Inheritance Link)라고 한다.⁴⁾

1.2. 의미네트웍으로 표현된 지식의 이용

의미네트웍을 활용하기 위해 추론은 여러가지 전략으로 적용되고 있다.⁵⁾ 의미네트구조의 능력은 컴퓨터내의 지식표현에 사용되는 다른 기법과 마찬가지로 문제를 풀기 위해 이들을 처리하는 프로그램의 능력에 의해 결정된다. 의미네트구조를 이용한 초기의 한 방법으로는 두 노드에서 동시에 함께 탐색을 시작하여 서로 만나는 곳을 조사함으로써 사물간의 관계를 찾아내는 방법이 있는데, 이것이 교차탐색(intersection search)에 의한 방법이다.⁶⁾

지식의 표현형식과 지식의 이용과 관련하여 고려되어야 할 의미네트웍의 특성은 다음과 같다.⁷⁾

- 사물, 사실, 속성, 행위 등을 나타내고자 할 때 의미네트모델에 의한 표현은 노드와 아크의 타입이 미리 정의되며, 노드간의 관계는 직접 표현된다.
- 어떤 일을 어떻게 처리하는가에 대한 지식을 나타내거나 사실에 대한 상대적 중요성을 표

2) 유석인, 전주식, 한상영(1986), 「人工智能」, (서울: 尙潮社), p.229.
 3) E.A. Feigenbaum (1981), The Handbook of Artificial Intelligence, (California: William Kaufmann Inc.), 1, p.181.
 4) 앞 글, p.182.
 5) 김재희, 박충식(1986), “인공지능의 지식표현 방식에 대한 비교 고찰,” 정보과학회지, 4, No.3, pp.31-40.
 6) 유석인, 전주식, 한상영(1986), p.231.
 7) 김재희, 박충식(1986), p.39.

현하기가 어렵다.

- 새로운 데이터와 기존의 데이터 사이에 모순이 있는지 쉽게 찾을 수 있다.
- 지식베이스 검색에서 하나의 데이터 구조가 다른 구조와 관계되어질 때 연결이 잘 이루어지며 유추나 귀납에 의한 추론이 가능하다.

2. 선 행 연 구

탐색중개시스템으로서 전문가시스템의 응용은 두가지로 구별된다.⁸⁾ 첫째는 온라인 목록에서 이용자가 찾는 문헌의 소재를 알려주기 위한 인터페이스의 제공이다. 검색되는 내용은 대개의 경우 단행본이 된다. 그리고, 둘째는 서지 데이터베이스의 탐색을 위한 질문식 작성에 도움을 주는 시스템이다. 검색되는 내용은 단행본 뿐만 아니라 학술잡지의 기사도 포함된다. 이 경우 이용자에게 탐색하여야 할 데이터베이스 뿐만 아니라 탐색을 위한 질문어와 관련된 정보도 제공한다. 이것은 곧 온라인 탐색중개시스템(Online Search Intermediary System)이라고 할 수 있다. 이와 관련하여 페이스(Paice)는 이상적인 탐색중개시스템이 갖추어야 할 특징들을 제시한 바 있다.⁹⁾ 영국 랑카스터(Lancaster) 대학에서는 탐색중개시스템을 개발하여 실행단계에 있으며 탐색용어와 관련된 실험적 시스템들이 연구된 바 있다.¹⁰⁾

이용자요구에 대한 정보서비스를 보다 효율적으로 하기 위해서도 전문가시스템이 응용되고 있다. 미국 메리랜드(Maryland) 주의 국립농학도서관(The National Agricultural Library)에서는 즉답형 참고질문에 대해서 이용자가 해답을 얻도록 도와주는 시스템을 개발하였다.¹¹⁾ 이 시스템은 기존의 소프트웨어패키지를 이용하여 특정형 질문에 해답을 제공한다. 또한 외부 데이터뱅크에도 연결이 되어 서지인용 데이

터베이스와 원문화일의 온라인 탐색이 가능하도록 하였다.¹²⁾

이외에도 정부간행물에 대한 이용자질문에 봉사하고 사서의 생산성을 강화하기 위하여 시스템이 개발되었다.¹³⁾ 이 시스템설계를 위하여 뉴욕주립대학(State University of New York) 내에 록우드(Lockwood)도서관의 정부간행물 부서에서 이용자 질문과 참고자료의 리스트가 수집 및 분석되었다.

3. 탐색활동에 이용되는 기억구조

탐색기술이란 수작업 또는 자동화된 정보시스템에서 탐색자에 의해 활용될 수 있는 방법, 경험, 기법, 전략, 계획 등을 의미한다.¹⁴⁾ 서지적

8) C. Paice (1986), "Expert Systems for Information Retrieval?" Aslib Proceedings, 38, No. 10, pp. 347-348.

9) 앞 글, p. 350.

10) Oddy의 THOMAS 프로그램과 Shoal의 탐색용어자문시스템이 대표적인 예이다. P. Shoal (1983), "Knowledge Representation in Consultation Systems for Users of Retrieval Systems," The Application of Mini & Micro computers in Information Documentation and Libraries, (Amsterdam: Elsevier Science Pub.), pp. 631-643.

11) S. T. Waters (1986), "Answerman, The Expert Information Specialist: An Expert System for Retrieval of Information from Reference Books," Information Technology and Libraries, 5, No. 3, pp. 204-211.

12) DIALOG, BRS Online Systems, AGRICOLA 데이터베이스 및 미국 화학회에서 발행되는 18종의 잡지 원문을 포함한다. 이용자의 질문과 관련하여 정의, 엔트리에 대한 설명, 사례 제공, 해결 등의 4가지 단계로 구성되었다.

13) Karen, F. Smith (1986), "Robot at the Reference Desk?" College & Research Libraries, 47, No. 5, pp. 486-490.

14) M. T. Bates (1981), "Search Techniques," ARIST, 17, p. 139.

탐색에 관한 모든 면을 포함하며, 그 연구범위는 상당히 넓다.

탐색자로서 사서는 이용자의 질문과 관련된 정보원을 탐색하는 과정에서 그가 기억하고 있는 여러가지 사실들을 참고하는 경우가 종종 있다. 그리고 각각의 사실들을 조직하기도 하는데, 이러한 과정에서 활용하고 있는 기억구조의 유형과 그 내용을 살펴보면 다음과 같다.¹⁵⁾

3.1. 어의적 기억 (Semantic Memory)

어의적 기억은 언어사용을 위해 필요한 기억으로, 이것은 지적인 디소러스 (mental thesaurus) 와 같아 사람이 단어에 대해 또는 구두에 의한 상징기호에 대해서 갖는 조직화된 지식이다. 단어에 대한 기억구조에는 단어의 의미, 대상, 단어간의 관계, 규칙, 공식, 단어를 조직하는 알고리즘, 개념관계 등이 포함된다.

3.2. 삽화적 기억 (Episodic Memory)

삽화적 기억이란 일시적인 사건이나 에피소드에 대한 정보를 사람의 두뇌에서 받아들여 축적한 것으로 이러한 사건은 시간과 공간적 관계에 대한 내용들이 된다. 삽화적인 기억은 탐색자의 특정한 경험을 상세히 회상하여서 요구된 질문과 연관을 지어보는 기억이다.

3.3. 스키마 기억 (Schematic Memory)

스키마타 (Schemata) 는 다른 기억보다도 더 자주 활용되는 기억구조이다. 이것은 사서가 탐색도구로서 자료의 종류, 자료내용의 범위, 배열, 크기, 유용성 등에 대하여 갖는 기억이다.

이러한 기억구조는 과거의 경험을 기초로 형성되며 사물, 장면, 사건에 대한 무의식적인 일련의 예측 등으로 구성된다. 이것은 어의적 기억과는 구별된다. 스키마타는 공간적 또는 시간적으로 조직된 구조를 갖고 그 유사성에 기초

하여 서로 연관이 된다.

