

패턴인식기법을 이용한 편목전문가시스템 설계에 관한 연구*

A Study on Design Of Cataloging Expert System Using Pattern Recognition Techniques

김현희(Hyun-Hee Kim)** 곽병희(Byung-Hee Gwak)***

□ 목 차 □

1. 서론	2.2.1 개념	3.4 알고리즘 설계
1.1 문제의 제기	2.2.2 지식표현기법	3.4.1 서지요소 자동인지 알고리즘
1.2 연구 목적과 필요성	2.3 편목 전문가시스템	3.4.2 편목 알고리즘
1.3 연구 내용과 방법	2.3.1 편목업무와 전문가 시스템	4. 편목전문가시스템 평가
1.4 연구를 위한 가정	2.3.2 선행연구	4.1 실험 데이터
1.5 연구 범위와 제한점	3. 편목전문가시스템 설계	4.1.1 실험문헌집단
2. 이론적 배경	3.1 시스템의 개관	4.1.2 검증문헌집단
2.1 패턴인식기법	3.2 지식베이스 설계	4.2 지식베이스
2.1.1 패턴인식의 개념	3.2.1 패턴인식 지식베이스	4.2.1 실험문헌집단
2.1.2 문자인식의 방법	3.2.2 편목 지식베이스	4.2.2 검증문헌집단
2.1.3 오프라인 문자인식 방법	3.2.3 전거화일	4.2.3 논 의
2.2 전문가시스템 개요	3.3 지식획득시스템	5. 결 론

초 록

본 연구에서는 표제면과 판권지의 서지요소의 레이아웃 특성과 구문적 특성을 이용하여 서지요소의 종류를 패턴인식 지식베이스와 전거화일들을 이용하여 자동 인지하고 인지된 서지요소를 한국문헌자동화목록형식(KORMARC)과 한국문헌자동화목록법(KORMARC) 기술규칙에 기초하여 KORMARC 형태로 출력해 주는 편목전문가시스템을 구축하였다. 이 시스템의 성능을 평가하기 위해서 패턴인식 지식베이스의 생성을 위해 분석한 155종의 실험문헌집단과 86종의 검증문헌집단을 이용하여 적중률을 조사해 보니 실험문헌집단의 경우는 94%, 검증문헌집단의 경우는 93%의 적중률을 나타냈다.

ABSTRACT

This study shows the design and implementation of cataloging expert system using pattern recognition techniques. This system attempts to demonstrate the feasibility of cataloging in KORMARC format from title page and copyright page without the intervention of humans. The prototype was implemented as a rule-based system in Turbo C. To demonstrate the function and capability of the system, experimental document-group and control document-group was analyzed. The hit ratio of experimental document-group is 94%. On the other hand, the hit ratio of control document-group is 93%, a little bit lower than the experimental group.

* 이 논문은 1994년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음(중간보고서).

** 명지대학교 문헌정보학과 *** 서울교육대학 도서관

■ 논문제출일 : 1994년 11월 3일

제 1 장 서 론

1.1 문제의 제기

도서관 편목 업무에서 최근 관심을 끌고 있는 주제는 도서관 자료의 편목자동화에 대한 가능성의 문제이다. 편목자동기법(cataloging automatic techniques)은 전통적인 편목 대상인 단행본의 주제 편목(subject cataloging)과 기술 편목(descriptive cataloging)에서 뿐만 아니라, 학술지 논문, 팸플렛, 전자우편 등의 자료들이 검색될 수 있는 접근점을 결정하는 데 실제적인 유용성을 지니고 있다.

자동 기술 편목에 관한 대다수의 프로젝트들이 표제면으로부터 문헌의 접근점을 자동적으로 추출해 내는 것과 관련이 있다. 이러한 프로젝트들 중 하나가 Weibel 등(1989)에 의해 연구된 것으로서, 이것은 표제면 상에 나타나 있는 서지 요소들을 자동적으로 확인하는 프로토타입 규칙 기반 시스템을 개발하였는데, 이것은 인간의 간섭없이 표제면에 나타난 서지정보로부터 기술편목의 가능성을 예시하였다. 또한 Jeng(1991a)은 표제면에 나타난 서지요소의 종류를 서지요소의 위치, 문자의 유형과 크기, 공간, 구두점과 같은 레이아웃 특성과 구문적 특성을 이용하여 서지요소를 자동 식별하는 패턴 인식기법에 의한 편목 전문가시스템의 가능성을 제시하였다.

본 연구에서는 표제면과 판권지의 서지요소의 레이아웃 특성과 구문적 특성을 이용하여 서지요소의 종류를 자동인지하고 인지된 서지요소를 한국문헌자동화목록형식(KORMARC)과 한국문헌자동화목록법(KORMARC) 기술규칙

에 근거하여 KORMARC 형태로 자동편목해주는 편목전문가시스템의 가능성을 제시하고자 한다.

1.2 연구목적과 필요성

국가 기간 전산망 사업의 일부분인 교육/연구 전산망사업의 추진으로 구축될 서지 네트워크의 분담 편목은 편목 작업을 좀 더 신속하고, 효율적으로 만들어 가게 될 것이다. 그러나 이러한 발전에도 불구하고, 편목은 기술을 요구하는 자원집약적 활동으로 남아있으며 더욱이 편목자가 점차 감소하고 있는 추세이다. 따라서 도서관 서비스를 향상시키기 위해서는 편목 작업을 자동화시키기 위한 노력이 무엇보다 중요하다 하겠다.

지금까지 연구개발된 편목 전문가시스템은 사람의 지적인 업무처리 과정을 전문가시스템과 나누어 수행하는 인간-기계간의 인터페이스에 의한 방법과 표제면을 패턴인식 등의 기법으로 읽어들이고 후 자동으로 표제, 저자, 출판사항 등의 데이터를 식별하는 방법 등 두 가지 유형이 있다.

본 연구에서 구축할 시스템은 두번째 유형으로 전자문헌은 물론 인쇄물 형태의 문헌 표제면과 판권지에 수록된 서지요소의 종류를 패턴인식기법에 의해 자동 인지하고 인지된 서지요소를 KORMARC포맷으로 출력하는 시스템을 구현하고자 한다. 이 시스템은 복잡하고 까다로운 편목업무의 내용을 지식베이스로 구축하여 활용케 함으로써 편목의 정확성과 일관성을 유지시키는데 기여할 것이다.

1.3 연구 내용과 방법

본 연구에서 설계할 시스템은 문헌 표제면과 판권지의 서지요소의 종류를 패턴인식기법에 의해 판독하고 판독된 서지요소를 KORMARC 포맷에 맞춰 출력하는 시스템이다. 먼저 서지요소를 자동판독하기 위해서 18개의 서지요소의 종류를 구분할 수 있는 패턴인식 지식베이스와 전자화일들을 구성하였다. 패턴인식 지식베이스는 12개의 서지요소¹⁾를 인지하는데 사용되었고 나머지 6개의 서지요소²⁾는 전자화일 및 간단한 알고리즘에 의해 인지하였다.

패턴인식 지식베이스는 크게 레이아웃 규칙(물리적)과 구문적 규칙으로 구분할 수 있다. 레이아웃 규칙이란 문자의 유형(한글 명조체, 한글 고딕체)과 문자의 크기, 데이터의 위치를 규정하여 서지 요소를 판독하는 규칙이다. 한편 구문적이라는 것은 표제면 상에서 저자명 다음에 오는 저작 역할어(저, 공저, 역) 등과 같은 어휘 단위를 이용하여 서지요소를 판독하는 규칙이다.

다음은 서지요소의 종류가 확인되면 이 데이터를 기초로 하여 KORMARC포맷³⁾에 맞춰 출력해 줄 수 있는 편목 지식베이스를 구성하였다. 이때 이용된 지식베이스는 한국문헌자동화목록형식(KORMARC)과 KORMARC 기술규칙에 기초하여 구성하였다.

시스템의 성능을 평가하기 위해서 패턴인식 지식베이스를 구축할 때 참조한 155종의 실험 문헌집단과 86종의 검증문헌집단을 이용하였다. 본 시스템에서 구축한 패턴인식 지식베이스와 편목 지식베이스는 생성 규칙을 이용하여 지식을 표현했으며 추론 방법은 전진추론을 이용하였다. 시스템을 운영하는 모든 프로그램은 Turbo-C언어를 이용하여 작성하였다.

1.4 연구를 위한 가정

이 연구에서 설정된 한 가지 가정은 제안된 알고리즘이 적용되는 표제면과 판권지가 기계가독형 형태(machine-readable format)로 존재해야 한다는 것이었다. 그러한 기계가독형 형태는 지금은 출판 과정에서 일반적으로 이용되고 있는 컴퓨터 식자(computer typesetting)의 부산물로 데이터베이스가 생성되므로 이 화일을 바로 이용하거나 인쇄물인 경우에는 문자인식시스템을 통해 기계가독형 형태로 변환시켜야 한다.

1.5 연구 범위와 제한점

본 연구에서 분석한 문헌들은 국내서 단행본으로 국한하였으며 단행본 중에서 총서명이 있는 경우는 하위총서명, 본총서의 편(part) 번호 및 서명이 있는 문헌은 제외시켰으며 회

1) 관제, 판정, 본서명, 부서명, 감제, 별서명, 대등서명, 원서명, 총서명, 총서번호, 권차서명, 판사항.

2) 발행사, 발행지, 저자명(단체명), 저작역할어의 서지요소는 전자화일들을 이용하여 인지하였고 발행년은 알고리즘(네개의 숫자로 구성되고 문자크기가 가장 크지 않은 단어를 발행년으로 인지)에 의해 인지하였고 또한 저자명(개인명)은 저작역할어를 활용하여 인지하였다.

3) 다음과 같은 표시기호(tag)들, 즉, 1xx(기본표목), 245(서명저자사항), 250(판사항), 260(발행사항), 4XX(총서사항), 500(일반주기), 507(원저자, 원서명에 관한 주기), 7xx(부출표목), 830(총서부출표목)의 데이터를 출력해 준다.

의명이 기본표목인 경우도 역시 제외시켰다. 또한 본 연구에서 사용한 문자인식시스템은 한 가지 문자유형과 크기로만 인식되기 때문에 원본을 그대로 재생시켜 주기 위하여 편집과정을 거쳐야 하는 제한점이 있으며, 분석할 문헌의 표제면과 판권지에 있는 문자열의 속성(문자의 유형, 크기 등)들을 추출하는데 패턴인식 프로그램에 의해서 자동인식되어야 하는데 본 연구에서는 수작업으로 수행하였다.

