

# 도서관·정보학에서의 인공지능의 응용에 관한 고찰

정 영 미\*

## <目 次>

- |                   |                         |
|-------------------|-------------------------|
| 1. 서 론            | 3.2 전문가시스템의 구성          |
| 2. 자연언어처리시스템      | 3.3 전문가시스템과 정보검색시<br>스텝 |
| 2.1 개요            | 3.4 온라인탐색 중개시스템         |
| 2.2 자연언어처리시스템의 구성 | 3.5 참고업무 전문가시스템         |
| 2.3 지식베이스의 구성     | 3.6 분류·편목 전문가시스템        |
| 2.4 자연언어 인터페이스시스템 | 3.7 색인·어휘통제 전문가시스템      |
| 2.5 질문응답시스템       |                         |
| 3. 전문가시스템         | 4. 결 론                  |
| 3.1 개요            |                         |

## 1. 서 론

인공지능에 관한 연구는 1950년대 후기부터 시작되었으며, 1970년 이전까지는 주로 자연언어의 기계번역에 치중한 것으로 나타나 있다. 이 시기의 연구성과는 그다지 성공적이지는 못하였으나 이후의 연구에 있어서의 새로운 방향을 제시하여 주었다는 점에서는 의의가 있다고 할 것이다.

1970년대에 들어서는 인공지능의 기초연구대상으로 지식의 중요성을 강조하기 시작하였으며, 지식표현기법에 관한 연구가 활발한 연구영역으로 대두되었다. 이 시기의 또 하나의 특성은 언어학, 논리학, 심리학 등을 비롯한 여러 다른 학문과의 상호교류가 활발해졌다는 점이다.

자연언어처리에 있어서 1970년 이전에는 주로 도식대조(template matching)와 같은 초보적인 분석기법이 사용되었으나, 1970년 이후에는 SHRD

---

\* 연세대학교도서관학과 부교수

LU<sup>1)</sup>와 같이 자연언어의 구문·의미분석을 성공적으로 처리한 시스템들이 나타나기 시작하였다. 또한 1965년에 개발된 DENDRAL에 이어 MYCIN 등과 같은 성공적인 전문가시스템들이 다수 개발되었다.

1980년대에 들어서 특히 인공지능의 응용영역에서 나타난 특징은 수많은 전문가시스템의 출현이며, 전문가시스템의 개발을 위한 소프트웨어들이 상품화되어 이러한 시스템의 구현을 용이하게 하고 있다는 점이다. 또한 컴퓨터시스템과의 자연언어인터페이스를 제공하는 상용시스템들이 개발되어 이용자에게 편의를 제공하고 있다.

최근 도서관·정보학 분야에서도 인공지능의 응용영역 가운데 특히 자연언어처리와 전문가시스템에 관한 연구가 활발해지고 있는 것을 볼 수 있다. 자연언어처리에 관한 연구는 자동색인과 자동초록의 작성을 위해 일찌기 1950년대부터 시작되었지만 인공지능이란 주제아래 논의되기 시작한 것은 1980년대에 들어서이며, 전문가시스템에 관한 논문들도 1980년대 초기에 나타나기 시작하였다. *Annual Review of Information Science and Technology*의 경우 '자동언어처리'라는 제목아래 1966년(v.1)부터 자연언어처리에 관한 논문이 수록되어왔으나 '인공지능'이라는 주제로는 1980년(v.15)에 최초의 논문<sup>2)</sup>이 게재되었다.

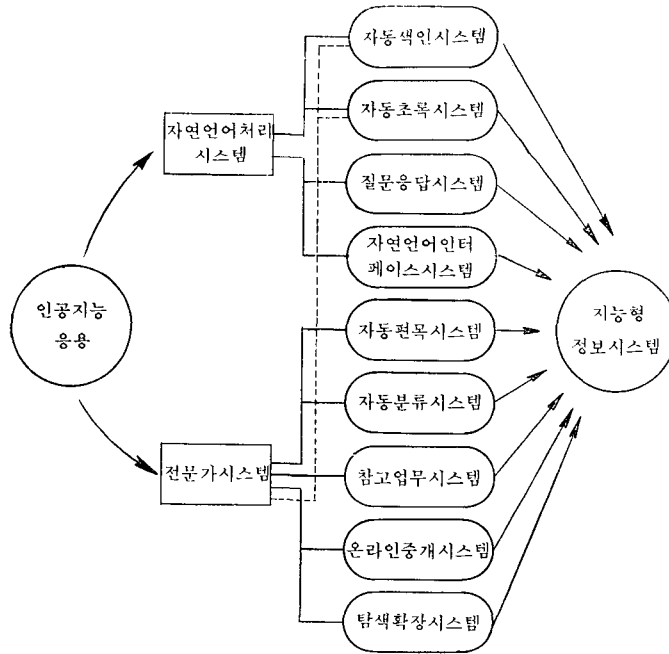
도서관·정보학 분야에서 자연언어처리에 관한 연구는 자동색인 및 자동초록시스템에서 시작되었고, 질문응답시스템에서 자연언어로 된 텍스트 및 질문의 처리와 해답의 생성을 위해 본격적으로 진행되어 왔다. 또한 일반적인 정보검색시스템에서 이용자의 편의를 위해서 제공하는 자연언어인터페이스시스템에 관한 연구가 수행되고 있다.

반면 전문가시스템은 인간이 수행하는 지적인 작업을 컴퓨터가 대신하도록 하는 시스템으로서 앞의 자동색인·초록시스템도 이러한 의미에서는 일종의 전문가시스템이며, 이외에 자동편목·분류시스템, 참고업무시스템, 온라인 탐색을 대신하는 온라인탐색 중개시스템, 온라인탐색 확장시스템 등이

1) T. Winograd, *Understanding Natural Language*. New York: Academic Press, 1972.

2) L.C. Smith, "Artificial Intelligence Applications in Information Systems," *ARIS T* 15 : 67-105 ; 1980.

전문가시스템으로 개발되고 있다. 이러한 모든 시스템들은 하나의 정보시스템 속에 포함되는 하부시스템이 될 수 있으며, 이러한 정보시스템을 ‘지능형 정보시스템’이라고 부를 수 있을 것이다. 앞의 내용을 요약해서 표현하면 아래 그림과 같다.



## 2. 자연언어처리시스템

### 2.1 개요

자연언어처리란 제약을 받지 않는 자연적인 언어로 쓰여진 글이나 발성된 말을 컴퓨터가 처리하는 모든 과정을 의미한다. 자연언어처리의 결과는 입력된 자연언어 문장이나 말의 올바른 이해와 문법에 맞는 문장 및 말의 생성에 있다고 볼 수 있다.

자연언어처리에는 음운분석단계, 형태소분석단계, 어휘분석단계, 의미분석

단계, 화용론적 분석단계로 진행된다. 발전된 자연언어처리시스템일수록 처리단계는 올라가며, 처리과정은 그만큼 복잡해진다. 일반적인 의미의 자연언어처리시스템은 문장분석수준이 구문분석 내지는 의미분석 단계에까지 이르고 있는 시스템을 말하고 있으며, 본고에서도 이런 시스템을 중심으로 하여 기술하고자 한다.

BASEBALL, SAD-SAM, STUDENT, PROTOSYNTHEX 등의 초기시스템은 주로 특정한 단어(구)의 도식대조 방식에 의해 문장을 분석하였으므로 구문분석은 비교적 초보적인 수준에 머물렀다.<sup>3)</sup> 좀 더 발전된 시스템들은 ATN 문법이나 변형문법 등을 이용하여 본격적인 구문분석을 시도하였으며, SHRDLU, ROBOT(후의 INTELLECT), LUNAR, PLANES 등은 구문분석과 의미분석을 병행한 시스템으로 평가되고 있다.<sup>4)</sup> 특히 적문법을 이용한 시스템들은 구문분석과 의미분석을 통합시킨 자연언어시스템으로 볼 수 있다.