탐색에서 실패의 주요 원인은 불완전하거나 적합하지 않은 탐색전략이 적용된 경우 또는 질문이 명확하지 않을 때 비롯된다. 문제를 구체적으로 표현하는 것은 어휘이다. 탐색에서 가장 중요한 점은 정보요구의 명확한 의미를 파악하는 것이다.¹⁶⁾ 그리고, 탐색자는 질문의 초기 단계에서 사용된 어휘나 문장에 너무 집착할 필요는 없다. 또한 탐색자가 용어에 대해서 명확하게 이해하지 못할 때에는 탐색도구로서 자료나 장치를 효과적으로 활용하여야 한다.

4. 탐색키의 선택과 탐색기법

4.1. 탐색키의 선택

탐색과정을 大別하여 볼 때, 세가지 지적인 활동으로 구성되는데, 그것은 질문어에 대한 정의, 탐색키의 선택, 피이드백에 의한 평가 등이다.¹⁷⁾ 이중에 탐색키의 선택은 탐색의 방향과 결과에 직접적인 영향을 주는 요인으로서 선택과 관련된 가능한 경우들을 살펴보는 것은 실제적 탐색에 도움이 된다.

후델 (Fidel) 은 전문탐색자에 의한 탐색활동에서 탐색키 결정에 관련된 선택과정과 탐색조건과의 관계를 연구하였다.¹⁸⁾ 이 연구를 통해서 알 수 있는 점은 온라인탐색자의 탐색키 선택과정은 결정구조 (decision tree) 로서 시스템설계를 할 수 있다는 점이다. 그리고 탐색시스템설계

15) S.D. Neill(1984), " The Reference Process and Certain Types of Memory : Semantic, Episodic, and Schematic," RQ, 23, No. 4, pp.417-423.

16) S.D. Knapp(1978), " The Reference Interview in the Computer-based Setting," RQ, 17, No. 4, p.323.

17) Raya Fidel(1986), " Towards Expert Systems for the Selection of Search Keys," JASIS, 37, No. 1, pp. 37-44.

18) 앞 글

에서 고려되어야 할 몇가지 사실들이 나타났는데 그것은 다음과 같다.

첫째, 탐색기를 선택하는 결정은 탐색의 결과와 밀접한 관계가 있다. 그러므로 탐색시스템에서 탐색과 관련된 기타의 다른 요소들이 설계되기 이전에 용어상의 문제가 해결되어야 한다.

둘째, 전문어와 비전문어 간의 구별은 매우 중요하다. 왜냐하면 용어가 비전문어일 때 탐색자는 자연어로 탐색기를 선택하는 경우는 드물고, 대부분 디스크립터를 선택한다. 디스크립터 중에 어느것도 원래의 개념과 정확히 매칭되지 않는 경우에도 이같은 사실은 마찬가지이다. 그러므로 비전문어에 대한 통제가 필요하다.

세째, 탐색에서 중요한 용어들의 속성이 더 연구되어야 한다. 이것은 통계적 방법을 사용할 수도 있는데, 예를 들어 디스크립터의 신뢰도는 용어가 디스크립터로 주어진 일관성 정도에 따라 측정될 수 있다.¹⁹⁾

4.2. 탐색 기법

탐색기법은 서지탐색과 참고탐색을 포함하며 수작업 및 온라인시스템에 모두 적용된다. 탐색기법은 심층적 탐색을 요하는 질문으로 탐색의 여러 단계가 필요하거나 또는 정보검색시스템의 검색결과에 대해서 더 만족스러운 정보를 찾는 경우에 필요하다.

일반적으로 탐색기법은 4가지의 유형으로 구별된다.²⁰⁾ 그중에 용어의 탐색기법으로 현재 사용되고 있는 방법을 간략히 살펴본다.^{21) 22)}

- 상위개념어 : 용어상에서 계층적으로 더 상위개념으로 확장시키는 기법이다. 탐색용어를 수정하기 위해 탐색자는 디소리스내의 포인터가 참고가 되거나 또는 탐색자 자신의 지식에 의존하기도 한다.

- 하위개념어 : 계층적으로 더 하위개념의 용

어로서 세부적 또는 종속적인 개념으로 탐색하는 기법이다.

- 관련어 : 용어를 조정하기 위하여 계층적으로 동등한 위치에 있는 용어들에서 탐색어를 선택하고 조정하는 기법이다.

- 유사어 : 인접한 용어 또는 관련된 용어로서 유사한 어휘의 표현이나 관련된 주제를 나타내는 용어들을 조사하여 탐색어휘를 추가시키는 기법이다. 유사어기법을 이용하는 것은 유사한 기입을 조사함으로써 탐색을 확장시킬 수 있기 때문이다.²³⁾

- 추적 : 더 확장된 탐색을 위해 탐색과정에

19) King 과 Bryant 는 용어의 일관성 측정은 디스크립터의 신뢰도 측정에 이용될 수 있다고 하였다.

Raya Fidel(1986), p.43 에서 재인용

20) 조정기법 (monitoring tactics): 탐색이 연속적이고 효율성을 유지하도록 하는 기법.

화일구조기법 (file structure tactics): 정보시스템에서 여러가지 방법의 화일구조를 이용하여 정보를 탐색하는 기법.

탐색식 기법 (search formulation tactics): 탐색식을 작성하거나 또는 재구성하는데 사용되는 기법.

용어탐색기법 (term tactics); 탐색식에서 특정 용어의 선정과 수정에 활용이 되는 기법.

M. J. Bates(1979), "Information Search Tactics," JASIS, 30, No. 4, pp. 205-214.

21) 앞 글, pp.211-212.

22) D. Fairhall은 Bates 의 구분과는 다르게 탐색기법을 구분하였다. 탐색자의 탐색과정을 우선 4단계 (locating, broadening, focusing, selecting)로 구분하였다. 그리고 이에 대해서 관련된 이제까지 연구된 기법들을 연관지었다.

D. Fairhall (1985), "In Search of Searching Skills," J1. of Information Science, 10, pp.112-114.

23) ORBIT 시스템에서는 유사어탐색을 위한 사전화일의 탐색명령어로 NBR(Neighbor 의 약어)

서 이미 찾은 용어들을 조사하는 방법이다.

- 대체 : 탐색어를 여러가지 방법으로 대체하거나 변경하는 방법이다. 구체적인 수정방법은 다음의 용어기법들을 통해서 적용된다.

- 어간조정 : 접두사, 접미사 등의 접사를 바꾸어 보는 기법으로서 절단기법을 통하여 여러가지 방법으로 이용될 수 있다.

- 재배열 : 시스템에서 용어들이 한 단어 이상으로 구성되었을 때 단어의 순서는 검색결과에 차이를 가져온다. 재배열은 탐색용어에서 단어의 순서를 일부 혹은 전부를 재배열하거나 전환하는 방법이다.

- 반의어 : 원하는 정보를 기술하는데 있어 논리적으로 반대가 되는 용어로서 탐색하는 기법이다.

- 철자 : 탐색용어에 대해서 철자상의 변화를 주어 탐색하는 방법이다. 즉, 철자의 정확한 일치에 중점을 둔 것이 아니고 검색의 재현을 증가시키기 위한 것이다.²⁴⁾

- 빈칸 : 연자기호(hyphen)가 포함된 단어나 빈칸을 갖는 단어는 탐색결과에 영향을 준다. 빈칸을 주는 문제는 단어 사이에 빈칸을 서로 다르게 주어 탐색을 하는 것으로서 이것은 자동화된 시스템에서 중요한 문제가 된다. 두가지 기본원칙이 적용되는데 단어별(word-by-word), 문자별(letter-by-letter) 방식으로 화일구조내에서 빈칸을 준다. 따라서 탐색자는 탐색에 사용될 화일규칙에 대한 이해가 있어야 한다.

이상에서 설명된 탐색기법은 탐색환경이나 탐색목적에 따라서 적절히 그룹별로 모아서 활용될 수 있다. 그리고 이와같은 기법외에도 이용자의 피이드백과 관련된 기법, 탐색자의 경험적 규칙에 의한 기법 등이 고려될 수 있다.