제 2 장 이론적 배경

2.1. 패턴인식 기법

2.1.1 패턴인식의 개념

사람은 태어나면서부터 많은 사물과 책을 접하며 체계적으로 교육을 받게 된다. 이러한 경험을 토대로 우리는 어떤 물체나 글자를 볼 때, 그 일부만 보아도 무엇인지 대충 알 수 있으며, 심지어는 자주 봐왔던 사람의 걷는 모습만 보고도 누구인지 짐작할 수 있다. 이처럼 학습과 경험을 거쳐 머리 속에 들어 있는 모든 사항을 조합하여 자신이 보고 있는 것이 무엇인지 판단하는 과정을 패턴인식(pattern recognition)이라고 한다(진성일, 임길택 1994). 패턴인식에서 패턴은 사람이 인식하려는 정보를 물리적으로 표현한 대상으로 정의할 수 있으며, 인식은 대상물의 존재를 아는 인지(cognition)의 과정에서 한걸음 더 나아가 구체적인 의미나 정보를 판정하는 것을 의미한다

(오영환, 1991). 오늘날 컴퓨터 과학이 발전함에 따라 사람들은 이러한 인식 과정을 컴퓨터에게 접목시키려는 시도를 하게 되었다.

2.1.2 문자인식의 방식

문서정보의 자동 입력에서는 사람이나 인쇄매체에 의해 작성된 문서를 카메라나 스캐너를 이용, 영상 형태로 컴퓨터에 입력하고 이를 분석하여 비문자 영역은 영상 압축 기법으로 그림 형태로 만들어 저장하고, 문자영역은 문자인식을 수행한 후 텍스트 화일로 저장된다. 문자인식은 문자 영상 정보를 얻는 방식에 따라 온라인(On-Line)인식과 오프라인(Off-line)인식으로 나뉘어진다.

온라인 인식에서는 디스플레이 테블릿(Tablet)⁴⁾ 위에 글자를 씌고 동시에 입력되는 글자의 순서 정보를 이용하여 인식한다. 오프라인 인식에서는 이미 필기가 되어 있는 필사본 또는 인쇄된 문서를 스캐너와 같은 영상 입력 장치로 입력한 후 문서 영상 내의 글자를 인식한다. 여기에서 대상 문자가 인쇄체인가 필기체인가에 따라 인쇄체 문자인식과 필기체 문자인식으로 나뉘어진다(진성일, 임길택 1994). 본 연구에서는 오프라인 문자인식방법에 의하여 표제면과 판권지에 있는 문자를 인식한다.

2.1.3 오프라인 문자인식 방법

오프라인 문자인식방법은 인식방법에 따라 원형 비교방법, 통계적 방법, 구조 분석방법,

4) 전자펜으로 글씨를 쓰거나 그림을 그리면 펜촉(x,y)의 이동을 추적하여 그 궤적을 따라 전자인크로 그려줄 수 있게 한 장치

신경망을 이용한 방법 등으로 나눌 수 있다.

원형 비교방법은 입력된 이미지의 영상을 저장되어 있는 모든 표준 패턴과 최소 단위로 비교하여, 같은 속성을 갖는 화소의 비율이 가장 높은 패턴을 선택하는 방법이며, 통계적 방법은 문자 영상의 통계적인 특징을 분석하는 방법으로 입력문자 영상으로부터 특징 벡터를 추출한 다음, 판별함수를 이용하여 특징 벡터가 특징 공간상에서 어느 부류에 속하는가를 알아냄으로써 입력패턴을 분류하는 방법이다. 구조분석의 방법은 문자의 구성원리에 입각하여 입력 문자를 구성하는 기본요소와 그들간의 연관성을 추출하여 문자를 인식하는 방법이며 (이성환, 1994), 신경망을 이용한 방법은 인지 과학적인 접근방법의 하나로서 인간의 두뇌구조를 분석하고, 그 처리 메카니즘을 규명, 그 구조를 컴퓨터로 모델링하는 방법인데, 최근에 들어와서 신경회로망의 학습 과정을 통해 문자 인식에 필요한 정보를 습득하는 연구가 많이 이루어지고 있다. 신경회로망은 패턴의 국부적인 잡음이나 손실 등으로 변형된 입력이 들어와도 이를 안전하게 처리할 수 있고, 신경회로망 자체의 구조적 손실로 인한 어려움도 잘 대처하는 결합 극복 능력이 있어 이러한 신경회로망을 이용한 문자인식에 관한 연구가 매우 활발히 연구되고 있다(권재욱, 1994).

2.2 전문가시스템 개요

2.2.1 개념

전문가시스템의 역사는 오래되지 않아 그 명확한 정의를 내리기 어려운 실정이고, 여러

학자들이 사용하는 용어와 정의에 있어서 통일성의 결여는 전문가시스템 연구를 하는데 직면하는 문제 가운데 하나로 대두되고 있다. 그 공통된 개념을 추출해 보면 전문가시스템이란 특정영역에서 잘 훈련된 인간전문가의 지식을 체계적으로 조직하여 컴퓨터에 그 지식을 이식시켜 인간전문가의 업무를 보조 또는 대신할 수 있을 정도의 업무수행 능력을 가진 일군의 컴퓨터 프로그램이라고 할 수 있다. 전문가시스템은 크게 추론엔진과 지식베이스로 분류하며, 전체적인 시스템은 지식베이스(knowledge base) 모듈, 추론엔진(inference engine) 모듈, 지식습득(knowledge acquisition) 모듈, 설명(explanation) 모듈, 사용자 인터페이스(user interface) 모듈의 다섯가지 구성요소로 분류된다.

2.2.2 지식표현기법

주어진 갖가지 문제, 혹은 그 문제해결을 위하여 유용한 정보나 지식을 컴퓨터상에서 어떻게 표현하는가 하는 지식 표현의 문제는 인공지능 연구에서 기본적인 문제로 인식되고 있다. 전문가시스템의 지식베이스를 구축하기 위해서 인간 전문가나 문헌 등을 통해 이끌어낸 지식이 컴퓨터상에 생성규칙, 의미망, 프레임, 기계학습에 의한 자동 유도규칙, 신경망 등의 형태로 표현하고 있다. 여기서는 가장 많이 이용되고 본 연구에서 사용한 생성규칙과 기계학습에 의한 신경망에 대해 설명한다.

1) 전통적인 기법-생성규칙

생성규칙은 가장 널리 이용되고 있는 지식 표현 기법이다. 먼저, If-Then 형식으로 규칙

을 표현하는 생성규칙은 斷言的 규칙과 確率的 규칙으로 구분해 볼 수 있다. 먼저, 정해진 대출 규칙을 단언적 규칙으로 표현한 예를 들어 보면 다음과 같다.

If 대출자가 도서를 반납하는데 있어서 날짜를 어기면

Then 그(녀)는 하루 늦는데 X원을 벌금으로 지불해야 한다.

다음은 추천도서가 이용자의 요구에 만족될 가능성을 고려하여 0-1 사이의 확신률을 부여한 확률적 규칙이다.

If 주제 = 교육학

and 정보 유형 = 학술지 논문

and 초록이 요구된다.

Then 추천도서는 'Current Index to Journals in Education' 이다(확신률:0.9).

If 주제 = 교육학

and 정보 유형 = 연구보고서

and 초록이 요구된다.

Then 추천도서는 'Research in Education' 이다(확신률:0.8).

If 주제 = 인공지능

and 정보 유형 = 연구보고서

and 초록이 요구된다.

Then 추천도서는 'Abstracts in Artificial Intelligence' 이다(확신률:0.9).

특히 이러한 확률적 규칙은 경험적, 불확실한 지식에 기초하여 어떤 결론을 이끌어 내야 하는 경우에 적합하다.

2) 기계학습-신경망

이제까지 사용된 기술은 지식 공학자로 하여금 인간의 지식을 생성규칙, 프레임 등 명백한 지식표현으로 변환시키는 것을 요구한다. 그러나 지식공학자가 인간 전문가로부터 지식을 이끌어 내는 것은 종종 어렵고 시간이 많이 드는 작업이며 또한 인간 전문가들이 자신의 경험적 지식을 쉽게 If-Then 진술문 등으로 표현하기도 쉽지 않는 일이다. 이러한 어려움을 극복하는 방법은 간단한 예제 데이터를 입력물로 하여 컴퓨터로 하여금 학습하도록 프로그래밍하여 출력물을 얻는 것이다. 이러한 기계 학습을 통해 지식을 표현하는 대표적인 방법에는 신경망 기술이 있다.

신경망이란 인간의 뇌구조를 모방내어 구성한 망으로, 학습이 가능하며 학습으로 얻어진 정보를 망 내의 신경세포(뉴런)들의 연결고리(시냅스)에 저장시킬 수 있다. 여기서 학습이란 주어진 목적에 맞게 각 신경세포 간의 연결 강도를 변화시켜 나가는 과정이다. 학습 과정에서 연결강도는 처음에 랜덤으로 시작해서 입력된 학습패턴에 대해 어떤 평가를 내리고 그 평가에 따라 연결 강도를 변화시켜 나간다. 학습을 한 신경망은 각 신경세포 간의 연결 강도, 연결 구성 등에 따라 출력을 내게 된다(진성일, 임길택 1994).

Mozer(1984)는 신경망 기술을 이용한 신경망 정보검색시스템을 구축하였는데 이 시스템은 133개의 디스크립터로 색인된 407개의 소규모 문헌들의 집합으로 구성되며 신경망 기술의 중요한 특징 중 하나인 학습 기능을 포함시키지 않았다.

학습기능을 이용한 좀 더 정교한 시스템에는 Belew(1989)에 의해 설계된 AIR라는 정

보검색시스템이 있다. AIR 시스템의 신경망은 키워드, 문헌, 저자의 세 단계 망으로 5,000개의 노드들을 갖는 1,600개의 문헌으로 구성된다. 이 시스템에서는 역빈도 가중치 체계를 이용하여 연결 가중치의 초기값을 계산하였는데 이 초기값은 이용자의 적합성 퍼드백에 의해 구성된 학습규칙에 따라 수정된다. 질문은 키워드 노드뿐만 아니라 저자, 문헌, 혹은 키워드 노드의 결합을 활성화함으로써 시스템에 입력된다. 활성화는 신경망을 통해 전파되고 시스템이 안정될 때 가장 높게 활성화된 노드들이 가장 적합한 정보를 나타낸다. 이 시스템은 관련 문헌, 저자 그리고 추론 키워드들을 출력물로 제시한다.

신경망 기술을 이용한 시스템은 인간 두뇌 상에서 병렬로 수행되는 과정을 직렬 기계상에서 시뮬레이트하는 컴퓨터 시스템이기 때문에 처리 속도가 느리다는 점이 매우 큰 문제가 된다. 그러나 처리 속도를 보강한다면 신경망 기술을 이용한 전문가시스템은 생성규칙, 프레임 등과 같은 복잡한 지식 표현 기법을 이용하지 않고 예제 데이터들과 학습 규칙에 의해 바람직한 결과값을 얻기 때문에 문헌정보학 분야는 물론 여러 분야에 유용하게 이용될 수 있을 것이다.