도서관·정보학 분야에서의 자연언어처리의 필요성은 특히 언어학적 기법을 이용한 자동색인 및 자동초록시스템에서 두드러지게 나타나고 있다. 최근에 이르러는 자연언어형태의 질문을 정형의 질의어로 변형시키는 정보검색시스템(자연언어 인터페이스시스템이라고도 함), 전문데이터베이스로부터 해당문장을 찾아내는 전문검색시스템, 자연언어텍스트 및 질문을 처리하고 해당도 자연언어형태로 생성이 가능한 질문응답시스템, 그리고 다양한 종류의 전문가시스템으로까지 그 응용영역이 확대되고 있다.

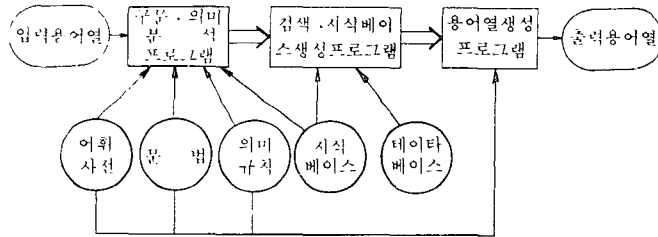
## 2.2 자연언어처리시스템의 구성

자연언어처리시스템인 자연언어 인터페이스시스템과 질문응답시스템은 자연언어 문장을 입출력시킨다는 점에서는 같은 구조를 갖는다. 그러나 자연언어 인터페이스시스템은 일반적으로 DBMS를 통한 데이터베이스의 검색을 위해 자연언어를 정형의 질의어로 변환시키는 과정만을 포함하는 시스템

3) H. Tennant, *Natural Language Processing*. N.Y.: PBL, 1981.

4) *ibid.*

인 반면, 질문응답시스템은 자연언어 텍스트의 구조화(데이터베이스의 생성을 의미함) 및 질문의 처리가 병행되며, 데이터베이스를 검색하여 해답을 제공하는 전 과정을 포괄하는 시스템인 것이다. 결국은 자연언어 인터페이스 시스템은 질문응답시스템이 이용자와 만나는 접촉부분을 구성한다고 보면 된다. 따라서 다음 그림에 나타나 있는 자연언어처리시스템의 구성도는 이러한 질문응답시스템의 맥락에서 이해하면 된다.



위의 그림에서 입력용어열은 문헌의 텍스트나 이용자의 질문이 되며, 출력용어열은 자연언어 형태의 문장이 된다. 기계번역시스템의 경우에는 입력용어열은 번역대상언어(소스언어)로 된 문장이며, 출력용어열은 번역된 언어(목적언어)로 된 문장을 의미한다. 물론 질문응답시스템의 경우에 출력용어열은 문장형태가 아니고, 데이터베이스에서 검색해낸 데이터 형태 그대로의 해답이 될 수가 있다.

구문·의미분석프로그램은 입력용어열을 받아들여 지식베이스에 대응되는 내부구조로 변환시키는 기능을 수행한다. 여기에는 품사분류를 위한 어휘사전과 구문분석을 위한 문법 및 의미해석에 필요한 의미규칙 등이 사용된다. 검색 및 지식베이스 생성프로그램은 지식베이스와 데이터베이스에 접근하여 지식을 검색하고, 추론하고, 새로운 지식을 추가하는 기능을 수행한다. 용어열생성프로그램은 지식베이스로부터 검색해낸 지식을 용어열 형태의 해답으로 생성하여 출력한다.

일반적인 정보검색시스템에서는 검색대상 정보의 집합을 데이터베이스라고 부르고 있다. 그러나 텍스트 형태 그대로가 아니라, 새로운 형태로 구조화시킨 지식을 소장하였다가 검색하는 질문응답시스템에서는 지식베이스란

용어를 사용하기도 한다. 위에서 설명할 전문가시스템에서도 전문적인 지식을 담고 있다는 점에서 지식베이스라는 용어를 사용하고 있다. 지식베이스의 사전적 정의를 살펴보면 다음과 같다. 지식베이스란 “규칙, 절차, 스키마, 또는 작업장소내 메타요소 등의 형태로 표현된 지식의 집합으로서 고도로 구조화되고 상호연결된 데이터베이스”를 의미한다.<sup>5)</sup> 이러한 지식베이스를 이용한 시스템을 일반적으로 지식기반시스템(knowledge-based system)이라고 부른다. 반면 인공지능에서의 데이터베이스는 “원하는 결과를 얻기 위해 지식베이스를 이용하여 작업할 대상이 되는 사물과 사건에 관한 데이터의 집합”으로 정의됨을 유의할 필요가 있다.<sup>6)</sup>

### 2.3 지식베이스의 구성

지식베이스는 자연언어처리시스템이나 전문가시스템 등 지식기반시스템의 기본이 되는 요소로서 지식을 표현하는 여러가지 기법에 의해 구성된다. 본 장에서는 지식베이스에 지식을 소장하는 지식표현기법에 관해 자연언어 문장을 구조화시켜 표현하는 예를 중심으로 간단히 살펴보고자 한다.

지식베이스내의 지식은 의미네트웍(semantic networks), 프레임(frames), 스크립트(scripts), 생성규칙(production rules), 논리적 표현 등의 형태로 표현되고 있다. 의미네트웍은 비교적 많이 사용되고 있는 구조로서 자연언어 문장의 표현에도 적합하다. 의미네트웍은 사물·행위·사건 등의 개념을 나타내는 노드(nodes : 의미점)와 이들간의 관계를 나타내는 아크(arcs : 연결호)로 구성된다. 예를 들어 “영회는 철수를 좋아한다”라는 문장은 의미네트웍 형태로 표현하면 아래와 같이 ‘영회’, ‘철수’, ‘좋아하다’는 각각 노



5) V.D. Hunt, *Artificial Intelligence & Expert System Sourcebook*. New York: Chapman & Hall, 1986. p.147.

6) *ibid.*, p. 83.

드가 되며, 문장속에서의 이들간의 구문적 관계는 아크가 된다.

문장을 의미네트웍으로 표현할 때 하나의 노드가 하나의 단일 개념만을 나타내지 않고 개념의 여러 단위를 표현하도록 하는 방법도 있다. 노와(Nowak) 등은 과학기술정보의 검색을 위한 전문가시스템에서 여러 유형의 노드를 설정하였다.” 여기에는 사물 등의 단일개념을 나타내는 단어노드, 단어들 이 형성하는 단어구노드, 완전한 문장형태의 사실노드 등이 포함된다.

프레임은 사물이나 사건과 같은 개념을 나타내는 지식표현기법으로서 하나의 프레임은 여러개의 슬롯(slots)으로 구성된다. 각 슬롯은 슬롯명을 가지며, 슬롯은 해당하는 값으로 채워진다. 스크립트는 프레임의 특별한 형태로서 일상적으로 일어나는 사건을 표현하기 위해 사용한다. 프레임구조는 여러가지 속성을 갖는 사물의 표현에 특히 적합하다. 디소러스를 지식베이스로 구성할 때 하나의 개념에 관한 프레임을 구성하고, 각 슬롯에는 동등관계, 제층관계, 연관관계를 갖는 다른 개념어들을 담으면 된다. 자연언어문장의 경우 앞의 예문을 간단한 프레임구조로 표현하면 다음과 같다. 아래의 행위자, 행위, 대상자 등은 슬롯의 이름이다.

문장 1

행위자 : 영화  
 행 위 : 좋아하다  
 대상자 : 철수

생성규칙은 전문가시스템에서 널리 사용되는 표현기법으로서 if 조건절과 then 행위절(if-then)의 형식으로 규칙을 표현한다. if 다음에 오는 조건이 만족되면 then 다음의 행위를 수행하도록 하는 것이다. 예를 들면 “if 영화는 철수를 좋아한다 and 철수는 영화를 좋아한다 then 영화와 철수는 친구이다”와 같은 표현이 가능하다. 다시 말해 if 다음에 and로 연결되는 두 가지 조건이 만족되면 then 다음의 문장이 결론이 되는 것이다. 또는 “if 비율확률이 50% 이상이다 then 우산을 가져가라”와 같이 표현한다.