III. 한글문헌의 색인어에 의한 시스템 베이스 설계

1. 실험주제 및 대상문헌

실험대상 주제는 화학공학분야의 세부 주제로서 ‘유체의 흐름’과 ‘연소’에 한정하였다.²⁵⁾ 실험대상문헌은 한국화학공학회에서 계간으로 발행되는 학술논문지인 화학공학에서 1977~1986년 까지 10년동안 발표된 논문으로 하였다.

2. 데이터의 수집

주제분야의 문헌집단에서 단어의 분포 현상 및 색인어선정 실험을 위해서 실험규모는 30건의 논문으로서 ‘유체의 흐름’과 ‘연소’분야에서 각각 15건으로 구성되었다.

각 논문에서 초록의 길이는 평균단어수 41개로서, 최대단어수 65개, 최소단어수 22개로 구성되었고, 단어의 추출과정에서 각 논문의 표제는 모두 포함시켰다. 각 논문당 18개의 단어를 추출하여 중복출현된 단어를 제외한 총 299개어가 표집크기로 결정되었다. 선행연구에서 단어유형이 358²⁶⁾, 350²⁷⁾ 개이었으므로 이에 준하고자 하였으나 초록의 길이가 짧은 논문이 3건 있었고²⁸⁾

24) 이것은 조정기법 영역에 속하는 CORRECT 기법과는 구별된다.

25) 주제의 구분은 Chemical Abstract: General Subject Index(1986)에서 화학공학부문에 대한 구분에 의한 것이다. 연소에 관련된 논문은 10년 동안에 발표된 총 논문수에 해당한다.

26) Harter, S.P.(1975.a,b), "A Probabilistic Approach to Automatic Keyword Indexing," JASIS 20, No.4, pp.197-206, 280-289.

27) 이태영, 정영미 (1982), "자동색인의 통제적 기법과 한국어 문헌의 실험," 도서관학, 제9집, pp.99-118.

28) 문헌당 단어의 수를 동일하게 추출하였고, 초록에 포함된 단어수가 22개, 24개의 논문이 3편이 된다.

중복출현된 단어를 삭제하는 과정에서 단어수가 줄었다.

3. 실험환경

문헌집단에서 색인어의 출현빈도에 의한 각 파라미터 값 산출은 VAX-11/780/VMS 로 PASCAL 언어를 사용하였다. 그리고 지식베이스 설계와 탐색은 IBM 5550 을 사용하였고, 프로그램 언어는 PASCAL 로 작성되었다.

4. 복수포아슨분포모형에 의한 색인어 실험

한 문헌에서 색인어로 유용한 주제어를 밝혀내는 것은 주제색인을 작성하는데 있어서 중요한 과제이다. 따라서 문헌내용에서 주제어의 출현현상이나 빈도수에 관하여 통계적 특성을 밝혀내고자 많은 연구자들은 여러가지 시도를 하였다.²⁹⁾

확률색인에서 단어는 일군의 문헌내에 분산적으로 분포되는 것으로서 포아슨분포함수로 설명된다. 즉, 한문헌에서 어떤 단어가 K번(K tokens) 출현할 확률 f(k)는 다음과 같은 방식으로 나타낼 수 있다.

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

이때 λ는 일군의 문헌집단에서 어떤 단어가 출현하는 평균수를 의미한다.

그런데 이같은 포아슨분포함수의 현상을 따르는 것은 비주제어가 갖는 특성임이 하터의 실험결과 밝혀졌다.³⁰⁾ 반면에 주제어는 이 함수모형을 따르지 않았으며, 주제어가 문헌에서 출현하는 분포현상으로 복스타인, 스완슨, 하터 등은 복수포아슨분포모형(2-Poisson Distribution)을 채택하였다.

복수포아슨분포모형은 다음과 같다.³¹⁾

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

문헌의 특정주제개념을 나타내는 주제어는 문헌집단에서 분산적으로 출현하지 않고 특정주제와 관련된 문헌에서 집중적으로 출현한다는 이론하에, 복수포아슨분포모형이 화학공학문헌의 주제어 분포현상 규명에도 적합할 것이라는 가설을 실험하였다.

표본문헌에서 단어의 실제 출현빈도수에 의해 λ₁, λ₂, π의 값을 299 개어에 대해서 계산하였으며, 그 가운데 예로 20개 단어의 문헌분포양상을 보면 표<1>과 같다.

<표 1> 20개 단어의 빈도 분포

빈도	단어유형	k 토큰을 갖는 문헌수														
		k	1	2	3	4	5	6	7	8	9	10	11	15		
6	교 반														1	
7	비뉴튼성															
5	슬러리															
9	액막															
22	유체															
20	가스화															
13	무연탄															
4	생성물															
43	연소															
4	전환율															
36	유동층															
4	유출															

29) M.Marion, R.Curtice, P.Jones 등은 문헌집단에서 출현되는 빈도에 의해 단어의 색인어로서 유용성을 판단하기 위해서 그들의 가설을 실험하였으며, A.Bookstein 과 D.R.Swanson 은 주제어의 집중적 출현현상을 통계적 특성에 기초하여 분석하였다.

30) S.P.Harter (1975 a), p.198.

31) 앞 글, p.200.

빈도	단어유형	k 토큰을 갖는 문헌수															
		k	1	2	3	4	5	6	7	8	9	10	...	15			
4	연료	1	1														
9	재연소	4				1											
7	조업	3			1												
8	관	1	1	1													
9	기체	2	1		1												
8	지역									1							
7	자유	1	1	1													
4	난류	1	1														

표집된 단어에 대해서 각각의 문헌과 문헌집단에서의 출현빈도에 의해 파라미터 값을 계산하였다. 계산과정은 다음과 같다.³²⁾

① 변수를 설정한다.

$$\lambda = \frac{R}{A}$$

- λ : 단어의 평균 출현빈도
- R : 단어의 총 출현빈도
- A : 단어가 출현하는 문헌수
- λ_1 : 특정개념을 중심주제로 하는 문헌 집단 I에서의 평균 출현빈도
- λ_2 : 특정개념을 주변주제로 하는 문헌집단 II에서의 평균 출현빈도
- π : 특정개념을 다룬 문헌이 문헌집단 I에 속하는 문헌의 비율

② 표본문헌에서 실제의 값을 구하여 복수포아손분포의 모멘트 산출공식에 대입한다.

- $\lambda_1, \lambda_2, \pi$ 의 실제의 값 계산
- $R_1 = \pi \lambda_1 + (1 - \pi) \lambda_2$
- $R_2 = \pi(\lambda_1^2 + \lambda_1) + (1 - \pi)(\lambda_2^2 + \lambda_2)$
- $R_3 = \pi(\lambda_1^3 + 3\lambda_1^2 + \lambda_1) + (1 - \pi)(\lambda_2^3 + 3\lambda_2^2 + \lambda_2)$

③ $M=R_1, L=R_2 - R_1, K=R_3 + 2R_1 - 3R_2$ 으로 하여 $M = \pi \lambda_1 + (1 - \pi) \lambda_2, L = \pi \lambda_1^2 + (1 - \pi) \lambda_2^2, K = \pi \lambda_1^3 + (1 - \pi) \lambda_2^3$ 의 식이 유

도된다.

④ $a = M^2 - L, b = K - LM, c = L^2 - MK$

일때 $a\lambda^2 + b\lambda + c = 0$ 의 방정식에서 두근이 λ_1, λ_2 가 된다.

π 는 ③에서 $\frac{M - \lambda_2}{\lambda_1 - \lambda_2}$ 가 되므로 색인어에 대

$\lambda_1, \lambda_2, \pi$ 의 값이 추정된다.

⑤ Z값은 $Z = \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1 + \lambda_2}}$ ³³⁾에서 λ_1 과 λ_2

의 값을 각각 대입하여 구한다.

표본문헌에 의해서 계산된 $\lambda_1, \lambda_2, \pi$ 의 값을 표<1>에 단어들에 의해서 보면 표<2>와 같다.