2.3 편목 전문가시스템

2.3.1 편목업무와 전문가시스템

도서관은 도서를 비롯한 다양한 정보자료를 수집, 정리, 보존하여 조사 및 연구에 기여하는 사회적인 시설이다. 방대한 정보자료를 수집,

정리, 보존하여 이용자들의 요구에 신속하고도 정확하게 접근할 수 있는 검색기법을 효과적으로 수행하여야 한다. 이러한 도구를 목록이라고 하며 서지요소를 작성하는 과정을 편목이라고 한다(한국도서관협회, 1994). 편목은 광의로는 분류를 비롯하여 목록 작성과 관련된 모든 절차를 말하며 협의의 편목은 단순히 서지요소의 기술만을 말한다.

목록은 사서가 이용자에게 도서관의 소장자료에 관한 정보를 전하는 매개체로 목록을 만들고 유지하는 과정은 아주 복잡하다. 현대목록법의 원칙과 관행이 확립된 19세기 이래로 목록법은 점진적으로 발전하여 왔다. 목록은 업무의 수행이 상당한 양의 전문적인 지식을 요구하기 때문에 전문가 시스템의 적용가능한 분야로 주목을 받아왔다(Davis, 1992).

또한 Davis(1986b)에 의하면 목록 업무는 첫째, 전문가를 필요로 하는 분야이며, 둘째, 많은 시간을 요하는 업무이며, 셋째, 인지적인 과정에 따라서 행해지는 업무이며, 넷째, 초보자에게 가르쳐줄 때 반복적인 교육이 필요하며, 다섯째, 많은 예산이 소요되는 분야이며, 여섯째, 상식을 요하지 않는 업무의 특성을 가지기 때문에 전문가시스템의 적용이 적합한 분야라고 지적하였다.

2.3.2 선행연구

편목 과정을 자동화하는 가능성에 대한 첫 연구는 Sandberg-Fox(1972)에 의해 이루어졌다. 그는 편목 업무의 인간에 의한 지적 처리 과정이 객관적이고 비직관적인 컴퓨터 기술에 의해서 시뮬레이트될 수 있는지를 검증하였고

또한 최근에는 표제면에서 서지 데이터들을 패턴인식기법에 의해 자동 인지한 후 인지된 데이터에 기초하여 AACR2 규칙에 따라 기본 기입을 선택하는 과정을 개념적 모델로 제시하였다(Sandberg-Fox, 1990).

Cockshutt 등(1983)은 접근점 선택을 위해서 AACR2 규칙을 분석하여 그에 대응하는 결정 논리를 순서도 형태로 나타냈다. 또한 Shaw 등(1981)은 저자목록이 적절한 기입과 표목에 도달하는데 있어서 일련의 논리적 단계를 포함한다는 것을 보여줄 목적으로 AACR2의 내용을 순서도로 표현했다. 자동편목에 대해서 Wellish(1980)은 다음의 세가지 이유에서 부정적인 견해를 제시하였다. 첫째는 현재의 목록 규칙들이 기계에 적합하게 구성되어 있지 않고, 가끔 모호하거나 비논리적인 점이 있고 둘째는 표제면의 불규칙적인 레이아웃은 관련된 정보를 인지하는 것을 어렵게 만들고 셋째는 표제면에 있는 서지 정보들이 광학 문자 판독기(OCR)에 의해서 읽혀질 수 있는 활자 형태로 인쇄되어야 한다는 점이다.

자동 편목의 연구자들에 의해 목록규칙의 모호함과 불일치성이 자동편목의 실행에 장애가 되고 있다는 점이 지적되고는 있지만 목록 규칙을 순서도, 개념적 모델 등으로 표현하는데 성공했다는 사실은 편목 전문가시스템을 개발하는 것이 실행가능하다는 것을 보여준다.

지금까지 연구개발된 편목 전문가시스템은 사람의 지적인 업무처리 과정을 전문가시스템과 나누어 수행하는 사람-기계간의 인터페이스에 의한 컴퓨터-보조 편목 방법과 패턴인식 등의 기법으로 표제면, 판권지 등의 서지요소의 종류를 자동 인지하여 목록 작업을 수행하는

완전자동 편목 방법 등 두가지 유형이 있다. 특히 후자의 경우는 전자출판 시스템과 연결된 완전한 편목 기능을 갖고 있는 전문가 시스템을 구현하는 것으로 텍스트가 온라인으로 처리되기 때문에 모든 편목 작업은 중개자의 지적 처리 없이 전문가시스템 편목 과정을 통해서 수행된다. 다음은 이 두 유형에 속한 중요한 연구들에 대해 설명한다.

1) 컴퓨터-보조 편목

Epstein(1987)은 초보자와 숙련자를 위한 편목 보조 시스템인 MITINET/marc 시스템을 개발하였다. 이 시스템은 MARC의 규칙과 절차로 구성된 지식베이스를 이용하여 이용자가 MARC에 대한 특별한 지식이 없이도 시스템의 지시에 따라 필요한 목록 정보를 입력하면 해당 문헌의 MARC 레코드를 자동 생성할 수 있도록 한다.

Borko(1987)에 의해 MAPPER라는 편목 전문가시스템이 개발되었다. MAPPER는 AACR2 규칙을 기초로 하여 MARC 형식에 맞게 지도를 편목할 수 있도록 편목자를 보조해 주는 시스템으로 EXSYS라는 전문가 셸을 이용하여 구축하였다.

Gibb(Gibb & Shariff, 1988)에 의해 CATALYST라는 편목전문가시스템이 개발되었다. CATALYST는 편목의 초보자와 전문가를 대상으로 기입요소 생성시 사용해야 할 적합한 AACR2의 규칙을 제시하는 시스템으로 ESP/Advisor라는 전문가 셸을 이용하여 구축하였다.

Hjerppe(Hjerppe & Olander, 1989)에 의해 수행된 ESSCAPE 프로젝트에서는

AACR2 규칙들을 전문가시스템 셸 즉 EMYCIN과 Expert-Ease로 적용할 수 있는지를 조사해 보았다. 이 프로젝트는 시스템을 구축하는 것보다 목록규칙을 실험, 증명, 그리고 이해하는 것을 더 강조하였다.

국내의 경우를 살펴보면 김희섭(1988)은 KORMARC 기본기입필드 생성을 위한 전문가시스템을 개발하였다. 이 시스템의 지식베이스는 프레임과 생성규칙을 이용하여 구축하였고 시스템은 Turbo-Prolog를 이용하여 구현하였다. 김명희(1988)는 KCR2와 AACR2에 의거하여 관련규칙을 프레임과 생성규칙을 이용하여 지식베이스로 구축한 편목전문가시스템을 구축하였고 시스템은 C언어로 구현하였다.

2) 완전자동편목

Harrison(1985)은 광학 판독 기술을 이용하여 카드 목록의 서지 자료를 MARC 형태 레코드로 변환하였다. Endres-Niggemeyer(Weibel, 1992)에 의해 수행된 AUTOCAT 프로젝트에서는 자연과학 분야의 정기간행물 논문들을 광학문자인식 기술을 이용하여 판독하고 자동 편목하는 시스템을 개발하였다.

Jeng(1991a)은 표제면의 서지요소의 종류를 서지 요소의 위치, 문자유형과 크기, 공간, 구두점 등과 같은 레이아웃 특성을 이용하여 자동적으로 식별하였는데 표제면을 분석한 결과 가장 중요한 서지요소들은 본서명, 저자, 발행사이며 두 서지요소간을 구분해 주는 가장 중요한 속성은 문자크기의 변화와 공간으로 나타났다.

Molto(Molto & Sevenonius, 1991)는 표제면의 저자명을 자동 인지하기 위해서 개인명 인지 알고리즘과 단체명 인지 알고리즘을 개발

하여 이용하였다. 단체명을 인지하는 절차는 표제면의 데이터가 LC 저자명 전거화일과 Bowker 발행사 명감을 이용하여 만든 단체명 전거화일에 있으면 해당 서지 요소를 단체명으로 인지하였다. 한편 개인명을 인지하는 절차는 데이터를 먼저 이름, 이니셜, 개인 직책으로 구성된 이니셜 요소 전거화일과 대조하여 일치되면 그 다음 데이터를 하나 또는 두 개의 공란, 쉽표 등과 같은 성 구분자(postname marker) 요소들과 대조하여 일치하면 그 데이터를 개인명으로 판정하였다.

Vizine-Goetz(Vizine-Goetz et al., 1990), Weibel(Weibel et al., 1989)에 의해 수행된 OCLC 프로젝트는 TEX 조판 시스템을 이용하여 기계가독형 형태의 표제면으로부터 서지 요소의 종류를 자동적으로 인식하는 작업의 가능성에 대한 조사이다. 이 시스템의 지식베이스는 서지 요소의 물리적 및 구문적 특성들과 편목자의 경험적 지식들을 기초로 16개의 생성규칙으로 표현하였다. 이 시스템의 인식 성공률은 81%로 16개의 소규모 생성규칙을 고려해 본다면 상당히 높은 성공률로 볼 수 있다.

앞의 OCLC 프로젝트는 광학 문자 인식 기술을 이용하여 텍스트를 판독하여 문헌의 구조적 구성요소들을 목록하고 색인하는 전문가시스템의 개발로 발전되었다(Weibel, 1992). 이 시스템의 목적은 문헌의 저자, 초록, 색인, 목차, 그림 등과 같은 기술적 편목 구성요소들을 자동으로 판독하여 문헌과 문헌의 일부를 검색, 디스플레이, 전달하는데 이용될 수 있는 데이터들을 생성하는 것이다. 이 시스템은 기술적 목록 엔트리를 생산하도록 보조하거나 목록

카드를 MARC 레코드 형태로 자동 변환하거나 또한 全文 검색 시스템, 개인 데이터베이스 등을 위한 문헌의 색인 작업 등 다양하게 이용될 수 있다.

제 3 장 편목전문가시스템 설계

3.1 시스템의 개관

본 연구를 통해 설계한 시스템은 문헌 표제면/판권지의 서지요소의 종류를 자동인지하고 인지된 서지요소를 KORMARC 형태로 출력해 주는 편목 전문가시스템으로 이 시스템은 크게 지식베이스, 지식획득시스템, 서지요소 자동인지 알고리즘, 편목 알고리즘, 전거화일들로 구성된다.