7) E.J. Nowak & B.F. Szablowski, “Expert Systems in Scientific Information Exchange,” *J. of Inf. Sci.*, 8(3) : 103-111 ; 1984.

앞의 세가지 지식표현기법은 표현된 사실을 직접 검색할 수 있도록 조직한다. 즉, 프레임구조의 특정한 슬롯의 내용이라든가 의미네트웍의 노드의 이름 등을 적절한 색인 등을 이용하여 직접 검색해낸다. 반면 논리적 표현 형식은 특정한 사실의 검색보다는 특정한 사실의 진위를 알아내는 데 주목적이 있다.<sup>8)</sup> 다시 말해 앞의 예문을 (좋아하다(영희, 철수))와 같이 술어 해석적으로 표현했을 때 “영희는 누구를 좋아하는가?”에 대한 해답을 검색하기보다는 “영희가 철수를 좋아하는가?”에 대한 ‘예, 아니오’ 형태의 해답을 구하게 된다.

논리적 표현은 명제해석, 명제해석의 확장인 술어해석, 그리고 일차 술어 해석 등을 이용하여 표현할 수 있는데 인공지능에서는 술어해석의 변형인 일차 술어해석이 가장 널리 이용되고 있다. 술어해석은 논리적 표현과 추론 규칙으로 구성되며, 추론규칙에 의해 지식베이스로부터 새로운 지식을 추론해 낼 수 있다. 술어해석적 표현은 술어(predicate)와 대상(argument)으로 구성되며, 단일명제들은 and(기호  $\wedge$ ), or(기호  $\vee$ ), not(기호  $\sim$ ), if~then(기호  $\rightarrow$ ) 등의 연결어를 사용하여 연결된다. “영희는 철수를 좋아하고 철수는 영희를 좋아한다”와 같은 복합문은 아래와 같이 2개의 단일명제가 and로 연결되어 표현된다.

(and (좋아하다(영희, 철수))

(좋아하다(철수, 영희))

또한 all(기호  $\forall$ ), some(기호  $\exists$ )과 같은 수량사 및 변수( $x, y$  등)를 이용한 표현이 가능하다. 즉 “모든 사람은 죽는다”라는 문장은 다음과 같이 표현이 가능하다.

forall (if (사 람  $x$ )

(죽는다  $x$ )

#### 2.4 자연언어 인터페이스시스템

자연언어 인터페이스시스템은 광의로는 자연언어를 사용하는 이용자와 인

8) P. Harmon & D. King, *Expert Systems*, New York: Wiley, 1985. p.47.



공언어를 사용하는 컴퓨터시스템 간에 인터페이스(접속수단)를 제공하는 모든 시스템을 의미한다. 즉 뒤에서 논의될 전문가시스템을 이용자가 자연언어를 사용하여 이용할 수 있는 경우에는 이 전문가시스템은 자연언어인터페이스를 제공하는 하부시스템을 포함하는 것이다. 그밖에 다른 유형의 컴퓨터시스템에서도 이용자와 시스템이 자연언어로 대화할 수 있도록 되어 있는 경우에는 이용자와 컴퓨터시스템 사이에 자연언어 인터페이스시스템이 존재하는 것이다.

정보검색시스템에서 일반적으로 사용하는 정형의 질문언어의 유형은 크게 두가지로 구분해 볼 수 있다. 첫째는 주로 서지정보를 검색하는 온라인 정보검색시스템의 불리안논리 탐색문 형태이며, 둘째는 DBMS를 이용한 데이터검색시스템에서 사용하는 정형의 질의어 형태이다. 관계형 DBMS에서 사용되는 언어인 QUEL이나 SQL 등을 예로 들 수 있다.

불리안논리 탐색문은 기본적으로 ‘탐색명령어+탐색어+불리안논리연산자+탐색어’의 형태이며, 이때의 탐색어가 주제어인 경우에는 자유키워드이거나 통제키워드(디스크립터)를 사용한다. 전문가시스템의 하나인 온라인중개시스템에서는 이용자가 형식적인 탐색문 작성을 피하도록 하기 위해 자연언어형태의 질문을 받아들여 불리안논리 탐색문으로 변환시키는 초보적인 자연언어인터페이스를 제공하기도 한다. CANSEARCH<sup>9)</sup>, CONIT<sup>10)</sup> 등의 온라인중개시스템은 처리과정은 다르지만 모두 자연어구 형태의 질문을 불리안논리 탐색문으로 변환시킨다.

CANSEARCH에서는 “Hyperthermia used in cancer therapy”와 같은 질문이 있으면 이용자가 메뉴방식에 의해 일반적인 주제어에서 특정한 주제어 순으로 탐색어를 지시해나감으로써 해당되는 프레임의 계층을 탐색해 나간다. CONIT이나 또는 ANNOD,<sup>11)</sup> CITE<sup>12)</sup> 등의 정보검색시스템은 입력된

9) S. Politt, “CANSEARCH: An Expert Systems Approach to Document Retrieval,” *Inf. Processing & management*, 23(2) : 119-138 ; 1987.

10) R.S. Marcus & J.F. Reintjes, “A Translating Computer Interface for End-User Operation of Heterogeneous Retrieval Systems. I. Design,” *JASIS*, 32(4) : 287-303 ; 1981.

11) L.M. Bernstein & R.E. Williamson, “Testing of a Natural Language Retrieval

자연어구(문장) 형태의 질문에서 불용어사전에 의해 키워드를 추출한 뒤 어간통제를 하여 불리안연산자로 연결시키거나 가중치를 부여하며, 디소러스에 의한 탐색어확장도 가능하게 하고 있다. 예를 들어 미국 국립도서관의 CITE/CATLINE 시스템에서는 “the use of biofeedback and relaxation in the treatment of stress headache”와 같은 질문이 입력되면 이 질문은 ‘relaxation technics, benzodiazepine tranquilizers feedback,...’ 등의 MESH 표목과 ‘biofeedback, relaxation, stress,...’ 등의 텍스트 키워드를 포함한 모두 13개의 탐색어로 변환된다.

위의 시스템들은 구문분석이 일어나지 않기 때문에 극히 초보적인 자연언어 인터페이스시스템이라고 볼 수 있다. 진정한 의미의 자연언어인터페이스는 질문의 형태가 보다 자연언어적이어야 하며, 구문·의미분석프로그램을 통해 키워드간의 관계도 보다 구체적으로 분석되어야 하는 것이다. 비스와스(Biswas) 등<sup>13)</sup>이 개발한 정보검색시스템에서는 ATN 문법을 이용하여 질문으로부터 탐색어와 이들간의 관계를 추출하도록 하였다.<sup>13)</sup> 예를 들어 “Retrieve a few very recent and important survey articles about multi-valued logic”과 같은 내용의 질문도 다소 변형된 형태의 불리안논리 탐색문으로 변환이 가능하도록 하였다.

반면 DBMS와의 인터페이스를 제공하는 시스템은 LIFER,<sup>14)</sup> LADDER,<sup>15)</sup> CO-OP<sup>16)</sup> 등 상당수가 있으며, INTELLECT<sup>18)</sup>와 같은 상용시스템도 개발

System for a Full Text Knowledge Base,” *JASIS*, 35(4) : 235—247 ; 1984.