<표 2> 20개 단어의 $\lambda_1, \lambda_2, \pi$ 의 값

단어유형	λ_1	λ_2	π
교반	$\frac{6}{2}$	0	$\frac{2}{30}$
난류	$\frac{4}{2}$	$\frac{5}{1}$	$\frac{2}{30}$
비뉴턴성	$\frac{7}{4}$	0	$\frac{4}{30}$
슬러리	$\frac{5}{2}$	0	$\frac{2}{30}$
액막	$\frac{9}{2}$	0	$\frac{2}{30}$
유체	$\frac{22}{9}$	0	$\frac{9}{30}$
가스화	$\frac{20}{5}$	0	$\frac{5}{30}$
무연탄	$\frac{13}{7}$	0	$\frac{7}{30}$
생성물	$\frac{4}{3}$	0	$\frac{3}{30}$

32) 앞 글, pp.202-203.

33) 문헌집단 I과 문헌집단II의 분포함수의 평균인 λ_1 과 λ_2 의 차이가 클수록 색인어로서의 가치는 크다. 따라서 색인어로서의 적합성 판정에 두 문헌집단 간의 중복도를 측정하는 Z값이 이용되었다. S.P.Harter (1975 a), p.204.

단어유형	λ_1	λ_2	π
연 소	$\frac{43}{11}$	0	$\frac{11}{30}$
전 환 율	$\frac{4}{2}$	0	$\frac{2}{30}$
유 동 총	$\frac{36}{7}$	0	$\frac{7}{30}$
유 출	$\frac{4}{1}$	0	$\frac{1}{30}$
연 료	$\frac{4}{2}$	0	$\frac{2}{30}$
재 연 소	$\frac{9}{5}$	0	$\frac{5}{30}$
조 업	$\frac{7}{4}$	0	$\frac{4}{30}$
관	$\frac{8}{3}$	0	$\frac{3}{30}$
기 체	$\frac{9}{4}$	0	$\frac{4}{30}$
지 역	$\frac{8}{1}$	0	$\frac{1}{30}$
자 유	$\frac{7}{3}$	0	$\frac{3}{30}$

이와 같은 실제 데이터의 값은 복수포아슨분포 모형에 의해 계산되었고, 예로서 20개 단어에 대한 값은 표 < 3 > 과 같다. 그리고 하터의 실험에서 색인어 적합성 판정기준으로 제시되었던 Z 값을 산출하였다.

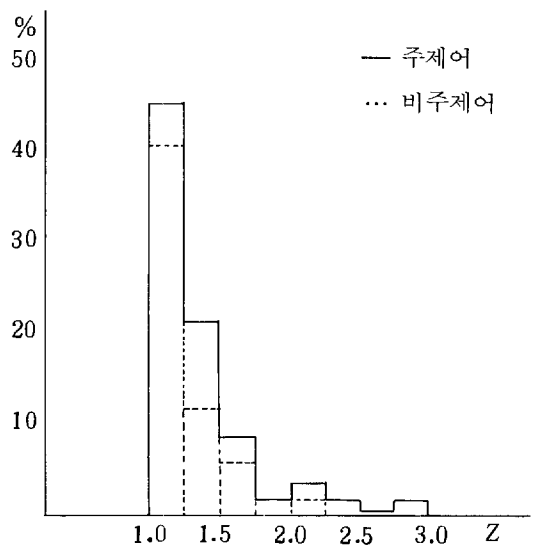
< 표 3 > 20개 단어의 $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\pi}, Z$ 의 값

단어유형	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\pi}$	Z
교 반	3.000000	0.000000	0.066667	1.732051
난 류	35.804173	43.598648	0.609533	D < 0 *
비뉴턴성	1.750000	0.000000	0.133333	1.322876
슬 러 리	2.500000	0.000000	0.066667	1.581139
액 막	4.500000	0.000000	0.066667	2.121320
유 체	2.444445	0.000000	0.300000	1.563472
가 스 화	4.000000	0.000000	0.166667	2.000000
무 연 탄	1.857143	0.000000	0.233333	1.362770
생 성 물	1.333333	0.000000	0.100000	1.154701
연 소	3.909091	0.000000	0.366667	1.977142
전 환 율	1.999999	0.000000	0.066667	1.414213

단어유형	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\pi}$	Z
유 동 총	5.142857	0.000000	0.233333	2.267787
유 출	4.000000	0.000000	0.033333	2.000000
연 료	1.999999	0.000000	0.066667	1.414213
제 연 소	1.000000	0.000000	0.033333	1.000000
조 업	1.750000	0.000000	0.133333	1.322876
관	2.666667	0.000000	0.100000	1.632993
기 체	2.250000	0.000000	0.133333	1.500000
지 역	8.000000	0.000000	0.033333	2.828427
자 유	2.333333	0.000000	0.100000	1.527525

* $D = \sqrt{\lambda_1 + \lambda_2}$

실험된 단어 299개어의 Z치에 따라서 주제어, 비주제어³⁴⁾의 분포현상을 도시하면 그림 < 1 > 과 같다.



< 그림 1 > 주제어, 비주제어의 Z치 분포

34) 주제어, 비주제어의 구분은 색인어가 디스크립터로 선정된 경우나 주제분야의 용어사전에 수록된 경우를 주제어로 구분하였다.

The American Chemical Society (1985), *Chemical Abstracts Index Guide part I, II* (Ohio: The American Chemical Society), 한국화학공학회 (1987), 「화학공학 술어집」 (서울: 한림상사) 등을 참고로 하였다.

5. 실험결과의 분석

그림 1 과 같이 나타난 결과를 분석해 보면 다음과 같다.

Z치가 1.5 이상인 경우에 절대다수가 주제어인 것은 하터의 심리학분야의 실험과 이태영, 정영미의 농학분야의 실험결과와 유사하다.³⁵⁾

그리고 Z=1 을 한계치로 했을 때 Z ≥ 1 에 포함되는 주제어는 전체 주제어의 74.4 % 에 해당하고, 비주제어는 약 28.2 %로서 재현율 74.4 %, 정확율 71.8 %로 예상할 수 있다. 이것은 한글문헌의 농학분야에서 66 %의 재현율, 70 %의 정확율과 비교할 때 정확율의 기대치가 유사한 결과이다. 반면에 Z치가 0 이하의 값으로서 이 중에 주제어는 25.6 %를 차지하고 있다. 이것은 심리학분야의 주제어중에 0.5 이하의 값이 약 25 %인 것과, 농학분야에서 25 %인 것 등 비율에 있어서는 유사하다. 그러나 선행실험에서 Z값의 기준이 0.5인데 비해 본 실험에서는 0 ≤ Z < 1.0 인 값은 나타나지 않고 Z < 0 을 기준으로 한 점이 상이한 결과이다.

6. 주제어의 색인작성

‘유체의 흐름’ 과 ‘연소’ 에 관련된 주제 개념들을 색인으로 작성하였다. 색인된 내용은 35 개의 주제어³⁶⁾ 와 관련하여 기술된 표본문헌내의 모든 문장들을 초록과 표제로부터 추출한 것으로 단어수 255 개어로 구성되었다. 색인작성에서 색인어간의 관계³⁷⁾는 연관색인의 9 가지 구분영역에 기초하였으며 실제 작성방법도 연관색인에 의한 것이다. 사용된 9 가지의 관계기호와 의미관계는 표 < 4 > 와 같다.

실제로 작성된 색인으로서 유동층과 석탄에 관한 내용을 예시하면 표 < 5 > 와 같다.

< 표 4 > 관계기호 및 의미관계

기 호	개 념 영 역
/0	동시인지 관계 공존, 일치, 미래의 행위, 시간 개념이 포함되기 이전의 관계
/=	동등성 관계 동등, 동의어, 고유명사, 상품명, 유사관계
/)	명료성 관계 모방, 대치, 상대적차이, 모델
/*	자활성 관계 자동차의 보어로서의 행위 여격, ~을 통하여
/+	차원 관계 시간, 공간적 위치, 변하기 쉬운 일시적 성질, 가변적 성질, 온도, 수, 무게, 양, 농축, PH, 비율, 속도
/-	행위 관계 자동차의 목적어로서의 행위 사물이나 행위에 영향을 미치는 일이나 행동
/;	연상 관계 추상적, 간접적 성질, 과거의 행위
/(<	종속 관계 전체와 부분관계, 사물의 고유한 성질, 물리적 성질, 종속관계
/:	기능상 의존성 관계 기능상 상호의존, 인과관계

35) 하터의 실험에서는 Z ≥ 1.5 인 경우에 있는 단어는 모두 주제어에 해당하였다.