이 시스템의 특징은 편목작성자의 개입 없이 표제면/판권지의 기계가독형 화일을 이용하여 서지요소를 KORMARC 형태로 출력해 주는 시스템으로 편목작성자와의 상호 작용이 필요치 않아 대화형 메뉴방식이 사용되지 않았다. 다음은 지식베이스 설계, 지식획득시스템, 지식베이스를 이용하여 서지 요소를 자동인지하고 KORMARC 형태로 출력해 주는 절차를 설계한 알고리즘 설계에 대해 설명한다.

3.2 지식베이스 설계

본 시스템에서 구축한 지식베이스에는 패턴

인식 지식베이스와 편목 지식베이스가 있다. 패턴인식 지식베이스는 표제면의 서지 요소의 종류를 판독하기 위하여 구축한 생성규칙이고 편목 지식베이스는 판독된 서지 요소들을 KORMARC 포맷으로 출력할 수 있도록 해 주는 생성규칙이다. 다음은 이 두 지식베이스와 서지요소의 자동인지와 편목에 사용된 전거화일들에 대해 설명한다.

3.2.1 패턴인식 지식베이스

패턴인식 지식베이스는 12개의 서지요소의 종류를 자동 인지하기 위해서 구축한 22개의 생성규칙으로 구성된다. 서지 요소의 각 항목별 확실도를 구하고 그 중 확실도가 가장 높은 항목을 선택케 하는 방식을 채택케 함으로써 전진추론방식을 이용하여 최종 결과값을 이끌어 내었다. 생성규칙 앞에 오는 소괄호속의 숫자는 생성규칙의 번호를 의미한다.

1) 본서명

- (1) 가장 처음에 온다 : 40
- (2) 글자크기가 가장 크다 : 150

2) 총서명⁵⁾

- (1) 가장 처음에 온다 : 30
- (3) 가장 마지막 문자가 아라비아숫자 또는 로마숫자이다 : 40
- (16) 문자열에 '총서', '전집', '전서', '선집', '시리즈', '서리이즈', '씨리즈' 라는

5) 총서명은 총서명과 총서번호로 구성된 문헌만을 분석하였으며 하위총서명, 본총서의 편(part) 번호 및 서명이 있는 문헌은 제외시켰다.

- 단어가 포함된다 : 150
- (17) 글자크기가 가장 크지 않다 : 30
- 3) 총서번호
- (15) 가장 크기가 큰 문자열 바로 앞에 위치한다 : 30
- (19) 바로 이전의 문자열에 '총서', '전집', '전서', '선집', '시리즈', '시리즈', '씨리즈' 라는 단어가 포함된다 : 50
- (20) 문자열이 아라비아숫자 또는 로마숫자로 구성된다 : 60
- 4) 판사항⁶⁾
- (1) 가장 처음에 온다 : 30
- (4) 마지막 글자가 '판'이다 : 50
- (5) 문자열이 판사항예외전거화일⁷⁾에 속해 있지 않다 : 20
- (6) '개정', '증보', '영인', '축쇄', '복사', '초訂' 이라는 단어가 문자열의 일부분이다 : 150
- 5) 관제⁸⁾
- (1) 가장 처음에 온다 : 70
- (7) 글자수가 10자 이상이다 : 20
- (15) 가장 크기가 큰 문자열 바로 앞에 위치한다 : 80
- 6) 관칭
- (1) 가장 처음에 온다 : 70
- (8) 글자수가 10자 미만이다 : 20
- (14) 판사항예외전거화일에 있는 문자열이다 : 20
- (15) 가장 크기가 큰 문자열 바로 앞에 위치한다 : 80
- 7) 대등서명
- (9) 외국어(영어, 일어)로 된 문자열이다 : 150
- (10) 번역 저작역할어⁹⁾가 없다 : 20
- 8) 원서명
- (9) 외국어(영어, 일어)로 된 문자열이다 : 150
- (11) 번역 저작역할어가 있다 : 20
- 9) 별서명
- (12) 문자열의 첫 두 글자가 '일명' 또는 '혹은' 등의 어귀로 시작한다 : 100
- 10) 부서명
- (13) 가장 글자크기가 큰 문자열(본서명)

6) 초판이나 제1판은 표제면 및 판권지에 나타나 있어도 이를 판차표시로 기술하지 않는다 (편목시 참조).

7) 판사항예외전거화일 = (대형판, 컬러에디션, 최신판, 컬러판, 영문판, 삼성판)

8) KORMARC 기술규칙에서는 본서명 앞에 기재되어 있는 文句중 총서명, 대등서명, 원서명, 저자명, 발행자명, 회차, 권차, 판차 등을 제외하고는 전부 관칭이나 관제로 취급한다고 규정하고 있다. KORMARC 기술규칙(예비노트보완판, 1985)과 단행본용 한국문헌자동화목록형식(1993)에서는 관칭은 본서명 앞에 기재되어 있는 문구가 10자 미만일 때 관칭으로 간주하고, 10자 이상 일 때는 관제로 규정하여, 관제와 관칭의 구분을 본서명 앞에 기재되어 있는 문구의 자수로 기준하고 있다. 한편 1991년 12월에 한국정보관리학회가 개정안으로 발표한 "한국문헌자동화목록법(KORMARC)에 관한 연구"에서 본서명 앞에 기재되어 있는 문구가 본서명에 종속적일 때는 관칭으로 규정하고, 본서명과 독립적일 때는 관제로 규정하고 있다. 그러나, 글자수에 의해 관제, 관칭을 구분하는 방법이 서지요소를 자동인식하는 데는 훨씬 용이한 방법이다.

9) 번역 저작역할어 = (옮긴이, 옮김, 역, 편역, 공역, 외역, 감역, 역저, 번역, 옮겨엮음, 역자, 옮겨엮은이, 편역자, 등역)

바로 다음에 온다 : 60

(17) 글자크기가 가장 크지 않다 : 30

(18) 저자명(개인명, 단체명) 바로 이전에 온다 : 70

11) 잡제명

(13) 가장 글자크기가 큰 문자열(본서명)

바로 다음에 온다 : 30

(14) 판사항예외전거화일에 있는 문자열이다 : 10

(17) 글자크기가 가장 크지 않다 : 20

(18) 저자명(개인명, 단체명) 바로 이전에 온다 : 40

12) 권차서명

(3) 가장 마지막 문자가 아라비아숫자 또는 로마숫자이다 : 40

(13) 가장 글자크기가 큰 문자열(본서명) 바로 다음에 온다 : 20

(17) 글자크기가 가장 크지 않다 : 20

(18) 저자명(개인명, 단체명) 바로 이전에 온다 : 50

(22) '제+숫자+권' 형식의 문자열이 포함된다 : 100

3.2.2 편목 지식베이스

앞의 패턴인식 지식 베이스에 의해 서지요소의 종류가 확인되면 그 다음은 KORMARC 기술규칙을 기반으로 한 편목 지식베이스를 이용하여 서지요소를 KORMARC포맷에 맞추어 출력한다. 다음은 KORMARC기술규칙과 편목 지식베이스에 대해 설명한다.

1) KORMARC기술규칙 :

세계서지통정과 자동화편목법을 목적으로 1971년에 발행된 ISBD(국제표준서지기술법)는 전 세계적으로 파급되기에 이르렀으며, 우리나라에서도 국립중앙도서관을 중심으로 ISBD의 적용을 다각적으로 토의한 결과, 1981년 1월을 기하여 국립중앙도서관의 모든 서지데이터는 ISBD에 준하여 입력하기로 결정하였다. 그러나 한국에서 출판되는 문헌을 기술함에는 많은 문제점이 제기되어, 이러한 한국적인 특성을 살리기 위해서 세부적인 처리방식을 규정할 필요가 생기게 되었다(국립중앙도서관, 1982). 이러한 필요성을 규정하고자 ISBD 세부규정안을 한국의 문헌적 특성에 입각하여 성안한 것이 이른바 1982년에 발행된 "국제표준서지기술법에 의한 한국문헌기술세부규칙, 제1집"이었다. 이어서 1983년에 제2집, 제3집의 내용을 보완하여 "한국문헌자동화목록법 기술규칙(단행본용) 예비노트판"을 발행하였으며, 1985년에 그동안의 "예비노트판"에 의하여 실제 데이터를 입력하면서 제기되는 문제점을 수정, 보완하여 "한국문헌자동화목록법 기술규칙(단행본용)-예비노트보완판"을 발행하여 기술규칙의 전거로 활용하고 있다(최영복, 1991).

〈표 1〉 표시기호 생성규칙의 예

-
- (1) IF 서명저자사항
THEN 표시기호 = 245
 - (2) IF 판사항
THEN 표시기호 = 250
 - (3) IF 총서부출표목 AND 통일총서명
THEN 표시기호 = 830
-

2) 편목 지식베이스 :

편목 지식베이스는 KORMARC기술규칙을 기초로 하여 구성한 것으로 다음 세가지 종류의 58개의 생성규칙으로 이루어진다.

〈표 2〉 지시기호 생성규칙 예

(1) IF 표시기호가 100 THEN 지시기호=1b
(2) IF 표시기호가 245 THEN 지시기호=10
(3) IF 표시기호가 260 THEN 지시기호=b

〈표 3〉 식별기호 생성규칙 예

(1) IF 표시기호가 100 AND 서지요소가 개인명 THEN 식별기호 =\$a
(2) IF 표시기호가 830 AND 서지요소가 총서명 THEN 식별기호 =\$a
(3) IF 표시기호가 830 AND 서지요소가 총서번호 THEN 식별기호 =\$v

3.2.3 전거화일

본 시스템에서는 총 8종의 전거화일을 구축하여 서지요소를 자동인지하고 인지된 서지요소를 KORMARC 형태로 출력하기 위해 이용하였다.¹⁰⁾ 다음 〈표 4〉는 전거화일명과 수록된 레코드 수를 기술한 것이다.

〈표 4〉 전거화일

화 일 명	레코드수
성씨전거화일	153
발행사전거화일	450
저작역할어전거화일	59
한국무저자명고전전거화일	31
단체명전거화일	50
발행지전거화일	9
판사항예외전거화일	6
역할어전거화일	70

3.3 지식획득시스템

본 시스템에서 구축한 지식베이스에는 패턴 인식 지식베이스와 편목 지식베이스가 있다. 패턴인식 지식베이스를 구축하기 위해서 먼저 155종의 실험 문헌집단의 표제면과 판권지의 서지 데이터를 분석하였고 이 실험 문헌집단에 포함되지 않는 도서관의 문헌들도 분석하였다. 또한 편목전문가들에게 의문이 있을 때 마다 논의했으며 참조한 문헌으로는 단행본용 한국 문헌자동화목록법기술규칙(1985년) 등이 있다. 편목 지식베이스는 단행본용 한국문헌자동화목록형식(1993년), 단행본용 한국문헌자동화목록법기술규칙(1985년) 등의 문헌들과 편목전문가들의 의견을 참조하여 구성하였다.