- 12) T.E. Doszkocs, “CITE NLM: Natural Language Searching in an Online Catalog,” *Inf. Technology and Libraries*, 2(4) : 364—380 ; 1983.
- 13) G. Biswas, et al., “Knowledge-Assisted Document Retrieval: 1. The Natural-Language Interface,” *JASIS*, 38(2) : 83—96 ; 1987.
- 14) G.G. Hendrix, “Human Engineering for Applied Natural Language Processing,” *Proceedings of the International Joint Conference on Artificial Intelligence*, 1977.
- 15) G.G. Hendrix, et al., “Developing a Natural Language Interface to Complex Data,” *ACM Trans. on Database. Systems*, 3(2) : 105—147 ; 1978.
- 16) S.J. Kaplan, “Cooperative Responses from a Portable Natural Language Query System,” *Artificial Intelligence*, 19 : 165—187 ; 1983.
- 17) S.J. Kaplan, “Designing a Portable Natural Language Database Query System,” *ACM Trans. on Database Systems*, 9(1) : 1—19 ; 1984.
- 18) L. Harris, “Experience with ROBOT in 12 Commercial Natural Language Database

되어 있다. 대부분의 인터페이스시스템들은 자연언어 질문만을 분석하여 줌으로써 DBMS는 데이터베이스내의 특정한 데이터항목을 직접적인 해답으로 제공하는 데 그치고 있다. CO-OP과 같은 시스템은 직접적인 해답이외에 다양한 형태의 간접적인 응답을 자연언어 형태로 제공하는 기능을 갖고 있다. 다시 말해 이용자와 시스템간에 자연언어 형태의 대화가 가능한 인터페이스를 제공하는 것이다. 이러한 시스템들은 앞에서 언급한 정보검색시스템이나 온라인 중개시스템이 거의가 광범위한 주제영역을 망라하는 것과는 달리 대부분 특정한 주제영역에 국한시켜 개발되었기 때문에 자연언어 질문의 보다 완전한 분석이 가능하다고 보겠다. 또한 데이터베이스를 구성하는 데이터항목들의 수가 적으며, 대개 특정한 값들을 갖도록 한정되어 있다는 것도 자연언어인터페이스를 보다 완벽하게 수행하도록 하는 이유가 된다.

DBMS에 인터페이스를 제공하는 시스템들은 사실상 질문응답시스템의 일부를 구성하는 것이므로 광의로는 질문응답시스템이라고 볼 수 있다. 따라서 여기에서는 특히 한글자연언어인터페이스시스템의 예만을 살펴보기로 한다. 한글 인터페이스시스템인 NHI 시스템은 한글질의어를 관계형 DBMS인 INGRES의 정형질의어인 QUEL로 변환시키는 시스템이다.<sup>19)</sup> NHI는 어휘분석프로그램, 구문분석프로그램, 변형프로그램의 세 부분으로 구성되며, DCG 문법을 사용하였다. 어휘분석프로그램은 입력된 자연언어문장을 단어 단위로 분류하여 토큰을 생성하며, 구문분석프로그램은 DCG를 이용하여 문장의 구문 및 의미를 분석하여 내부구조를 생성한다. 이 내부구조를 변형 프로그램에 의해 정형의 질의어로 변환한다. 이 정형의 질의어는 INGRES에 의해 데이터베이스 탐색에 사용된다. 예를 들어 “전산과 학생은 누구인가”라는 질문은 아래와 같은 형태의 정형질의어로 변환된다.

range of R1 is 학생  
retrieve(R1. 학생명)

Applications,” *Proceedings of the International Joint Conference on Artificial Intelligence*, 1979.

19) 김성기, 자연한글 질의어처리를 위한 인터페이스의 설계 및 구현에 관한 연구. 서울대학교 대학원 석사학위논문, 1985.

where R1. 학생 = '전산학과'

## 2.5 질문응답시스템

본격적인 자연언어처리가 가능한 질문응답시스템으로는 잘 알려진 LUNAR 시스템을 비롯하여 PLANES,<sup>20)</sup> REQUEST<sup>21)</sup> 등이 있다. 위의 세 시스템은 모두 자연언어질문을 ATN 문법이나 변형생성문법에 의해 분석한 뒤 여러개의 메타영역으로 구성된 데이터베이스를 검색하여 응답하는 시스템으로서, 텍스트를 구조화시킨 메타데이터베이스를 검색하는 LSP 시스템<sup>22)</sup>이나 노와 등<sup>23)</sup>의 시스템과는 차이가 있다. LUNAR 시스템의 메타데이터베이스는 월석표본에 관한 일곱개의 메타영역으로 구성된 단순한 테이블 형태의 화일이며, REQUEST는 소규모의 비지베스통계 메타데이터베이스가 검색대상이고, PLANES는 비행기의 비행 및 보수에 관한 대규모의 관계형 메타데이터베이스를 검색대상으로 하고 있다. 이 시스템들은 대개 질문의 구문·의미분석, 논리적 질의어 생성, 메타데이터베이스검색, 해답의 생성의 네 단계로 구성되며, 해답은 단순한 데이터리스트, 또는 그래프나 테이블 형태로 출력된다.

LSP 시스템은 의료관계보고서의 텍스트를 테이블 형태로 구조화시켜 메타데이터베이스를 구성한 뒤 입력되는 질문을 논리적 질의어형태로 변환한 뒤 해답을 검색하는 시스템이다. 이 시스템은 앞의 시스템들이 단순한 메타항목으로 구성된 메타데이터베이스를 검색하는 것과는 달리 보고서의 텍스트를 질문분석과정에서와 마찬가지로 분석하여 메타데이터베이스를 구조화시켰다는 데 그 특성이 있다. 앞에서 언급한 노와 등의 시스템도 과학기술체통의 문헌텍스트를 의미네트웍 형태로 표현하여 메타데이터베이스를 구성하려고 시도하였다.

20) D.L. Waltz, "An English Language Question Answering System for a Large Relational Database," *Comm. of the ACM*, 21(7) : 526-539 ; 1978.

21) W.J. Plath, "REQUEST: A Natural Language Question-Answering System," *IBM J. of Res. & Dev.*, 20(4) : 326-335 ; 1976.

22) R. Grishman & L. Hirshman, "Question-Answering from Natural Language Medical Data Bases," *Artificial Intelligence*, 11 : 25-43 ; 1978.

23) Nowak, et. al., op. cit.

### 3. 전문가시스템

#### 3.1 개 요

전통적인 도서관과 기타 정보시스템의 사서 및 정보전문가는 모두 특정한 분야 내지는 업무에 관한 전문적인 지식을 갖추고 있는 전문가임에 틀림없다. 분류담당사서, 편목담당사서는 각각 분류, 편목에 있어서의 전문가이며, 색인·초록 서비스기관의 색인·초록자는 색인·초록 대상주제의 주제전문가이자 색인·초록기술 전문가인 것이다. 또한 참고사서나 정보전문가는 특정한 주제분야의 참고자료나 정보자료에 관한 전문가이며, 온라인탐색 전문가는 데이터뱅크와 최종이용자 사이에서 효과적인 정보탐색을 대신해 주는 전문가인 것이다.

도서관·정보학 분야에서는 일찌기 컴퓨터가 실용화된 이래 색인·초록·분류 등의 업무를 자동화하여 전문가로서의 기능을 수행하는 컴퓨터시스템을 개발하여 왔으나 전문가시스템이라는 용어는 사용하지 않았다. 대신 자동색인시스템, 자동초록시스템, 자동분류시스템이라고 불려온 것은 주지의 사실이다.

전문가시스템(expert systems)이란 용어는 인공지능을 이용하는 전산학 분야에서 사용되기 시작한 것으로서 “특정한 전문영역에서의 문제해결을 위해 인간전문가의 사고과정을 모방하여 만들어진 지식기반 컴퓨터시스템”이라고 정의한다.<sup>24)</sup> 이러한 전문가시스템은 가장 중요한 특성이 전문적인 지식을 이용한다는 점에서 그냥 지식기반시스템이라고 부르기도 한다.