36) 주제어 판정기준으로 Z 치 1.5 이상을 한계치로 하여 1.5 이상의 값을 갖는 단어들이 채택된 범위이다.

37) 진조합 용어열색인에서는 다양한 구분적 장치를 사용하고 있다. 개념관계를 표시하기 위해 대표적인 것으로 관계기호가 있으며, 연구된 시스템으로는 페리데인의 연관색인과 Syntol 시스템이 있다.

< 표 5 > 색인작성의 예

석탄 / : 가스화 / : 생성물 / ; 공기 / +유량 /
 석탄 / : 가스화 / : 생성물 / ; 수증기 / +유량 /
 석탄 / : 가스화 / : 생성물 / ; 온도 / ; 반응 /
 석탄 / : 가스화 / +유동층 / +상압
 석탄 / (준강점탄 /
 석탄 / : 전처리 / +반응기 / +유동층 / +상압 /
 석탄 / : 가스화 / +반응기 / +유동층 / +상압 /
 석탄 / : 가스화 / +유동층 / +상압 /
 석탄 / : 가스화 / : 생성물 / ; 공기 / 유량 /
 석탄 / : 가스화 / : 생성물 / ; 수증기 / +유량 /
 석탄 / : 가스화 / : 생성물 / ; 온도 / 반응 /
 석탄 / : 가스화 / : 생성물 / +유동층 / +상압 /
 가스화 / : 생성물 / (일산화탄소 / +농도 /
 가스화 / : 생성물 / (수소 / +농도 /
 가스화 / : 생성물 / (일산화탄소 / +상대비 /
 가스화 / : 생성물 / (이산화탄소 / +상대비 /
 가스화 / (석탄 / +전환율 /
 석탄 / +전환율 / ; 회분 /
 반응 / (가스화 / 석탄 / (입자 , 단일 /
 반응 / (가스화 /) 모사 , 수식 /
 반응 / (가스화 / : 석탄 / (무연탄 / (입자 , 단일 /
 반응 / (가스화 /) 모사 , 수식 /
 조업 / +변수 / ; 석탄 / -공급 / +양 /
 조업 / +변수 / ; 석탄 / (입자 / +크기 /
 조업 / +변수 / ; 공기 / * 유동화 / +속도 /
 조업 / +변수 / ; 온도 / +유동층 /
 연소 / +효율 / ; 증가 /
 연소 / +효율 / ; 석탄 / (입자 / +크기 /
 연소 / +효율 / ; 공기 / +비율 /
 유동층 / ; 특성 , 수역학적 /
 유동층 / * 순환 /
 연소 / : 석탄 /
 연소 / ; 특성 /

흐름 / : 석탄 / +유동층 / * 순환 /
 흐름 / : 공기 / +유동층 / * 순환 /
 흐름 / -측정 /
 연소 / : 석탄 /
 연소 / ; 효율 /
 연소 / ; 온도 / +분포 / +유동층 /
 석탄 / : 흐름 / ; 특성 /
 석탄 / +입자경 /
 석탄 / : 연소 /) 실험 /
 석탄 / +입자 /

7. 지식베이스 설계

색인어에 대한 의미네트웍이 컴퓨터내에서 구성되었으며 데이터 구조로 링크드리스트(linked list) 구조와 배열(array)을 사용하였다. 총 255 종류의 노드와 한 노드당 링크수는 최소 1 종류에서 최대 8 종류로 구성되었다. 그리고 255 개의 색인어와 해당문헌은 도치화일로서 작성되었으며 표 < 6 > 과 같다. 이것은 노드 탐색시 컴퓨터내의 주소를 찾는 경우와 의미네트웍을 이용한 문헌검색 효율과 비교하기 위해서 만들어졌다.

컴퓨터내의 의미네트웍을 구성하기 위해 작성된 알고리즘은 다음과 같다.

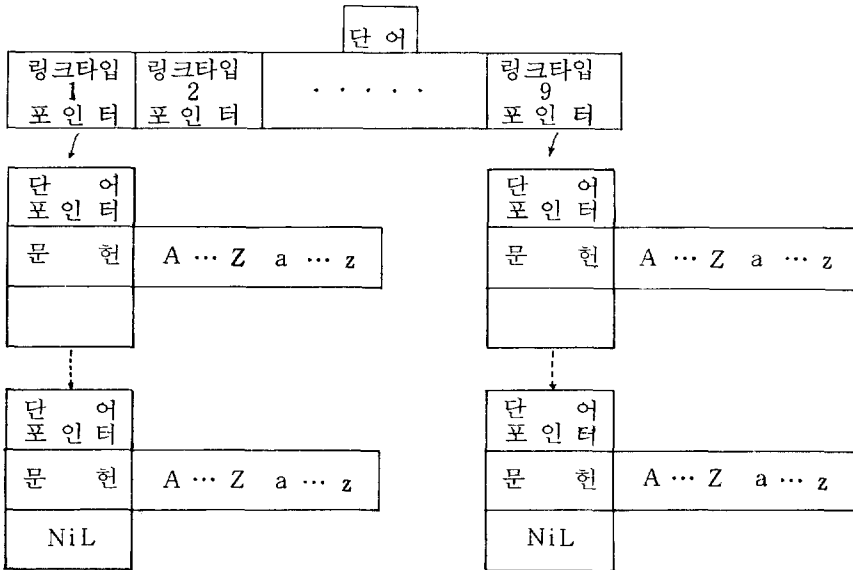
```

Procedure make-network;
/* linktype : type of link,
index      : contain the index of word,
word       : keyword,          */
read(linktype, word, index);
WHILE NOT eof Do
    read(linktype, word, index);
    IF linktype ≠ 0 THEN
        find-parent(word);
    IF have-not-parent THEN
        allocate(node);
    
```


33	균	일	26	B	F	i
34	균	일	성	49	H	O g
35	그	을	음	64		i
36	극	한	치	226	K	
37	근	사	112	A		
38	기	구	148			o
39	기	상	135			yz
40	기	술	231			d
41	기	체	50	H I J	N	
42	기	포	31	I	N g	
43	기	화	70			I
44	길	이	183	A		
45	나	선	형	18	G	
46	난	류	27	B	H	g

특기사항: '유체의 흐름'에 관한 문헌은 A~Z까지 문헌번호를 주었고 '연소'에 관한 문헌은 a~z까지 문헌번호를 주었다. 이것은 색인어 실험에서 문헌집단 I, II로 구별을 위한 것이다.

데이터의 구조는 그림 < 4 > 와 같다.



< 그림 4 > 컴퓨터내의 데이터 구조

8. 탐색방법 및 결과

탐색을 하기 위해서 세워진 탐색전략은 다음의 두가지로 구분하였으며 사용된 질문식은 표 < 7 > 과 같다.

첫째, 탐색자가 선택한 특정 탐색어에 대해 이와

관련된 모든 링크와 관련된 탐색어를 출력한다.

둘째, 2개 또는 3개의 탐색어에 관련된 내용을 갖는 문헌을 검색하며, 이때 개념간의 관계 기호를 주어 탐색한다.