3.4 알고리즘 설계

다음은 지식베이스들과 전거화일들을 이용하여 표제면과 판권지의 서지요소를 인지해 주

10) 각 전거화일의 용도는 [3.4 알고리즘 설계]에서 자세히 설명되었다.

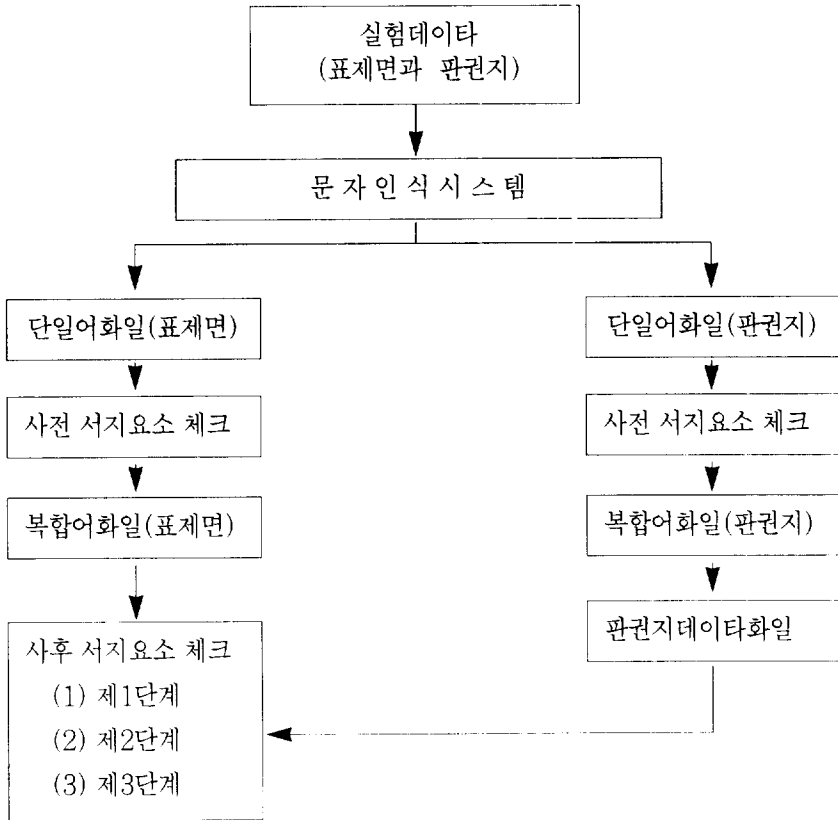
는 서지요소 자동인지 알고리즘과 인지된 서지 요소를 KORMARC 형태로 출력해 주는 편목 알고리즘을 설계하였다.

3.4.1 서지요소 자동인지 알고리즘

먼저 표제면의 서지 요소를 판독하고 부족한 정보가 있으면 판권지의 서지 요소를 판독하였다. 서지요소 자동인지 절차를 순서도로 표시하면 <그림 1>과 같다.

(1) 실험 데이터를 선택
155종의 실험 문헌을 선정하였다. 선정방식은 서울교육대학교 도서관 서고에서 무작위로 각 주제분야별로 1990년 이후의 단행본을 추출하였다. 선정된 실험 문헌의 표제면과 판권지를 실험 데이터로 이용하였다.

(2) 단일어화일(표제면) 생성
본 연구에서는 표제면과 판권지의 서지 데이터를 문자인식 시스템인 하이아트 글눈 2.1 프로그램을 이용하여 기계가독형화일인 표제



<그림 1> 서지요소 자동인지 절차 순서도

면화일로 변환하였다.¹¹⁾ 표제면 화일에서 공란과 구두점을 단어 구분자로 하여 단어를 추출하고 각 단어에 속성(문자유형, 문자크기, 위치¹²⁾을 수록하여 표제면 단일어화일을 구성하였다. 다음 <표 5>는 문헌번호 6과 14에 대한 표제면 단일어화일의 일부이다. 마침표(.)는 단어구분자로 사용하였으나 영어대문자 다음의 마침표는 개인명(영문) 이니셜의 한 부분으로 인지하여 붙여서 추출하였다.

(3) 事前 서지요소체크(표제면)

표제면의 단일어화일을 입력화일로 하여 저작역할어, 발행지, 발행년, 발행사의 서지요소를 체크한다. 서지요소의 종류가 확인된 결과가 <표 6>에 나와 있다.

(A)저작역할어, 발행사, 발행지는 전거화일을 이용하여 체크하였다. 저작역할어¹³⁾가 서지요소의 종류를 구별하는데 중요한 데이터가 되기 때문에 표제면에 '공 저'로 나타나 하나의

<표5>표제면 단일어화일1

문헌번호	단어순서	단 어	문자유형	위치크기	위 치
6	1	도서관경영론	고	38	24
6	2	R.Stueart	명	12	38
6	3	/	명	12	38
6	4	B.Moran	명	12	38
6	5	공저	명	12	38
6	6	임명순	명	12	38
6	7	/	명	12	38
6	8	오동근	명	12	38
6	9	공역	명	12	42
6	10	계명대학교출판부	필	12	158
14	1	학술연구총서	고	10	24
14	2	43	고	10	24
14	3	인공지능의	고	19	28
14	4	철학	고	19	28
14	5	이초식	고	11	42
14	6	고려대학교출판부	고	11	190
14	7	1993	고	11	193

11) 문자인식시스템은 문서를 키보드로 입력하지 않고, 스캐너로 입력받아 한글, 한자, 영어, 숫자, 특수기호, 도표까지도 자동으로 인식하여 아래아한글과 같은 기존의 워드프로세서에서 읽어 들어 원하는대로 편집할 수 있는 문자인식 소프트웨어이다. 우리나라에서 문자인식시스템의 개발에 관한 연구가 진행된 것은 80년대 후반으로 89년에 삼보컴퓨터에서 한국과학기술원과 공동으로 초보적인 수준의 학습 기능을 갖춘 한글문서인식시스템인 '실눈'을 개발하여 상품화에는 성공하지 못했으나 당시 열악한 연구 환경하에서 개발된 최초의 상업용 문자인식시스템이라는 점에서 의의가 있다. 현재 상용화되어 시판되고 있는 대표적인 문자인식시스템으로는 한국인식기술의 '하이아트(HiArt)글문'과 삼흥시스템의 '뉴로(OCR)', 코테크의 '리텍스', 한국 하이테크의 '르네상스', 헨디소프트의 '아리랑' 등이 있다.

12) 표제면 상단과 해당 서지요소의 거리를 mm로 표시한다.

13) 저작역할어리스트<부록 1> 참조

저작역할어가 공란때문에 두 개의 단어로 분리되어 단일어화일을 구성한 경우는 현재 단어가 '공' 일 경우 바로 다음에 오는 단어가 '저'인 경우는 '공저'로 문자열을 합병하여 하나의 저작역할어로 인지하도록 알고리즘을 구성하였다. 저작역할어는 세분하여 22-24, 35-37의 숫자코드들¹⁴⁾을 할당하였다.

(B)발행년(44)은 네개의 숫자로 구성되고 문자크기가 가장 크지 않으면서 단어순서가 1이 아닌 문자열을 발행년으로 인지하였다.

(C) 확인되지 않은 서지요소에는 '99' 코드를 할당하였다.

(4) 복합어화일 생성

사전 서지요소 체크에서 구분이 안될 경우는 서지요소 코드가 '99'이면서 글자유형과 글자크기가 같고 사이간격이 50mm이내에 있는 문자열을 결합하여 복합어화일을 만든다(<표 7> 참조).

(5) 事後 서지요소 체크

복합어화일을 입력 데이터로 하여 패턴인식 지식베이스와 전거화일, 판권지 데이터화일을 이용하여 事後 서지요소 체크 작업을 한다.

(A) 데이터

<표 6> 표제면 단일어화일2

문헌번호	단어순서	단 어	문자유형	위치크기	위 치	
6	1	도서관경영론	고	38	24	99
6	2	R.Stueart	명	12	38	99
6	3	/	명	12	38	99
6	4	B.Moran	명	12	38	99
6	5	공저	명	12	38	23
6	6	임명순	명	12	38	99
6	7	/	명	12	38	99
6	8	오동근	명	12	38	99
6	9	공역	명	12	42	36
6	10	계명대학교출판부	필	12	158	43
14	1	학술연구총서	고	10	24	99
14	2	43	고	10	24	99
14	3	인공지능의	고	19	28	99
14	4	철학	고	19	28	99
14	5	이초식	고	11	42	99
14	6	고려대학교출판부	고	11	190	43
14	7	1993	고	11	193	44

14) 저작역할어구분표(<부록 2> 참조)

15) 패턴인식 지식베이스에 대한 자세한 설명은 [3.2 지식베이스 설계]를 참조하시오.

(a) 패턴인식 지식베이스 : 본서명, 부서명 등 12개의 서지요소를 구분하는 22개의 규칙으로 구성되어 있다.¹⁵⁾

(b) 판권지 데이터화일 : 판권지 데이터화일은 문서인식 시스템에 의해 기계가독형태의 판권지화일을 다음과 같은 과정을 거쳐 생성한 화일이다.

- ① 판권지의 단일어화일을 입력 데이터로 하여 발행년은 문자열의 첫 네 글자가 숫자로 구성되어 있는 문자열을 먼저 뽑은 다음, 그 중에서 값이 가장 큰 것을 발행년으로 간주한다.
- ② 발행사 전거화일을 이용하여 발행사를 확인하였다.
- ③ 역할어 전거화일을 이용하여 역할어를 체크하였다. 역할어 전거화일은 59개의 저작 역할어에 발행사와 발행인을 나타내는 발행역할어 11개¹⁶⁾를 첨가하여 모

두 70개의 역할어들로 구성되어 있다. 역할어 인지 알고리즘은 역할어 전거화일에 있는 문자열을 인지하거나 판권지에 '역 자'로 나타나 하나의 역할어가 분리되어 구성된 경우는 표제면에서 처럼 현재의 단어가 '역'일 경우 바로 다음에 오는 단어가 '자'인 경우는 '역자'로 문자열을합병하여 역할어로 인지하도록 구성되었다.