전문가시스템에 관한 연구는 1964년 미국의 스탠포드대학교에서 시작되어 최초의 전문가시스템인 DENDRAL이 1965년 개발된 것으로 나타나 있다.<sup>25)</sup> 화학분야의 전문가시스템인 DENDRAL과 함께 초기의 전문가시스템으로 잘 알려져 있는 것은 의료진단분야의 MYCIN, 지질학분야의 PROSP

24) P.S. Sell, *Expert Systems-A Practical Introduction*. New York: Wiley, 1985. p.15.

25) *ibid.*, p.7.

ECTOR 등이다.

전문가시스템에 관한 연구개발은 초기 시스템 이후 최근 들어 급속히 확산되어 1985년경 그 수가 50여개를 넘는 것으로 보고되었으나,<sup>26)</sup> 1986년의 출판물에서는 중요한 전문가시스템만도 그 수가 100개에 달하며, 대상 분야 또한 다양한 것으로 나타나 있다.<sup>27)</sup>

도서관·정보학분야에서 전문가시스템이 본격적으로 논의되고 또한 응용되기 시작한 것은 1980년대에 들어서인 것으로 보인다. 1983년 데이비스(Davies),<sup>28)</sup> 클라크(Clarke)와 크로닌(Cronin)<sup>29)</sup>이 전문가시스템을 도서관·정보업무와 관련시켜 논의한 것을 시작으로 하여 1984년 *J. of ASIS*에는 인공지능에 관한 특징으로 전문가시스템에 관한 논문<sup>30)</sup>이 실렸으며, 1987년에는 *Information Processing and Management*가 '전문가시스템과 도서관·정보학'에 관한 특집호(v. 23, no. 2)를 내놓기에 이르렀다.

도서관·정보학분야의 전문가시스템은 1980년 이전에 간장염(hepatitis)에 관한 지식베이스의 개발이 시작된 후<sup>31)</sup> 다양한 영역에 걸쳐 개발되고 있다. 앞에서 언급한 사서 및 정보전문가가 수행하는 모든 업무가 전문가시스템의 대상이 되고 있으며, 이들 인간전문가를 대체하거나 도와줄 수 있는 컴퓨터시스템이 개발목적이 되고 있다.

### 3.2 전문가시스템의 구성

전문가시스템은 고도의 전문성이 요구될 정도의 어려운 문제를 해결하기 위한 것으로서 전문적인 지식과 추론절차를 사용하는 컴퓨터프로그램들로

26) *ibid.*

27) Hunt, *op. cit.*, pp. 37-39.

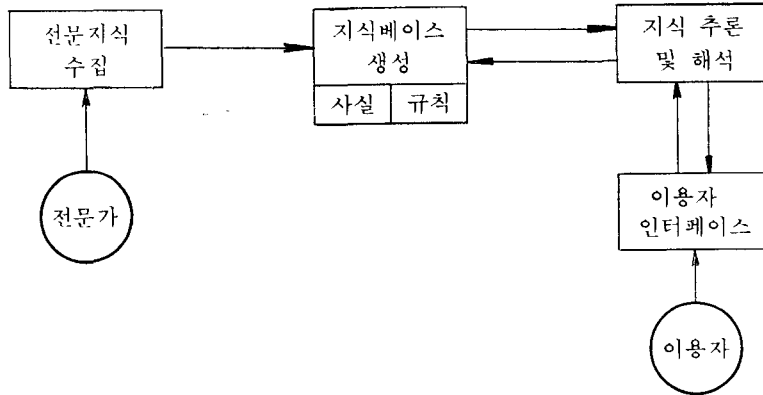
28) R. Davis, "Documents, Information or Knowledge? Choices for Librarians," *J. of Librarianship*, 15(1) : 47-65 ; 1983.

29) A. Clark & B. Cronin, "Expert Systems and Library/Information Work," *J. of Librarianship*, 15(4) : 277-292 ; 1983.

30) N.S. Yaghamai & J.A. Maxin, "Expert Systems: A Tutorial," *JASIS*, 35(5) : 297-305 ; 1984.

31) L.M. Bernstein, et al., "The Hepatitis Knowledge Base: A Prototype Information Transfer System," *Annals of Internal Medicine*, 43(1) : 165-222 ; 1980.

구성된다. 따라서 전문가시스템을 구성하는 가장 기본적인 요소는 지식을 담고 있는 지식데이터베이스(즉, 지식베이스)와 이 지식베이스를 이용하여 문제를 해결하는 추론프로그램이다. 여기에 이용자의 문제를 처리하기 위한 작업장으로서 이용자인터페이스장치가 추가된다. 아래 그림은 이러한 기본적인 요소를 보여주는 전문가시스템의 간단한 구성도이다.<sup>32)</sup>



위의 구성요소 가운데 전문지식의 수집은 가장 어려운 과정이라고 볼 수 있다. 사실(facts) 형태의 전문지식은 주로 해당분야의 전문가나 전문가의 출판물을 통해 입수하지만 실험적·판단적·상식적인 지식을 포함하는 'heuristic' 지식은 전문가들도 쉽게 정의하기가 힘들기 때문이다. 전문가시스템에서는 이렇게 지식을 수집하고 수집된 지식을 컴퓨터속에 적절히 표현하는(즉, 지식베이스를 생성하는) 작업과 추론과정을 설계하는 작업을 담당하는 사람을 '지식엔지니어'라고 부르고 있다. 결국 지식엔지니어는 도서관에서 주제별 정보자료의 수집을 담당하는 주제전문사서와 정보자료의 내용을 분석하여 검색이 용이하도록 적절히 조직하는 분류·편목·색인전문가에 해당하는 사람이라고 볼 수 있다.

지식추론·해석기는 지식베이스 관리시스템과 추론시스템으로 구성된다.

32) Hunt, op. cit., p.27.

지식베이스 관리시스템은 지식베이스의 조직·통제·갱신·탐색 등의 작업을 수행하며, 추론시스템은 지식베이스내의 지식을 이용하여 문제의 해결에 필요한 추론을 수행한다.

전문가시스템의 이용자 인터페이스는 이용자가 정보를 제공받는 일반적인 정보검색시스템과는 달리 다음의 세가지 이용자모드로 운영된다.<sup>33)</sup> 첫째는 문제에 대한 해답을 제공함으로써 이용자는 시스템의 ‘고객’이 되며, 둘째는 시스템의 전문지식을 개선 내지는 확장시킴으로써 이때의 이용자는 시스템의 ‘선생’으로서의 역할을 한다. 셋째는 지식베이스내의 지식을 이용하여 미래의 전문가를 교육시키는 것으로서 이용자는 ‘학생’으로서의 역할을 한다.

### 3.3 전문가시스템과 정보검색시스템

앞에서도 언급했듯이 전문가시스템의 개념자체는 도서관·정보학분야에서 전혀 새로운 것이 아니다. 실제로 언어학적 지식을 담아둔 지식베이스를 이용하여 자연언어질문을 처리하고, 질문에 대한 응답으로서 사실정보를 검색해내는 질문응답시스템은 지식베이스(언어학적 지식베이스와 전문주제영역의 사실정보 데이터베이스)를 이용한 시스템으로서 일종의 전문가시스템으로도 볼 수 있다. 실제로 노왁 등이 발표한 전문가시스템은 과학·기술분야의 전문적인 사실정보를 검색해내는 질문응답시스템으로서 저자들은 이러한 시스템을 새로운 세대의 정보검색시스템이라고 보고 있다.<sup>34)</sup>

비스와스(Biswas) 등이 개발한 지식베이스를 이용한 문헌검색시스템은 주제개념과 이들간의 관계로 구성된 지식베이스 및 서지정보 데이터베이스로 구성되며, 자연언어질문을 처리하는 정보검색시스템으로서 주제개념 및 이들 개념간의 관계에 관한 전문적인 지식을 이용하며, 추론기능을 갖는다는 점에서 일종의 전문가시스템의 범주에 속한다.<sup>35)</sup>

33) *ibid.*, p. 28.