< 표 7 > 질문조건 및 질문식

질문조건	질문식
1. “유체”에 관한 모든 용어와 용어간의 관계를 알고 싶으며 해당문헌도 알고 싶다.	유체
2. ① 재연소반응과 연소와의 관계에 대한 문헌을 찾고자 한다.	반응, 재연소, 연소
② 연소효율에 공기가 주는 영향에 관한 문헌을 알고 싶다.	연소, 효율, 공기
③ 유동층연소로에서 입자의 현상에 관한 문헌을 찾는다.	입자, 연소로, 유동층
④ 원통을 통과하는 유체의 유동에 관한 문헌을 찾는다.	유체, 유동, 원통
⑤ 입자의 표면온도에 관한 문헌을 찾고자 한다.	입자, 표면, 온도
⑥ 표면온도와 일산화탄소와의 관계에 대한 내용의 문헌을 찾는다.	표면, 온도, 일산화탄소
⑦ 조업변수와 온도와의 관계에 대한 문헌을 찾고자 한다.	조업, 변수, 온도
⑧ 슬러리수송에 미치는 속도의 영향에 대한 문헌을 찾는다.	슬러리, 수송, 속도
⑨ 유속이 기체에 미치는 영향에 관한 문헌을 찾는다.	유속, 기체
⑩ 석탄의 전환율과 수증기와의 관계에 대한 문헌을 찾는다.	석탄, 전환율, 수증기

첫번째 방법에 의한 탐색은 컴퓨터 화면에 나오는 메시지에 대해 탐색어의 번호를 줌으로서 결과를 출력할 수 있으므로 별도의 탐색프로그램은 작성되지 않았다. 두번째 방법에 의한 문헌 검색을 위해 작성된 알고리즘은 다음과 같다.

```

Procedure search 2;
/* words are sorted by ascending order.
   binsrch mean binary search */
/* print procedure print its node recursively by preorder traversal */
FOR i← 1 TO 3 DO
    read ( data [ i ] );
    read ( linktype1 , linktype 2 );
    add1← binsrch ( data[1]);
    pointer ← node [add1]. linktype1
    WHILE pointer ≠ NIL DO
    
```

```

IF data[2] = pointer.data THEN
    write (pointer.index);
    pointer← node[add 2].link-
        type 2;
WHILE pointer ≠ NIL DO
    IF data[3]=pointer.data
        THEN
            write(pointer.index);
        ENDIF
        pointer← pointer.next
    ENDIF pointer ← pointer.next
    pointer ←pointer.next
ENDWHILE
    
```

질문조건은 다음의 세가지 경우에 의해서 구분되어 ①~⑩까지의 질문식이 작성된 것이며 실제 2명의 화학공학자의 의견이 고려된 것이다.

첫째, 지식베이스 내에 포함된 탐색어로서 개념간의 직접 관련이 있는 내용의 질문식(①, ②, ⑤, ⑥, ⑦, ⑨, ⑩)

둘째, 탐색어간의 관계가 한가지 이상의 관계를 갖는 질문식(③, ⑧)

셋째, 지식베이스에 포함된 탐색어이기는 하나 동일 문헌에서 관계를 갖고 있지 않은 경우(④)

이상과 같은 질문내용으로 지식베이스에 대한 탐색 및 그 결과는 각각 표<8>, <9>, <10>과 같다.

< 표 8 > 질문식에 사용된 관계기호

질문번호	도치 화일	의미네트워
①	AND, AND	(, ;
②	"	+, :
③	"	+, +
④	"	(, *
⑤	"	+, +
⑥	"	+, ;
⑦	"	+, ;
⑧	"	-, :
⑨	"	:
⑩	"	+, ;

< 표 9 > 적합문헌 및 검색된 문헌

질문번호	적합문헌	의미네트워	
		도치화일 검색된문헌	검색된문헌
①	o, z	a, k, o, y, z	o, z
②	e, g	e, g	e, g
③	e, j	d, e, j	j
④	해당문헌없음	C, J, K, N	-
⑤	a, c	a, c	a, c
⑥	a	a, c, k	a
⑦	e	e	e
⑧	B, F	B, F	B
⑨	H	H, I	H
⑩	b	b, c	b

< 표 10 > 검색효율 비교

(비율: 백분율)

질문번호	도치 화일		의미네트워	
	재현율	정확율	재현율	정확율
①	100	40	100	100
②	100	100	100	100
③	100	66.7	50	100
④	0	0	0	0
⑤	100	100	100	100
⑥	100	33.3	100	100
⑦	100	100	100	100
⑧	100	100	50	100
⑨	100	50	100	100
⑩	100	50	100	100
평균	90	64	80	90

탐색에서 사용된 질문식에 대한 출력결과는 각각 표 <11>, <12>와 같다.

< 표 11 > '유체'에 관한 모든 관련용어와 관계에 대한 출력결과

Node = 유체 /

CURRENT NODE : 유체 /

- 13 LINK TYPE = 1 유동 / CKOX
- 13 LINK TYPE = 1 교반 / G
- 13 LINK TYPE = 1 흐름 / JN
- 13 LINK TYPE = 1 이동 / K
- 13 LINK TYPE = 1 플라스틱 / K
- 13 LINK TYPE = 1 불성 / K
- 13 LINK TYPE = 1 층 / M
- 13 LINK TYPE = 1 유속 / O
- 13 LINK TYPE = 4 고점도 / G
- 13 LINK TYPE = 4 의소성 / G
- 13 LINK TYPE = 4 침성 / CKOX
- 13 LINK TYPE = 4 비압축성 / CKX
- 13 LINK TYPE = 4 비뉴턴 / G
- 13 LINK TYPE = 4 비뉴턴성 / KOX
- 13 LINK TYPE = 4 속도 / OX

13 LINK TYPE = 4	압력장 /	X	30 LINK TYPE = 6	원통 /	J
13 LINK TYPE = 4	온도 /	X	30 LINK TYPE = 7	근사 /	A
13 LINK TYPE = 4	고점성 /	OX	30 LINK TYPE = 7	모사, 수치 /	C
13 LINK TYPE = 4	안정성 /	M	30 LINK TYPE = 7	해석, 수치 /	O
13 LINK TYPE = 4	수평면 /	M	30 LINK TYPE = 7	방정식 /	A
13 LINK TYPE = 4	압력 /	O			
13 LINK TYPE = 4	점도장 /	O	CURRENT NODE : 유속 /		
13 LINK TYPE = 5	교반 /	G	43 LINK TYPE = 3	압력 /	H
13 LINK TYPE = 5	수송 /	B	43 LINK TYPE = 3	층 /	H
13 LINK TYPE = 5	충전 /	O	43 LINK TYPE = 3	기체 /	H
13 LINK TYPE = 5	주입 /	O	43 LINK TYPE = 4	변화 /	O
13 LINK TYPE = 6	전진 /	O	43 LINK TYPE = 7	실험, 수치 /	O
13 LINK TYPE = 7	모사, 수치 /	X	CORRENT NODE : 압력 /		
13 LINK TYPE = 7	실험, 수치 /	O	46 LINK TYPE = 2	시간 /	O
			46 LINK TYPE = 4	강하 /	H
CURRENT NODE : 유동 /			46 LINK TYPE = 4	수평 /	O
30 LINK TYPE = 2	유속 /	O	46 LINK TYPE = 4	입구 /	O
30 LINK TYPE = 2	표면 /	O	CURRENT NODE : 시간 /		
30 LINK TYPE = 4	관 /	N	151 LINK TYPE = 2	체류 /	I
30 LINK TYPE = 4	원통 /	JK	151 LINK TYPE = 6	경과	O
30 LINK TYPE = 4	평판 /	A	CURRENT NODE : 수평 /		
30 LINK TYPE = 4	표면 /	D	228 LINK TYPE = 4	분포 /	O
30 LINK TYPE = 4	조건 /	D	CURRENT NODE : 분포 /		
30 LINK TYPE = 4	칼렌더 /	X	61 LINK TYPE = 2	균일성 /	Og
30 LINK TYPE = 4	해석 /	X	61 LINK TYPE = 3	유동층 /	g
30 LINK TYPE = 4	모양 /	O	61 LINK TYPE = 4	층 /	c
30 LINK TYPE = 5	해석 /	K	61 LINK TYPE = 4	유동층 /	g
30 LINK TYPE = 6	관 /	CN	61 LINK TYPE = 7	실험, 수치 /	X
30 LINK TYPE = 6	배열 /	C			

< 표 12 > 2개 및 3개의 탐색어에 관한 문헌검색 결과

B > se Please wait.....