- ④ 사전 체크를 한 후 글자유형과 글자크기가 같으면 문자열을 결합하여 판권지의 복합어화일을 구성한다.
- ⑤ 전거화일을 이용하여 발행사를 다시 한번 체크한다.
- ⑥ 저자명(개인명, 단체명)을 인지한다. 저자명 인지 알고리즘은 단체명 화일에 있는 복합어는 단체명으로 인지하고 개인 저자군은 저작역할어 바로 다음에 오는

〈표 7〉 표제면 복합어화일

문헌번호	단어순서	복 합 어	문자크기	서지요소코드
6	1	도서관경영론	38	99
6	2	R.Stueart*/*B.Moran	12	99
6	3	공저	12	23
6	4	임명순*/*오동근	12	99
6	5	공역	12	36
6	6	계명대학교출판부	12	43
14	1	학술연구총서*43	10	99
14	2	인공지능의*철학	19	99
14	3	이초식	11	99
14	4	고려대학교출판부	11	43
14	5	1993	9	44

16)만든곳, 만든이, 발행, 발행및인쇄처, 발행소, 발행인, 발행자, 발행처, 판권소유, 퍼낸곳, 퍼낸데.

복합어를 개인저자군으로 인지하였다.

- ⑦ 서지요소가 발행사(43), 발행년(44), 저자명(3(단체명), 77(개인저자군)), 저작역할어(22-24, 35-37)로 인지된 서지요소만을 추출하여 판권지 데이터화일을 생성하였다.

(B) 사후 서지요소 체크 과정

(a) 제1단계 : 표제면의 복합어화일을 분석하여 불용어를 제거하고 발행사, 발행지, 저자명을 인지하였다.

- ① 불용어리스트¹⁷⁾에 나온 단어는 제거한다.
- ② 발행사전거화일을 이용하여 발행사를 다시 한번 체크한다.
- ③ 발행사전거화일에 수록된 각 발행사의 발행지를 체크한다.
- ④ 저자명(개인명, 단체명)을 인지하였다. 저자명 인지 알고리즘은 단체명인 경우는 전거화일을 이용하였고, 개인명인 경

우는 저작역할어 바로 앞에 오는 문자열로서 문자크기가 가장 크지 않은 문자열을 개인저자(군)으로 인지하여 77의 코드를 할당한다. 이 절차에 의해서 앞의 <표 7>의 복합어 "R. Stueart*/*B. Moran", "임명순*/*오동근"의 서지요소 코드가 '99'에서 '77'로 바뀐다[<표 9> 참조].

(b) 제2단계 : 판권지 데이터화일을 이용하여 표제면에서 누락된 정보를 보완하였다. 판권지 데이터화일은 발행년, 발행사, 저자명, 저작역할어의 데이터를 수록하고 있다. 다음 <표 8>는 판권지 데이터화일의 일부이며 <표 9>는 표제면에서 누락된 데이터를 판권지 데이터화일에서 보완한 통합 복합어화일이다. <표 9>를 체크해 보면 문헌 6의 발행년이 판권지 데이터화일로부터 첨가되었고 문헌 14의 경우는 저작역할어의 생략으로 저자명을 인지하지 못한 것을 판권지 데이터화일을 통해 확인하였다.

<표 8> 판권지 데이터화일

문헌번호	단어순서	복 합 어	문자크기	서지요소코드
6	1	1993년	10	44
6	2	역자	10	35
6	3	임*명*순*&*오*동*근	10	77
6	4	계명대학교출판부	10	43
14	1	1993년	9	44
14	2	지은이	9	22
14	3	이*초*식	9	77
14	4	고려대학교출판부	9	43

17) 바로 뒤의 문자열이 발행사이며 다음의 불용어리스트에 있는 단어는 불용어로 처리하였다. 불용어리스트 = {도서출판, (주)도서출판, 사단법인, (주), 주식회사, 재단법인, 첨단 과학기술도서출판}. 이 연구에서 활용하지는 않았지만 개인명 앞에 오는 신분을 나타내는 단어들(예, 대학교수)이나 소속을 나타내는 단어들(예, xx대학교)도 불용어로 처리하는 것이 바람직하다.

물론 표제면 복합어화일에서 아예 생략된 저자명은 판권지 데이터화일을 통해 추출하였다.

(c) 제3단계 : 패턴인식 지식베이스를 이용하여 12 종류의 서지요소를 인지하였으며 다음의 <표 10>은 그 결과이다.

<표 10>에서 “도서관경영론”에 본서명코드 10이 할당되었는데 패턴인식 지식베이스를 활용하여 확신도가 가장 높은 본서명코드가 최종적으로 할당되었다. 다음 <표 11>은 “도서관경영론”에 할당된 서지요소코드들과 그것들의 확신도이다.

3.4.2 편목 알고리즘

앞의 표제면과 판권지의 자동인지 절차를 통해서 서지요소들의 종류를 확인한 후 편목 지식 베이스와 전거 화일들을 이용하여 KORMARC 포맷에 맞춰 다음 표시기호들

1XX(기본표목), 245(서명저자사항), 250(판사항), 260(발행사항), 4XX(총서사항), 500(일반주기), 507(원저자, 원서명에 관한주기), 7xx(부출표목), 830(총서부출표목)의 데이터를 출력한다. 편목은 바로 앞 단계의 출력물인 통합 복합어화일을 다음과 같은 절차에 의해 수정보완하여 KORMARC포맷으로 출력하였다.

(1) 제 1 단계

- ① 개인저자군 세분 : 개인저자군으로 인지된 서지요소중 저자가 2명 이상으로 구성된 문자열은 ‘·’, ‘/’, ‘&’, ‘,’, ‘ ’ 등의 특수부호를 문자열 구분자로 하여 분리한 다음 세분된 서지 코드를 할당해 준다. 이때 특수부호의 생략 등으로 저자간의 구분이 명확하지 않다면 성씨화일을 이용하여 개인명을 인지하였다. 즉 문자열의 글자수가 24자이면서 첫 글자(예,

<표 9> 통합 복합어화일 1

문헌번호	단어순서	복 합 어	문자크기	서지요소코드
6	1	도서관경영론	38	99
6	2	R.Stueart*/*B.Moran	12	77
6	3	공저	12	23
6	4	임명순**오동근	12	77
6	5	공역	12	36
6	6	계명대학교출판부	12	43
6	7	1993년	10	44
14	1	학술연구총서*43	10	99
14	2	인공지능의*철학	19	99
14	3	이초식	11	77
14	4	고려대학교출판부	11	43
14	5	1993	9	44

〈표 10〉 통합 복합어화일 2

문헌번호	단어순서	복 합 어	문자크기	서지요소코드
6	1	도서관경영론	38	10(본서명)
6	2	R.Stueart*/*B.Moran	12	77
6	3	공저	12	23
6	4	임명순*/*오동근	12	77
6	5	공역	12	36
6	6	계명대학교출판부	12	43
6	7	1993년	10	44
14	1	학술연구총서*43	10	48(총서명)
14	2	인공지능의*철학	19	10
14	3	이초식	11	77
14	4	고려대학교출판부	11	43
14	5	1993	9	44

김, 이) 또는 첫 두 글자가(예. 사공) 한 국성씨화일에 있다면 개인명으로 인지하였다.

② 세분된 서지 코드 할당 : 다음은 개인저

〈표 11〉 서지요소들의 확신도

도서관 경영론	
서지 요소	확신도
본서명	190
관칭	90
관제	70
부서명	70
판사항	50
권차서명	50
잡제	40
총서명	30
원서명	20

자군이외의 서지 요소들을 세분하여 세분된 서지 코드를 할당해 준다. 먼저 총서명인 경우는 문자열의 끝부분의 숫자를 떼어내어 총서명과 총서번호를 분리하여 수록해 주고 세분된 서지 코드를 할당해 준다. 본서명으로 인지된 문자열중에서 한국무저자명고전 전거화일에 있는 서명과 일치하면 이 문자열의 서지 코드를 10(본서명)에서 7(통일서명)로 변경하였다. 또한 총서명이 "연구논총", "일반자료", "연구보고서", "학술연구" 등으로 구성된 경우는 "총서명(발행기관명)" 형식으로 표시하여 서지코드 53(통일총서명)을 할당한다. 다음 〈표 12〉는 서지요소 코드를 세분하여 할당한 경우이다.¹⁸⁾

(2) 제2단계 : 〈표 12〉의 통합 복합어화일 3을 입력데이터로 하여 편목지식베이스에 의

18) 세분된 서지 코드표 (〈부록 3〉 참조)

〈표 12〉 통합 복합어화일 3

문헌번호	단어순서	복 합 어	서지요소코드
6	1	도서관경영론	10 (본서명)
6	2	R.Stueart	2(기본기입-개인명)(영문)
6	3	B.Moran	14(두번째 원저자명)(영문)
6	4	공저	23(복수원저저작역할어)
6	5	임명순	25(첫번째 번역자명)
6	6	오동근	26(두번째 번역자명)
6	7	공역	36(복수번역저작역할어)
6	8	계명대학교출판부	43(발행사)
6	9	1993년	44(발행년)
14	1	학술연구총서	48(총서명)
14	2	43	49(총서번호)
14	3	인공지능의*철학	10(본서명)
14	4	이초식	1(기본기입-개인명)(한글)
14	5	고려대학교출판부	43(발행사)
14	6	1993	44(발행년)
14	7	학술연구총서 (고려대학교출판부)	53(통일총서명)

〈표 13〉 KORMARC 출력물

문헌번호	#6	
100	1b	\$aStueart, R.%
245	10	\$a도서관경영론/\$d R.Stueart: \$eB.Moran공저: \$e임명순:\$e오동근공역%
260	bb	\$a대구:\$b계명대학교출판부, \$c1993%
700	1b	\$a Moran, B.\$e 공저%
700	1b	\$a임명순,\$e 공역%
700	1b	\$a오동근,\$e 공역%
문헌번호	#14	
100	1b	\$a이초식%
245	10	\$a인공지능의*철학/\$d 이초식저%
260	bb	\$a서울:\$b고려대학교출판부, \$c1993%
490	10	\$a학술연구총서 :\$v43%
830	00	\$a학술연구총서(고려대학교출판부):\$v43%

〈표 14〉 실제 목록완성에서 작성한 KORMARC 출력물

문헌번호	#6	
100	1b	\$aStueart,R.%
245	10	\$a도서관경영론/\$dR.Stueart:\$eB.Moran공저; \$e임명순:\$e오동근공역%
260	bb	\$a대구:\$b계명대학교출판부,\$c1993%
300	bb	\$a401p.:\$c23cm%
507	10	\$tLibrarymanagement%
653	bb	\$a도서관경영%
700	1b	\$a임명순,\$e공역%
700	1b	\$a오동근,\$e공역%
문헌번호	#14	
100	1b	\$a이초식%
245	10	\$a인공지능의*철학/\$d이초식저%
260	bb	\$a서울:\$b고려대학교출판부,\$c1993%
300	bb	\$a305p.:\$c23cm%
490	10	\$a학술연구총서:\$v43%
653	bb	\$a인공지능%
830	b0	\$a학술연구총서(고려대학교출판부):\$v43%

해 다음 〈표 13〉과 같은 출력물을 얻었으며, 〈표 14〉는 실제 편목업무의 목록완성에서 작성한 KORMARC 출력물이다.