34) Nowak, et al., *op. cit.*

35) Biswas, et al., *op. cit.*



정보검색시스템을 전문가시스템으로 볼 때의 지식베이스는 대체적으로 자연언어질문의 처리에 필요한 사전 등을 포함하는 언어학적 지식과 적절한 탐색어선정에 필요한 디소러스 등으로 구성되고 있다. 또한 추론기능도 검색기능과 탐색어확장기능, 그리고 검색된 문헌을 적합도에 따라 순위를 매기는 기능 등으로 구성되어 기존의 검색시스템과 크게 다를 바가 없다. 결국 이러한 정보검색시스템은 같은 내용에 새로운 이름을 붙인 것으로서 전문가시스템이라기 보다는 오히려 지식베이스를 이용하는 '지능형 정보검색 시스템'이라는 명칭이 보다 타당한 것으로 보인다.

#### 3.4 온라인탐색 중개시스템

도서관·정보학 분야에서 개발된 초기의 전문가시스템은 거의가 대규모 데이터뱅크의 온라인탐색을 위한 전단(front-end) 또는 중개(intermediary) 시스템인 것으로 나타나 있다. 즉 최종이용자가 데이터베이스 탐색기법에 관한 전문적인 지식이 없이도 단말기상에서 직접 데이터베이스를 탐색하도록 해주는 시스템으로서 최종이용자와 데이터뱅크간의 중개자인 탐색전문가의 역할을 대신하는 시스템인 것이다.

DIALOG, BRS, ORBIT, MEDLINE 등의 대규모 상업적인 데이터뱅크들은 정형의 질문언어를 사용하여 탐색할 수 있으며, 또한 다양한 탐색기법의 효과적인 사용여부에 따라 검색의 효율성이나 탐색비용면에서 차이가 생기게 된다. 따라서 정보요구를 갖고 있는 최종이용자가 탐색훈련을 받지 않은 채 직접 데이터베이스를 탐색하는 것 보다는 탐색전문가에 의한 대리탐색이 효율적이었다. 온라인탐색 중개시스템은 최종이용자가 탐색중개인을 거치지 않고 직접 온라인 탐색을 할 수 있도록 한다는 점에서 실용성이 매우 높은 시스템이라고 하겠다.

지금까지 개발된 온라인탐색전문가 시스템들은 시스템의 수준에 따라 제공하는 기능이 다양하며, 대략 다음과 같은 자동기능을 포함하고 있다.

- (1) 탐색문의 형성 및 탐색전략의 수행
- (2) 데이터뱅크와의 연결기능

- (3) 데이터베이스의 선택기능
- (4) 자연언어질문의 처리기능
- (5) 이용자와의 대화기능
- (6) 탐색문 및 탐색결과 출력형식의 변환기능(공통된 탐색언어에 의한 다수의 데이터뱅크에의 접근기능)

초기의 시스템으로 의학분야의 탐색 중개시스템인 Paperchase는 1981년에 발표되었고,<sup>36)</sup> 1982년 발표된 FRED는 훨씬 발전된 형태로서 메뉴방식과 자연언어 질문처리를 결합한 고도의 대화형 시스템이다.<sup>37)</sup>

MIT에서 개발되어 1981년 발표된 CONIT(Computer for Networked Information Transfer)<sup>38)</sup>과 1982년 발표된 상업적 전단프로세서(front-end processor)인 OL'SAM은 모두 MEDLINE, DIALOG, ORBIT 등 여러개의 데이터뱅크들을 하나의 공통된 언어를 통하여 탐색하도록 하는 온라인 중개시스템이라는 특성을 갖는다.<sup>39)</sup> CONIT, OL'SAM과 같은 중개시스템은 자기 다른 탐색언어(명령어 및 탐색문 형식)를 사용하는 기존 데이터뱅크의 이용자들이 각각 다른 언어를 익혀야 하는 불편함을 해소시켜주는 이용자편의시스템이라고 하겠다. 이러한 시스템들은 이용자로 하여금 한가지 언어만을 사용하도록 하고 이것을 각 데이터뱅크의 고유한 언어로 변환시킨다. 또한 탐색결과 역시 공통된 한가지 형식으로 출력시켜준다. 이러한 변환기능을 수행하는 중개시스템은 여러개의 데이터뱅크의 컴퓨터시스템에 대한 전단시스템으로서의 기능을 수행하는 것으로서 대개 TELENET, TYMNET 등의 부가통신망을 통해 원하는 데이터뱅크에 이용자를 연결시킨다.

1979년과 1982년에 각각 발표된 중개시스템인 IIDA(Individualized Instruction for Data Access System)는 다른 중개시스템과는 다소 다른 특성을

36) G.L. Horowitz & H.L. Bleich, "Paperchase: A Computer Program to Search the Medical Literature," *New England J. of Medicine*, 305(16) : 924-930 ; 1981.

37) M.I. Crystal & G.E. Jacobson, "FRED, a Front End for Databases," *Online*, 27-30 ; 1982.

38) Marcus & Reintjes, op. cit.

39) D.E. Taliver "OL'SAM: An Intelligent Front-end for Bibliographic Information Retrieval," *ITL*, 1(4) : 317-326 ; 1982.

갖는다.<sup>40)41)</sup> IIDA는 강의모드와 보조모드의 두가지 모드로 운영되며, DIALOG과 연결되어 실험되었다. 강의모드는 컴퓨터보조강의(CAI) 프로그램을 이용하여 이용자에게 DIALOG의 기본적인 탐색기법을 가르쳐주며, 또한 실제 배운 것을 실습해보도록 해준다. 보조모드에서 이용자는 DIALOG을 실제로 탐색하게 되는데 이 모드에서는 탐색문 구성상의 오류나 비효율적인 탐색방법을 지적하는 메시지가 이용자에게 출력된다.

MEDLINE 데이터베이스의 탐색중개시스템으로 개발된 CANSEARCH는 앞의 다른 중개시스템과는 달리 전문가시스템으로서의 원칙을 이 시스템의 설계 및 구현에 적극적으로 응용하였다는 점이 주목할 만하다.<sup>42)43)</sup> CANSEARCH는 의사들의 암치료에 관한 문헌탐색을 돕기 위해 만들어진 중개시스템으로서, 지식베이스는 MeSH(Medical Subject Headings) 디스커스와의 여러 종류의 규칙들로 구성되어 있다. 규칙들은 디스커스내의 개념·용어의 선택 및 출력규칙, 탐색문 생성규칙 등을 포함한다. 규칙들은 if-then 형식의 생산규칙으로 표현되었다. 실질적인 MEDLINE의 탐색어(즉, 컴퓨터내부의 메시지)는 블랙보드형태의 작업장소에 기록되는데 블랙보드는 메시지의 기록, 복사, 삭제 등의 기능을 위해 사용된다. CANSEARCH의 궁극적 목적은 규칙들을 사용하여 MEDLINE 데이터베이스 탐색에 적합한 탐색문을 생성해 내는 데 있다.

### 3.5 참고업무 전문가시스템

전통적인 의미의 참고사서는 도서관의 이용자가 원하는 정보를 담고 있는 참고자료에 대한 안내를 해주는 전문가이므로 이 전문가의 기능을 대신하는 전문가시스템의 개발은 극히 당연한 응용이라고 보겠다. 최근에 와서는 참고사서가 수행하는 기능이 크게 확장되어 기본적인 참고자료에 대한 안내

40) C.T. Meadow, "The Computer as a Search Intermediary," *Online*, 3(3) : 54-59 ; 1979.

41) C.T. Meadow, et al., "A Computer Intermediary for Interactive Data Base Searching. 1. Design," *JASIS*, 33(5) : 325-332 ; 1982.

42) Politt, op. cit.

43) A.S. Pollitt, "A Front-end System: An Expert System as an Online Search Intermediary," *Aslib Proceedings*, 36(5) : 229-234 ; 1984.