1 반응	INDEX 1 = akoyz	재연소	INDEX 2 = oz	연소
2 연소	INDEX 1 = eg	효율	INDEX 2 = eg	공기
3 입자	INDEX 1 = j	연소로	INDEX 2 = dej	유동층
4 유체	INDEX 1 = CKOX	유동	INDEX 2 = J	원통
5 입자	INDEX 1 = ac	표면	INDEX 2 = ac	온도
6 표면	INDEX 1 = ac	온도	INDEX 2 = a	일산화탄소

7조업	INDEX 1 = e 변수	INDEX 2 = e 온도
8슬러리	INDEX 1 = B 수송	INDEX 2 = B Fg 속도
9유속	INDEX 1 = H 기체	
10석탄	INDEX 1 = bc 전환율	INDEX 2 = b 수증기

9. 평 가

지식베이스에 대한 탐색 결과는 용어의 제공 능력, 검색의 효율, 기타의 효용성 등에 관하여 평가하였으며 그 내용을 요약하면 다음과 같다.

첫째, 탐색어 확장을 위한 용어의 제공능력은 도치화일에서는 기대할 수 없다. 그러나, 의미네트웍은 원래의 탐색어와 관련된 다른 용어 및 그 관계를 추가할 수 있도록 탐색어를 제공한다.

둘째, 10개의 질문식에 대한 문헌검색에서 평균재현율은 도치화일 탐색이 90%로서 의미네트웍 탐색의 80%보다 높았다. 평균정확율은 도치화일의 64%에 비해 의미네트웍이 90%로서 월등히 높았다.

세째, 의미네트웍에서 표현된 개념관계가 탐색어간의 관계기호로서 사용된 경우 검색의 정확율을 높일 수 있었다.(⑥,⑨,⑩) 반면에, 동일한 문헌에서 색인어간의 관계가 두가지 이상일 때는 도치화일과 의미네트웍 탐색의 정확율은 일치하였다.(②)

네째, 한 문헌에서 직접적인 관계를 갖고 있지 않은 색인어에 대해서 탐색할 때는 탐색자가 상위 혹은 하위의 용어를 입력한다. 그리고 이와 관련된 모든 용어를 제공받은 뒤에 검색을 하면 검색의 재현율에 향상을 가져올 수 있다.(④) 이것은 도치화일이 검색의 목적에 주로 사용될 수 있는 반면에, 지식베이스는 검색 전이나 검색도중에 질문식 수정을 위해서 이용자에게 탐색용어에 관한 정보를 제공한다. 이용자는 탐색전략에서 이를 활용할 수 있다.

IV. 결 론

정보의 탐색활동은 직접탐색이나 대리 탐색의 경우 모두 적합성이 가장 높으며, 또한 관련된 모든 정보를 검색하는 것에 가장 중요한 목적이 있다.

이를 위해서는 정보의 색인과정에서 부터 적합정보의 검색에 이르기까지 모든 구성요소들이 서로 밀접한 관계에 있으며 또한 각각의 중요성을 갖는다. 이용자의 질문어를 처리하기 위해서는 어떤 종류의 정보시스템이든지 소장된 데이터베이스가 과학적인 방법으로 분석 및 가공되어야 할 것이다.

본 연구에서는 정보의 탐색과정에서 이용되는 여러가지 전략 및 방법 중에 탐색용어 및 용어간의 관계에 대한 지식이 명확하였을 때, 검색 결과에 향상을 가져올 수 있음을 고려하여 이를 위한 실험적 연구를 실시하였다.

실험과정에서와 실제의 탐색에서 밝혀진 결과를 다음과 같이 요약함으로써 결론을 맺고자 한다.

첫째, 색인어 적합성 판정을 위해서 복수포아슨분포모형의 한글문헌에 대한 실험은 선행실험결과와 검색의 정확율의 면에서 비슷하였다. 그러나, Z치에 의한 주제어 및 비주제어의 분포현상은 $Z < 1$ 인 경우는 상이한 결과로 나타났다.

확률분포를 이용한 이 기법은 본 연구에 한정하여 볼 때 주제어, 비주제어 판정을 위해서는 비교적 효과적이었다. 그러나, 계산과정이 다소 복잡한 점과, 또한 적합문헌 및 부적합문헌의 구분

과정에서 완전히 이원적인 결정이 어려운 논문이 2건 있었다. 이 경우는 두가지 세부주제의 내용을 응용한 논문으로서 표본문헌 선정시 제외되었다.

둘째, 지식베이스에 포함된 의미지식을 탐색 과정에서 활용하였을 때 검색효율의 향상을 가져올 수 있었다.

우선, 탐색자가 탐색어에 관하여 불충분한 정보를 가진 경우, 실제 질문식 작성에 도움이 되는 용어 및 용어간의 관계를 제공받을 수 있다. 그리고, 의미네트웍을 적절히 확장함으로써 개념간의 의미관계를 이용할 수 있는데 이것은 정보요구와 관련이 있는 문헌검색의 재현을 높일 수 있도록 한다.

세째, 용어에 관한 탐색과정을 통해서 탐색자는 탐색어의 선정과 탐색주제의 정확한 표현이 가능하다. 이것은 문헌검색을 위해서 적합한 탐색용어와 관계기호를 사용하여 검색의 정확율을 높여 주게 된다.

참 고 문 헌

金榮澤, 李京煥, 禹治水, 李東翰, 朴永培 (1985), 컴퓨터 시스템 개요, 서울: 正益社.

김영택, 김종상, 이석호 (1986), "시맨틱 네트 워크를 이용한 관계 DBMS의 확장," 정보관리학회논문지 13, №1 pp.51 ~ 60.

김재희, 박충식 (1986), "인공지능의 지식표현 방식에 대한 비교 고찰", 정보과학회지, 4, №3, pp.31 ~ 42.

김태수 (1984), "연관색인법의 이론과 실제", 정보관리학회지, 1, №1, pp.25~42.

_____, 사공철 (1987), 정보관리론, 서울:

구미무역

노옥순 (1986), "정보서비스의 새로운 추세와 적용문제에 관한 고찰," 도서관학, 제 13 집, pp.78 ~ 91.

랑카스터, F.W. (1979), 정보검색시스템, 김태승, 윤구호 (共譯), 서울: 구미무역, 1985.

우동진 (1987), "통계적기법에 의한 한글자동 색인의 연구" 정보관리학회지, 4, №1, pp.47 ~ 86.

이태영, 정영미 (1982), "자동색인의 통계적 기법과 한국어 문헌의 실험," 도서관학, 제 9 집, pp.99 ~ 118.

電子時報社 出版部 (1985), CP/M 입문, 서울: 電子時報社.

정영미 (1980), "색인이론과 실제", 연세논총, 17, pp.21 ~ 35.

_____, (1983), "Shannon 정보이론과 도서관학, 정보학적 해석에 관한 연구", 연세논총, 20, pp.83 ~ 101.

_____, (1987), 정보검색론, 서울: 정음사

장재경 (1986), "우리말 문헌정보 검색을 위한 지식베이스 설계에 관한 연구", 정보관리학회지, 3, №1, pp.70 ~ 95.

學園社 (編) (1972), 철학대사전, 서울: 學園社.

한상영, 전주식, 유석인 (1986), 人工知能, 서울: 尙潮社.

韓國化學工學會 (編) (1987), 화학공학 술어집, 서울: 한림상사

_____, (1984), 화학공학총서인, 제 1 집.

大須賀節雄 (1986), 知識ベース入門, 東京: オーム社.