헌집단은 무작위로 추출된 155개 문헌의 표제면과 판권지를 이용했으며, 또한 실험문헌집단을 이용하여 생성한 패턴인식 지식베이스의 질적 타당성을 검증하기 위한 검증문헌집단은 86개 문헌의 표제면과 판권지를 대상으로 하였다. 다음은 실험문헌집단과 검증문헌집단의 데이터를 분석한 결과이다.

제 4 장 시스템 평가

4.1 실험 데이터

본 연구에서 사용된 실험데이터는 실험문헌집단과 검증문헌집단으로 구성되는데 실험문

4.1.1 실험문헌집단

먼저 실험대상 문헌으로서 번역서 50종, 국내서 105종으로 구성된 155개 문헌의 표제면

과 판권지에서 편목에 필요한 모든 서지요소를 분석해 보았다. 우선 표제면을 분석해 보니 본서명은 모든 문헌에 포함되어 있었고 저자명이 생략된 문헌은 국내서인 경우는 105개 문헌 중 15개 문헌에서 누락되었으며 번역서의 경우는 50개 문헌 중 원저자가 생략된 경우는 2개의 문헌이 있었다. 발행년은 단지 21개 문헌에 수록되었고 발행사는 12개 문헌을 제외한 모든 문헌에 실려 있었다. 표제면에서 누락된 저자명, 발행사, 발행년 등은 판권지를 통하여 얻을 수 있었다. 표제면과 판권지에서 모두 누락된 데이터는 원서명의 경우 40개 문헌에서 누락되었고 원저자명은 36개 문헌에서 누락되었으며 발행사도 1개 문헌에서 누락되었다.

4.1.2 검증문헌집단

다음은 번역서 27종, 국내서 59종으로 구성된 86개 검증문헌집단의 표제면과 판권지에 수록되어 있는 서지요소들을 분석해 보았다. 표제면에서 본서명은 모든 문헌에 포함되어 있었고 저자명이 생략된 문헌은 국내서인 경우는 59개 문헌 중 18개 문헌에서 누락되었고 번역서의 경우는 27개 문헌 중 원저자가 생략된 문헌은 없었고 번역자가 생략된 문헌은 2개의 문헌이 있었다. 표제면에서 발행사와 발행년이 7개 문헌에서 누락되었으나 대부분 표제면에서 누락된 정보는 판권지를 통하여 얻을 수 있었다. 표제면과 판권지에서 모두 누락된 데이터는 번역서의 경우 원서명은 22개 문헌에서 누락되었으며 원저자명은 19개 문헌에서 누락되었고 발행사가 누락된 문헌도 5개 문헌이 있었다.

4.2 지식베이스

패턴인식 지식베이스를 평가하기 위해서 먼저 지식베이스 생성을 위하여 분석된 실험문헌 집단의 서지요소들을 알고리즘에 의해서 인지하였고 다음은 검증문헌집단의 서지요소들을 인지하였다.

4.2.1 실험문헌집단

실험문헌집단의 경우 155개의 표제면에서 838개의 서지요소를 추출하였으며, 패턴인식 지식베이스와 전거화일들을 이용하여 분석한 결과 49개의 서지요소를 제외한 789(94%)개의 서지요소가 정확하게 인지되었다. 인지되지 못한 서지요소들을 분석해 보면, 판사항을 관칭으로 오인한 경우 9개의 서지요소, 저작역할어가 생략되어 개인저자군을 부서명으로 오인한 경우 8개, 문자의 유형과 크기가 달라서 한 서지요소로 구성된 문자열이 분리되어 잘못 인식된 경우 7개, 저자명 앞에 소속기관이나 학위, 발행사를 수식하는 어구 등을 부서명으로 오인한 경우 6개, 관칭을 부서명으로 오인한 경우와 권차서명을 저자군으로 오인한 경우가 3개의 서지요소에서 발견되었다. 그 외, 총서명을 대등서명으로, 판사항을 부서명으로, 원저자명을 원서명으로, 본서명을 관칭이나 관제, 대등서명으로 인식하는 등 다양한 경우가 발견되었다.

4.2.2 검증문헌집단

검증문헌 집단의 경우 86개 문헌의 표제지

에서 401개의 서지요소를 추출하여, 패턴인식 지식베이스와 전거화일들을 이용하여 분석한 결과 28개의 서지요소를 제외한 373(93%)개의 서지요소가 정확하게 인지되었다. 인지되지 못한 서지요소를 분석해 보면, 서로 다른 종류의 서지요소들이 50mm 간격이내에 있으면서 문자의 크기와 유형이 같은 경우 하나의 서지요소로 인지되는 경우가 13개 문헌, 저자군을 부서명으로 오인한 경우 6개, 본서명을 부서명으로 오인한 경우 4개, 잡제를 대등서명이나 관칭으로 오인한 경우 3개, 관칭을 관제로 오인한 경우 2개 등 다양한 경우가 있었다.

4.2.3 논의

실험문헌집단과 검증문헌집단의 서지요소를 분석한 결과, 저작역할자가 생략되어 원저자, 번역자 등의 구분이 명확하지 않으면 판권지의 서지데이터를 자동 판독하여 대부분 정확한 서지요소를 인지할 수 있었다. 그러나 부서명과 잡제간의 구분¹⁹⁾은 먼저 편목작성자가 의미분석을 한 다음에 서지요소 판별이 가능하기 때문에 패턴인식기법에 의해 인식하는데 한계가 있었다.

본 연구의 결과를 OCLC 프로젝트의 인식률과 비교해 보면, OCLC 프로젝트는 16개의 생성규칙을 사용하여 81%의 인식률을 보였으나 본 연구에서는 22개의 생성규칙을 이용하여 93%의 인식률을 보였다. 본 연구에서 인식

성공률이 OCLC 프로젝트보다 높은 것은 첫째 OCLC 프로젝트보다 생성규칙의 수가 많으며, 둘째 OCLC는 표제면만을 인식했으나 본 연구에서는 표제면에서 인식할 수 없는 데이터를 관련지를 통하여 인식하여 인식률을 높일 수 있었으며, 셋째 국내서는 서양서보다 표제면의 레이아웃이 단순하기 때문에 추측된다.

제 5 장 결 론

본 연구에서 설계한 패턴인식기법을 이용한 편목 전문가시스템은 초보 또는 전문 편목자가 국가 표준 서지포맷인 KORMARC포맷에 따라 목록작업을 할 수 있도록 조인해 주는 시스템이다. 이 시스템은 문헌의 표제면과 관련지를 자동으로 판독하여 편목자들이 표제면과 관련지를 읽고 서지요소들을 확인하는 과정을 생략하여 신속하게 편목작업을 할 수 있도록 도와 줄 뿐 아니라 KORMARC에 익숙하지 않은 편목자들이 쉽게 KORMARC포맷을 배울 수 있도록 도와준다. 이 시스템의 효과로는 다음과 같은 점을 들 수 있다.

- 1) 편목에 소요되는 경비와 시간을 절약할 수 있다.
- 2) 편목업무의 표준화작업을 수행할 수 있다.
- 3) 국가 표준포맷인 KORMARC포맷을 활용한다.

19) 한국문헌자동화목록법 기술규칙에 의하면, 부서명과 잡제는 본서명의 뒤 또는 그 아래에 위치하여 본서명과 관련된 어귀로서, 본서명과 대체할 수 있는 서명이거나 그 어귀가 독립하여 하나의 서명을 이룰 수 있는 어귀는 부서명으로 규정하고, 본서명과 대체할 수 없는 어귀로서 그 어귀가 단순한 출판의 경위 또는 동기 등을 나타내는 것일 때에는 잡제로 취급한다고 규정하고 있다. 따라서 부서명과 잡제의 한계는 해당되는 서명이 표목으로 부출 될 만한 가치가 있는지 여부로 결정한다.

- 4) 초보자들의 교육프로그램으로서의 역할을 한다.
- 5) 1)에 의해서 사서들이 정보서비스(참고 서비스 등)에 좀 더 많은 시간을 할애할 수 있다.

본 시스템이 실제 업무에서 좀 더 효율적으로 활용되기 위해서는 먼저 문자유형과 크기를 그대로 아스키 화일로 재생시켜 주는 문자인식률이 향상된 문자인식시스템이 개발되어야 한다. 현재 활용하는 문자인식시스템은 한가지 문자유형과 크기를 유지하여 스캐너 등을 통해 아스키화일로 전환해 주는 시스템으로 화일을 원본 그대로 재생시켜 주기 위해서 편집과정을 거치게 된다. 또한 서지요소 자동인지의 모든 절차가 자동화되기 위해서는 본 연구에서 수작업으로 수행한 문자열 속성의 추출과정을 수행해 주는 프로그램이 개발되어야 한다.

다음으로 국내문헌은 표제면에 대부분 발행년, 발행지를 생략하고 있는데 기본 서지 요소들을 표제면에 수록하는 기준 설정이 필요하다고 생각된다. 또한 판권지에 자료의 형태사항(표시기호 300), 번역본의 원서명/원저자명을 수록하는 기준 설정이 있다면 바람직할 것이다.

이 연구에서 활용한 표제면과 판권지외에 패턴인식기법에 의하여 서지요소를 인식할 수 있는 대상으로 도서출판사 목록작성에 필요한 사항을 일정한 형식에 의해 표현한 CIP (Cataloging In Publication)²⁰⁾이다. 즉 CIP

화일이나 문헌에 인쇄된 CIP 정보는 문자인식시스템을 통해 기계가독형 화일로 전환한 다음 패턴인식기법으로 서지요소의 종류를 자동 판독한 후 KORMARC포맷으로 출력하는 방안이다. 국내에서는 아직 CIP시스템을 채택하고 있지 않지만 국립중앙도서관의 통제하에 조만간 실현될 수 있을 것으로 사료된다.

끝으로 국내 문헌에 적용하고 있는 목록기술규칙에는 KCR3, KORMARC 기술규칙이 있으나 한국도서관협회와 국립중앙도서관이 각각 독자적으로 개정 및 제정함으로써 한 국가내에 이원화된 목록규칙이 존재하기 때문에 목록규칙의 표준화에 장애요인으로 대두되고 있다. 그러므로 이 두 목록기술규칙을 하나로 통합하여 서지요소의 자동인지에 적절한 표준화된 한국문헌자동화목록법 기술규칙으로 재구성된다면 완전자동편목시스템의 구축이 훨씬 용이해질 것이다.

〈참고문헌〉

〈국내문헌〉

- 강한배. 1992. 문자인식에 의한 장서관리시스템. 석사학위논문, 숭실대학교 정보과학대학원.
- 고순주, 김화수. 1993. 인공지능의 이론과 실제. 서울 : 집문당.