이외에 정보자체의 제공과 정보를 알려줄만한 전문가·전문기관에 대한 안내까지 포함하고 있다. 또한 온라인 탐색전문가로서의 기능까지 수행함으로써 단순한 참고자료에서 벗어난 전문적인 보고서나 논문에 대한 서지를 제공해주고 있는 것은 주지의 사실이다.

미국 국립농학도서관에서 개발한 Answerman이라는 참고업무 전문가시스템은 전문가시스템 개발용 소프트웨어패키지(shell이라고 함)를 이용하여 IBM/PC 상에 구현시킨 것으로서 비교적 간단한 시스템이다.<sup>44)</sup> Answerman은 이용자의 참고질문에 대한 응답으로서 각 질문의 해답이 들어 있을 참고도서를 알려준다. 이 시스템은 외부 데이터뱅크에 온라인으로 연결함으로써 참고도서 이외의 자료의 서지정보나 전문정보를 제공할 수 있도록 확장이 가능하다. Answerman의 지식베이스에는 정보유형(지도, 서지, 사전, 신문 등 16종류), 주제, 직업/조직체 등에 관한 세가지 종류의 정보가 변수로 들어 있고, 각 변수의 조합에 따른 적절한 참고도서의 코드가 해답으로 소장되어 있다. 예를 들어 정보유형이 '디렉토리', 주제는 '정치학', 직업이 '국회의원'으로 입력이 되면 해답으로 출력되는 참고도서는 *Congressional Directory*가 된다.

영국에서 개발된 PLEXUS는 공공도서관에서 사용하도록 만든 참조(referral) 전문가시스템이다.<sup>45)</sup> 이 전문가시스템의 개발목적은 사서와 최종이용자 모두에게 도움을 주기 위한 것이며, 참조정보가 필요한 상황에서 인간참고사서가 수행하는 가장 지적인 업무를 모방하도록 하였다. PLEXUS가 선택한 주제영역은 비교적 공공도서관 이용자의 정보요구가 많은 원예분야로 한정하였다. 이 시스템의 중요한 특성은 고도의 대화형시스템이라는 점이다. 입력된 이용자의 원질문은 시스템에 의해 탐색개념이 추가되거나, 시스템이 이용자에게 질문에 답하게 함으로써 보완된 탐색문으로 변환된다. 탐색결과는 시스템과 이용자에 의해 평가되며, 초기의 탐색전략이 만족스럽

44) S.T. Waters, "Answerman, the Expert Information Specialist: An Expert System for Retrieval of Information from Library Reference Books," *ITL*, 5(3): 204-212; 1986.

45) A. Vickery & H.M. Brooks, "PLEXUS-The Expert System for Referral," *Inf. Processing & Management*, 23(2): 99-118; 1987.

지 못한 결과를 가져왔을 때는 적절한 행위를 수행하도록 하였다.

PLEXUS의 지식베이스는 참고사서가 갖고 있는 전문지식을 담고 있으며, 여기에는 다음과 같은 여러가지 유형의 지식이 포함된다. 즉, 도서관학/도서관에 대한 지식, 정보검색기법에 관한 지식, 주제영역에 관한 지식, 참고자료에 관한 지식, 지식전반의 구조에 관한 지식, 도서관 이용자에 관한 지식, 탐색문형성에 관한 지식 등이다. 이 지식베이스내의 지식은 여러명의 전문가와 정보학문헌으로부터 수집되었으며, 지식들은 자기 적절한 지식표현기법에 의해 소장되는데 생산규칙, 프레임구조, 의미네트웍 등이 두루 이용되었다.

### 3.6 분류·편목 전문가시스템

정보자료의 분류기호매김과 주제명의 부여, 그리고 목록의 기술은 전통적으로 사서가 수행하는 지적인 작업으로 인식되어 왔다. 그러나 컴퓨터의 대중화에 따라 도서관분야에서도 이러한 지적인 작업을 컴퓨터가 대신하도록 해보려는 시도가 시작되었다. 그 결과 실험적인 자동분류, 자동색인, 온라인목록의 생산 등의 새로운 기술이 개발되었던 것이다. 그러나 이 단계의 시스템들이 거의가 컴퓨터기술의 단순한 응용에 불과하다고 본다면 다음 단계는 인공지능의 개념을 이용한, 즉 컴퓨터에 내장된 전문지식을 이용하여 지적인 작업을 수행하는 전문가시스템으로의 발전이 될 것으로 보인다.

자동분류는 정보자료 텍스트에 나타난 주제어를 선택한 후에 수행되는 작업이므로 실제로는 자동색인의 다음 단계라고 볼 수 있다. 다시 말해 텍스트에 어떠한 단어가 어떠한 관계로 출현했는가에 따라 자료의 주제분류가 가능하기 때문이다. 지금까지의 자동분류에 관한 연구는 특정한 주제에 관한 분류항목과 이에 대응하는 주제어들을 단서로 하여 문헌을 특정항목에 분류하는 유형과 문헌들이 갖는 주제어의 유사도에 따라 같은 집단으로 분류하는 문헌집단화(클러스터링)로 구분된다. 모두가 주제어에 기초하여 문헌과 분류항목, 또는 문헌과 문헌간의 유사도를 측정하는 과정에 의해 분류하였다.

전문가시스템으로의 분류시스템은 실제로 목록에 기입되는 분류기호의 선택에 사용되는 분류표의 내용을 지식베이스에 담고 있어서 일련의 규칙에 의해 어떤 경우에 어떤 분류기호를 선택할 것인지를 결정하거나 도와주는 시스템이 될 것이다. 예를 들어 '도서선택'이나 '수서' 등의 단어가 특정한 수식어구가 없이 표제에 나타난 경우에는 if-then 규칙(If 도서선택 or 수서 then 025.2)에 의해 025.2를 분류기호로 부여하며, 앞의 주제어 이외에 특정한 주제를 나타내는 단어가 나타나 있는 경우에는 025.27을 부여하도록 하는 것이다.

온라인목록은 편목전문가가 단말기에 앉아 스크린에 출력되는 형식에 맞춰 일일히 목록데이터를 입력시키며, 이용자는 다양한 접근점에 의해 목록 레코드를 찾아볼 수 있도록 한 것이므로 목록의 작성은 여전히 인간전문가에 의해 이루어진다. 편목전문가는 AACR(영미목록규칙)과 같은 목록규칙을 도구로 사용하여 기본기입·부출기입과 같은 접근점을 선택하고 저자명·판용표제 등의 목록데이터는 기술규칙에 따라 일관성 있게 기술하여야 한다.

편목전문가시스템은 분류전문가시스템과 마찬가지로 목록도구인 목록규칙을 지식베이스로 만들고 각 경우(조건)에 맞는 규칙(if~then 규칙)을 찾아 적절한 목록규칙을 적용하는 것이다. 편목전문가시스템은 첫 단계에서는 인간 편목전문가를 보조하는 도구로 사용할 수 있을 것이다. 즉 단말기에 데이터 입력양식만을 출력하는 것이 아니라 지식베이스내의 목록규칙을 이용하여 해당되는 조건(예를 들면 저자의 경우 익명의 저자, 공저자, 전집 등)을 이용자가 선택하던 각 경우의 표목형식을 자동으로 선택하여 준 뒤 해당 데이터를 채워넣게 하거나, 또는 출판지나 출판사를 모르는 경우에는 자동으로 s.l.이나 s.n.이 해당되는 항목을 채우도록 하는 것이다. 이러한 전문가시스템의 목록데이터는 프레임구조로 표현하는 것이 용이하며, 목록규칙은 if-then 규칙형식으로 표현한다.

한 단계 더욱 발전한 전문가시스템에서는 각 목록데이터를 직접 입력하는 대신 표제지를 패턴인식 등의 방법으로 읽어들이고 후 자동으로 저자, 표제, 출판사 등의 데이터를 식별하도록 하는 것이다. 이때에는 표제지를 제대로

인식하는 데 필요한 규칙들이 컴퓨터에 소장되어야 할 것이다.