- American Chemical Society(1986), Chemical Abstract : general subject index, 105, No. 26.
- Bates, Marcia J. (1979), "Information Search Tactics," *JASIS* 30, No. 4, pp. 205-214.
- Bawden, David(1986), "Information Systems and the Stimulation of Creativity," *Jl. of Information Science*, 12, pp. 203-216.
- Brookes, Bertram C. (1980), "The Foundations of Information Science : Part 1. Philosophical Aspects," *Jl. of Information Science*, 2, pp. 125-133.
- (1986), "Jason Farradane and Relational Indexing," *Jl. of Information Science*, 12, pp. 15-18.
- Bookstein, A. and D. R. Swanson. (1975), "A Decision Theoretic Foundation for Indexing," *JASIS*, 26, No. 1, pp. 45-50
- and D. Kaft (1977), "Operation Research Applid to Document Indexing and Retrieval Decision," *Jl of ACM.*, 24 No. 3, pp. 418-427
- and D. R. Swanson(1974), "Probabilistic Models for Automatic Indexing," *JASIS*, 25, No. 5, pp. 312-318.
- Borko, H. (1977), "Toward a Theory of Indexing," *Information Processing & Management*, 13, No. 6, pp. 355-365.
- (1978), *Indexing Concepts and Methods*, New York : Academic Press.
- Bottle, R. T. (1983), "Changes in the Communication of Chemical Chemical Information II : An Updated Model," *Jl. of Information Science*, 6, pp. 109-113.
- Cercone, N. and G. McCalla(1984), "Artificial Intelligence : Underlying Assumptions and Basic Objective," *JASIS*, 35, No. 5, pp. 280-290.
- Childers, Thomas (1980), "The Test of Reference," *Library Journal*, 105, No. 8, pp. 924-928.
- Cohen, Paul. R. and Edward A. Feigenbaum (1982), *The Handbook of Artificial Intelligence*, 3, California : William Kaufmann.
- Corcoran, M., Richard C., and Denn C. (1980), "Subject Specialist Searching Chemical Abstracts on SDC," *Proceedings of the 43rd ASIS Annual Meeting*, New York : ASIS, 17, pp. 345-346.
- Craven, Timothy C. (1980), "NEPHIS Index Production from a General Concept Network," *Proceedings of the 43rd ASIS Annual Meeting*, New York : ASIS, 17, pp. 180-182.
- Cross, G. R. and C. G. DeBessonet(1985), "Representation of Legal Knowledge for Conceptual Retrieval," *Information Processing & Management*, 21, No. 1, pp. 35-44.
- Duke, John(1985), "Information Management and the Consultant : towards the Infallible Expert," *Aslib Proceedings*, 37, No. 3, pp. 157-161.
- Fairhall, Donald (1985), "In Search of Searching Skills," *Jl. of Information Science* 10, pp. 111-123.
- Farradane, J. E. L. (1970), "Analysis and Organization of Knowledge for Retrieval," *Aslib Proceedings*, 22, No. 12, pp.607-616.
- (1980), "Knowledge, Information, and Information Science," *Jl. of Information Science*, 2, pp. 75-80.
- and D. Thompson(1980), "The Testing of Relational Indexing Procedures by Diagnostic Computer Programs," *Jl. of Information Science*, 2, pp. 607-616.
- Fenichel, Carol Hansen(1980), "The Process of Searching Online Bibliographic Databases : a Review Research," *Library Research*, 2, No. 2, pp. 107-127.
- (1980), "Intermediary Searchers' Satisfaction with the Results of Their Searches," *Proceedings of the 43rd ASIS Annual Meeting*, New York : ASIS, 17, pp. 58-60.
- Fidel, R. and C. Travis(1981), "Subject Analysis," *ARIST*, 17, pp. 125-127.
- Fidel, R. (1886), "Towards Expert Systems for the Selection of Search Keys," *JASIS*, 37, No. 1, pp. 37-44.
- Gevarter, William B. (1984), *Artificial Intelligence, Expert Systems, Computer Vision and Natural Languages Processing*, Noyes
- Godert, Wilfried(1980), "Subject Headings for Mathematical Literature," *Jl. of Doc.*, 36, No. 1, pp. 11-23.

- Harter, S. P. (1975a), "A Probabilistic Approach to Automatic Keyword Indexing," JASIS, 26, No. 4, pp. 197-206.
- (1975b), "A Probabilistic Approach to Automatic Keyword Indexing part II : an Algorithm for Probabilistic Keyword Indexing part II : an Algorithm for Probabilistic Indexing," JASIS, 26, No. 4, pp. 280-289.
- Jahoda, Jerald (1974), "Reference Question Analysis and Search Strategy Development by Man and Machine," JASIS, 25, No. 3, pp. 139-144.
- Jones, K. P. and C. L. Bell (1986), "An Intelligent Retrieval System," Aslib Proceedings, 38, No. 1, pp. 71-79.
- Keen, E. Michael (1977), "On the Processing of Printed Subject Index during Searching," JI. of Doc., 33, No. 4, pp. 266-276.
- Knapp, Sara D. (1978), "The Reference Interview in the Computerbased Setting," RQ, 17, No. 4, pp. 320-324.
- Lancaster, F. Wilfrid (1979), Information Retrieval Systems : Characteristics, Testings and Evaluation, New York : John Wiley & Sons.
- Lunin, L. F. and L. C. Smith (1984), "Artificial Intelligence Concept, Techniques, Application Promise," JASIS, 35, No. 3, pp. 277-319.
- Maron, M. E. (1979), "Depth of Indexing," JASIS, 30, No. 4, pp. 224-228.
- Morrows, Deanna I. (1976), "A Generalized Flow-chart for the Use of ORBIT and other Online Interactive Bibliographic Search Systems," JASIS, 27, No. 1, pp. 57-62.
- Meadow, Charles T. and Pauline Cochrane (1981), Basics of Online Searching, John Wiley & Sons, Inc.
- Molholt, P. A. (1986), "The Information Machine : a New Challenge for Librarians," Library Journal, 111, 16, pp. 47-52.
- Neill, S. D. (1975), "Farradane's Relations as Perceptual Discriminations," JI. of Doc., 31, No. 3, pp. 144-157
- (1984), "The Reference Process and Certain Types of Memory : semantic, episodic, and schematic," RQ, 23, No. 4, pp. 417-423.
- Norris, C. (1981), "MeSH-The Subject Heading Approach," Aslib Proceedings, 33, No. 4, pp. 153-159.
- Nowak, Elibieta J. and Bogumil F. Szablowski (1984), "Expert Systems in Scientific Information Exchange," JI. of Information Science, 8, pp. 103-111.
- Oddy, R. N. (1977), "Information Retrieval Through Man-Machine Dialogue," JI. of Doc. 33, No. 1, pp. 1-14.
- Paice, C. (1986), "Expert Systems for Information Retrieval," Aslib Proceedings, 38, No. 10, pp. 347-350.
- Politt, A. S. (1984), "'Front-end' system : an Expert System as an Online Research Intermediary," Aslib Proceedings, 36, No. 5, pp. 229-234.
- Powell, R. R. (1985), Basic Research Methods for Librarians, New Jersey : Ablex pub.
- Robertson, S. E. (1981), "Probabilistic Models of Indexing and Searching," Information Retrieval Research, Butterworth.
- Salton, G. and M. J. McGill (1983), Introduction To Modern Information Retrieval, New York : McGraw-Hill.
- Salton, G. (1983), "Some Characteristics of Future Information Systems," The Application Mini and Micro Computers in Information, Documentation, and Libraries, ed. C. Keren, Elsevier Science pub.
- Schwartz, C. and L. M. Elisemann (1986), "Subject Analysis," ARIST, 21, pp. 37-62.
- Shoval, P. (1983), "Knowledge Representation in Consultation Systems for Users of Retrieval Systems," The Application of Mini and Micro Computers in Information, Documentation, and Libraries, ed. C. Keren, Amsterdam : Elsevier Science Pub.
- (1985), "Principles, Procedures, and Rules in an Expert Systems for Information Retrieval," Information Processing & Management, 21, No. 6, pp. 475-487.
- Smith, K. F. (1986), "Robot at the Reference Desk?" College and Research Libraries, 47, No. 5, pp. 486-490

- Smith, Linda C. and Amy J. Warner (1984), "A Taxonomy of Representations in Information Retrieval System Design," *Jl. of Information Science*, 8, pp. 113-121.
- Sparck-Jones, Karen (1984), "Proposals for R & D in Intelligent Knowledge Based Systems (IKBS)," *Jl. of Information Science* 8, pp. 139-147.
- Waterman, D. A. (1986), *A Guide to Expert Systems*, Massachusetts : Addison Wesley.
- Waters, S. T. (1986), "Answerman, the Expert Information Specialist : an Expert System for Retrieval of Information From Library Reference Books," *Information Technology and Libraries*, 5, No. 3, pp. 204-212.
- Yaghmai, N. S. and Jacqueline A. M. (1984), "Expert Systems : A Tutorial," *JASIS* 35, No. 5, pp. 297-305.