1987년 현재 미국의 UCLA에서 개발중인 MAPPER는 AACR2를 사용하여 지도를 편목하는 전문가시스템이다.<sup>46)</sup> 이 시스템의 개발목적은 편목사서가 AACR2의 적절한 규칙을 선택하여 MARC 형식에 맞게 지도를 편목하도록 보조하는 컴퓨터시스템을 설계하는 것으로 되어 있다. 이 시스템의 구현을 위해서 IBM/PC에 사용가능한 셸(shell)인 EXSYS라는 패키지를 선택하였고, 전체적인 시스템은 이용자 인터페이스모듈, 지식베이스모듈, 추론모듈, 설명모듈로 구성되도록 설계하였다.

### 3.7 색인·어휘통제 전문가시스템

색인·어휘통제 전문가시스템은 자동색인시스템과 자동탐색확장시스템의 새로운 이름에 지나지 않는다. 전통적으로 색인전문가에 의해 수행되던 색인작업(통제언어색인에서는 어휘통제작업을 포함함)을 컴퓨터가 수행하도록 하려는 시도는 1957년 룬(Luhn)의 연구<sup>47)</sup> 이후 계속되어 왔다. 텍스트에 출현한 단어의 빈도를 이용하는 통계적 기법 이외에 언어학적 분석을 도입한 기법이 제시되었으나 큰 성과를 가져오지는 못하였다. 현재의 DIALOG 등의 상용검색시스템에서는 대부분 기능어를 제외한 나머지 단어를 색인으로 선택하는 불용어제거기법이 가장 많이 사용되고 있다. 어휘통제를 통해 디스크립터를 색인으로 부여하는 통제언어색인은 여전히 인간전문가에 의해 수행되고 있다. 어휘통제를 자동으로 실행하는 시스템에서는 기계가독형 디소러스나 동의어사전 등을 통해 어간이 동일한 단어의 동의어처리, 유사한 개념을 갖는 단어의 표준화 등의 작업을 수행하고 있다.

어휘통제 기능은 탐색어확장에도 이용된다. 유사한 개념의 단어들을 한 개의 디스크립터로 변환시킨 뒤 탐색한다든가 유사한 개념어 아래 색인된 문헌들을 모두 검색하는 등의 동의어 통제를 통한 탐색어확장과, 개념의 계

46) H. Borko, "Getting Started in Library Expert Systems Research," *Inf. Processing & Management*, 23(2) : 81-88 ; 1987.

47) H.P. Luhn, "A Statistical Approach to Mechanized Encoding and Searching of Library Information," *IBM J. of Res. & Dev.*, 1(4) : 309-317 ; 1957.

총관계 및 연관관계를 이용한 탐색어확장이 있다.

그렇다면 이러한 자동색인·어휘통제시스템과 전문가시스템의 차이는 어디에서 찾을 것인가? 첫번째 차이는 사용하는 용어에서 나타난다. 즉 디소러스라는 용어 대신 지식베이스라는 용어를 사용한다든가, 자동색인 대신 지식기반색인(knowledge-based indexing)이라는 용어가 사용되고 있다. 두 번째의 차이는 색인어나 탐색어의 선택에 있어서 추론기능을 이용하는 데 있다. 단순히 디소러스·사전 등에 들어 있는 단어와의 대조기능에서 한 단계 더 나아가 의미가 모호한 단어의 처리라든가 하나의 단어가 특정한 문맥에 나타날 경우의 처리방법이 추론에 의해 결정된다. 이것이 가능하기 위하여 지식베이스에는 단순한 디소러스·사전의 수준을 넘어서는 보다 구체적인 용어간의 개념관계와 추론규칙이 정의되어야 할 것이다.

장재경의 논문에서는 사전지식베이스를 용어 및 이들 용어간의 개념관계로 구성하였으며, 문헌을 분석한 결과의 지식베이스(정보검색 용어로는 색인화일에 해당함)는 색인어 및 이들 색인어간의 구문적 관계로 구성하였다.<sup>48)</sup> 위의 지식베이스는 의미네트웍 형태로 표현하였으며, 탐색시 추론기능에 의해 이용자의 질의어를 확장하도록 하였다.

의학분야 문헌의 색인을 위해 최근개발중인 지식기반색인시스템에서는 프레임구조를 이용하여 지식베이스를 구성하였다.<sup>49)</sup> 이 색인시스템은 MEDLINE 색인자들로 하여금 각 색인대상 문헌에 대해 특정한 유형(신체부위, 절차, 증세 등)의 주제어를 각각 해당되는 프레임슬롯에 채워넣도록 한 뒤 추론기능에 의해 부표목을 갖는 완전한 MeSH 표목으로 변환하도록 하였다.

#### 4. 결 론

도서관·정보학 분야에서는 일찍이 1950년대부터 자동색인과 자동초록의

48) 장재경, 우리말 문헌정보검색을 위한 지식베이스설계에 관한 연구. 연세대학교 대학원 석사학위논문, 1985.

49) S.M. Humphrey & N.E. Miller, "Knowledge-based Indexing of the Medical Literature: The Indexing Aid Project," *JASIS*, 38(3) : 184-196 ; 1987.



작성을 위해 컴퓨터를 이용하였으며, 이것은 자연언어로 쓰여진 텍스트를 처리하고자 하는 시도인 동시에 색인전문가나 초록전문가가 수행하던 지적 작업을 컴퓨터가 대신하도록 한 것으로서 실질적으로는 이때부터 인공지능의 응용이 시작되었다고 할 수 있다. 그러나 도서관·정보학에서 인공지능을 본격적으로 응용하기 위한 논의는 1980 년대에 들어서부터이며, 이제는 국내에서도 이에 관련된 연구가 시작되고 있다. 도서관·정보학에서의 인공지능의 응용은 이 분야와 관련된 전문적인 사서 및 정보전문가의 역할의 변화를 요구하고 있으며, 이에 따른 관련학과의 교과과정의 수정도 불가피한 단계에 이르렀다.

자연언어처리가 가능한 정보시스템과 각종 전문가시스템의 개발은 색인·초록전문가, 온라인 탐색전문가, 편목·분류전문가, 참고사서(주제전문가) 등의 기존 업무를 크게 감소시킬 것으로 보인다. 하지만 의료진단 전문가시스템이 의사를 대신할 수 없는 것처럼 사서 및 정보전문가를 전문가시스템이 완전히 대신하리라고는 예측할 수 없다. 그러나 이러한 전문가 시스템들은 사서 및 정보전문가를 보조하는 시스템 이상으로서의 역할을 할 수 있을 것이 거의 확실하다.

그렇다면 앞으로의 사서 및 정보전문가는 기존의 업무이외에 보다 지적인 업무에 종사할 시간과 필연성을 갖게 될 것이 분명하지 않은가? 전문가시스템이나 자연언어처리가 가능한 지식기반시스템의 핵심이 되는 요소가 지식이며, 이 지식의 수집이나 표현은 전통적으로 도서관·정보학의 가장 중요한 주제가 되어왔음을 생각하면 이제 이들이 해야 할 일은 명확해진다. 즉, 사서 및 정보전문가들은 지식기반시스템의 지식베이스를 제작하고 활용시키는 '지식엔지니어'로서의 변신을 시도해야 할 것이다. 따라서 도서관·정보학을 전공하는 이들의 의식변화와 더불어 교과과정의 적극적이고 실제적인 내용의 변화가 요청된다.

## Artificial Intelligence Applications in Library and Information Science

Young Mee Chung

In this paper, artificial intelligence applications in library and information science are reviewed. Especially, natural language processing and expert systems are represented as the two major application areas.

In natural language processing, natural language interface systems and question-answering systems are discussed in detail with some specific examples. In the second part of the paper, online search intermediary systems, reference expert systems, classification and cataloging expert systems are described as possible expert systems to be developed in libraries and information systems. As a conclusion, implications of the artificial intelligence applications for librarians and information scientists are suggested.