20) CIP(Cataloging in Publication)란 용어는 출판중목록정보로 표기하며 CIP프로그램에서는 출판사로 하여금 사전정보를 국가서지기관에 제공하도록 하고 있는데, 이 정보는 인쇄서지와 상호교환 테이프봉사, 온라인 데이터베이스에 수록되도록 출판예정도서의 레코드를 작성하기 위하여 사용된다. 이 CIP 데이터는 대개 도서의 표제지의 이면(Verso)에 인쇄되는데, 그 정보는 개인과 단체에 대한 전거가 되는 표목, 주제명표목, 분류번호에 중점을 두고 있다.

- 국립대학도서관협의회. 1992. 대학도서관업무
편람. 서울 : 일신출판사.
- 국립중앙도서관. 1985. 한국문헌자동화목록법
기술규칙 : 예비노트 보완판. 서울 :
동도서관.
- 국립중앙도서관. 1991. 한국문헌자동화목록법
(KORMARC)에 관한 연구 : 단행본
용 기술규칙. 서울 : 동도서관.
- 국립중앙도서관. 1993. 한국문헌자동화목록형
식 : 단행본용. 서울 : 동도서관.
- 권기정. 1991. 패턴분류를 위한 적응학습 알
고리즘 개발. 석사학위논문. 전북대학
교 대학원.
- 권아영. 1994. "데이터 입력의 새 장을 여는
문자인식시스템," Computermart 1
월호.
- 권재욱. 1994. "문서자동입력기술", 컴퓨터 4
월호.
- 그레들리 E., 홉킨슨 A. 1993. 서지정보의 상
호교류. 정필모, 현규섭, 오동근 공역.
서울 : 아시아문화사.
- 김기태, 이상용. 1991. 인공지능. 서울 : 대웅.
- 김남석. 1986. 도서편목법. 대구 : 계명대학교
출판부.
- 김명희. 1989. 편목업무를 위한 전문가시스템
에 관한 연구. 석사학위논문, 연세대학
교 대학원.
- 김영희. 1993. 목록에 있어서 기본기입론에
관한 연구. 석사학위논문, 전남대학교
대학원.
- 김정현. 1994. 문학류 분류를 위한 전문가시
스템의 개발에 관한 연구. 박사학위논
문, 중앙대학교 대학원.
- 김진형. 1994. "문자인식의 현황과 전망." 컴
퓨터 2월호.
- 김치우. 1990. "한국목록규칙 3판의 특성과 호
울성," 경성대학교 논문집 제11집 3
권.
- 김현희. 1989. "도서관에서의 전문가시스템
응용(1)," 도서관 제45권 5호.
_____. 1989. "도서관에서의 전문가시스템
응용(2)," 도서관 제45권 6호.
- 김호건. 1994. KCR3와 KORMARC의 비교
분석. 석사학위논문. 중앙대학교대학
원.
- 김희섭. 1988. KORMARC 기본기입필드 생
성을 위한 전문가 시스템. 석사학위논
문, 경북대학교 대학원.
- 대한출판문화협회. 1994. 1994 한국출판년
감. 서울 : 동협회.
- 박신흥. 1991. 출판학사전. 서울 : 경인문화
사.
- 산업기술정보원. 1992. "제조업에서 전문가시
스템의 개발동향," 조사연구보고 제76
호.
- 신봉기. 1994. "온라인 필기 인식 기술," 컴퓨
터 3월호.
- 신용수. 1991. 서지데이터베이스 탐색식 작성
을 위한 전문가시스템 설계. 석사학위
논문, 중앙대학교 대학원.
- 신은자. 1988. 참고업무의 전문가시스템에 관
한 실험적 연구. 석사학위논문. 연세대
학교대학원.
- 안계성. 1994. 온라인 데이터베이스 선정을
위한 전문가시스템 설계. 석사학위논
문, 중앙대 학교 대학원.

- 오동근. 1991. "MARC포맷과 표기법," 정보관리학회지 제8권 1호. . 1991. 편목구축과 MARC 포맷에 있어서 동양 자료의 서지적 기술에 관한 비교 분석. 박사학위논문. 중앙대학교 대학원.
- 오영환. 1991. 패턴인식론 : 문자, 음성, 화상. 서울 : 정익사.
- 워트맨 D.A.(Waterman,D.A). 1992. A Guide to Expert System. 최세일, 유옥성 역. 서울 : 명성출판사.
- 이름한자 2854자 대사전. 서울 : 삼성서적, 1991.
- 이성찬. 1994. 패턴인식의 원리 : I권. 서울 : 홍릉과학출판사.
- 이수상. 1992. 대학도서관목록의 소급변환에 관한 연구. 석사학위논문, 부산대학교 대학원.
- 이재철교수정년기념논문집 간행위원회 편. 1994. 문헌정보학논총. 서울 : 구미무역.
- 정영미. 1990. "문헌정보학 영역 지식기반시스템에서의 지식표현," 정보관리학회지 제7권 2호
- 정환묵. 1993. 인공지능. 서울 : 동일출판사.
- 조충호. 1993. 인공지능개론. 서울 : 홍릉과학출판사.
- 진성일, 임길택. "신경회로망을 이용한 문자인식(1)," 마이크로소프트웨어 7월호.
- 진성일, 임길택. "신경회로망을 이용한 문자인식(2)," 마이크로소프트웨어 8월호.
- 청량정필모박사화갑기념논문집 편집위원회 편. 1990. 청량정필모박사화갑기념논문집. 서울 : 중앙대학교 문헌정보학과 동문회.
- 최영복. 1991. "KORMARC의 현재와 미래," 한국정보관리학회지 제8권 1호.
- 최원태. 1994. 정보조사제공 전문가시스템의 구축에 관한 연구. 박사학위논문, 연세대학교.
- 한국과학기술원 주최. 1993. 전거시스템의 현황과 발전방향에 관한 세미나. 대덕 : 동기술원.
- 한국도서관협회 편. 1994. 도서관정보편람. 서울 : 동협회.
- 한국도서관협회 편. 1990. 한국목록규칙 : 3.1판. 서울 : 동협회.
- 한국정보관리학회. 1993. 정보관리강좌. 서울 : 동학회.
- 현규섭. 1992. "자동화목록법(MARC)의 저자 표시 독립 원칙," 도서관 제47권 4호.
- 현규섭. 1993. "자동화목록법(MARC)의 서명 속성 분해," 도서관 제48권 6호.

< 외국 문헌 >

- Belew, R. K. 1989. Adaptive information retrieval : using a connectionists representation to and learn about documents. In Belkin, N. J. and Van Rijsbergen, C. J. eds., Proceedings of the Twelfth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Cambridge, Massachusetts, USA June 25-28 1989, New York,

- Association for Computing Machinery.
- Borko, H. 1987. "Getting Started in Library Exprt System Research," *Information Processing & Management* Vol.23.
- Bramer, M.A. 1982. "A Survey and Critical Review of Expert System Research," In D. Michie ed., *Introductory Readings in Expert System* (Gordon & Breach), pp.3-29. (A. Clarke and B. Cronin, "Expert Systems and Library/Information work," *Journ of Librarianship*, Vol.15, No.4(Oct. 1984), p.280에서 재인용
- Clarke, A and Cronin, B. 1983. "Expert Systems and Library Information Work," *J. of Librarianship* Vol.15, No.4.
- Cockshutt, Margaret E., Cook, Donald and Schabas, Ann H. 1983. "Decision Logic for AngloAmerican Cataloguing Rules, Chapter 21 : Choice of Access Points," *Library Resources and Technical Services* Vol.27.
- Davis, R. 1984. "Outlines of the Emerging Paradigm in Cataloging," *Information Processing & Management* Vol.23, No.4.
- _____. 1986a. *Expert System and Cataloging : New Wine in old Bottles?* In : F. Gibb ed., *Expert Systems in Libraries*. London : Taylor Graham.
- _____. 1986b. *Cataloguing as a Domain for an Expert Systems*. In R. Davies ed. *Intelligent Information Systems : Progress and Prospects*. Chichester, West Susses : Ellis Horwood.
- _____. 1992. *Expert Systems and Cataloging*. In *The Application of Expert Systems in Libraries and Information Centres*. London : BowkerSaur.
- Davis, R. and James, B. 1984. "Towards an Expert System for Cataloging : Some Experiment Based on AACR2," *Program* Vol.18, No.4.
- Epstein, H. 1987. "An Expert System for Novice MARC Cataloguers," *Wilson Library Bulletin* Vol.62.
- Ercegovac, Z. 1986. *Artificial intelligenc and cataloging reasoning*. *Proceedings of the 7th National Online Meeting* New York, April 10-12.
- Ford, Nigel. 1991. *Expert System and Artificial Intelligence : an information manager's guide*. London : LA.
- Gibb, F. and Shariff, C. 1988. "CATALYST : an Expert Assistant for Cataloging," *Program* Vol. 22.

- Harrison, M. 1985. "Retrospective conversion of card catalogues into full MARK format using sophisticated computercontrolled visual imaging techniques." Program Vol.19, No.3.
- Hill, J. S. 1985. "Wanted : Good catalogers," American Libraries Vol., No.10.
- Hjerppe, R. & Olander, B. & Marklund, K. 1985. Project ESSCAPEExpert system for simple choice of access points for entries : Applications of artificial intelligence in cataloging. Paper presented at International Federation of Library Associations(IFLA) 51st General Conference. Chicago.
- Hjerppe, R. & Olander, B. 1989. "Cataloging and Expert Systems : AACR2 as a Knowledge Base." JASIS Vol.40.
- Jeng, L. H. 1986. "An Expert System for Determining Title Proper in Descriptive Cataloging : A Conceptual Model," Cataloging & Classification Quarterly Vol.7.
- _____. 1988. The language of a title page. In the 51th proceedings of the annual conference of the American Society for Information Science, October 25-28, at Atlanta, George.
- _____. 1991a. "Knowledge Representation of the Visual Image of a Title Page," JASIS Vol.42., No.2.
- _____. 1991b. "The Structure of a Knowledge Base for Cataloging Rules," IPM Vol.27, No.1.
- Molto, M. and Svenomius, E. 1991. "Automatic Recognition of Title Page Names", IPM Vol.27.
- Mozer, M. C. 1984. Inductive information retrieval using parallel distributed computation. Institute for Cognitive science, University of California, San Diego. ICS Report No.8460.
- Sandberg-Fox, A. 1972. The Amenability of a Cataloging Process to Simulation by Automatic Techniques. Ph.D. Dissertation, Urbana, IL : University of Illinois.
- _____. 1990. Selection of Main Entry : A Conceptual Model, In Riggs, D. E. ed. Expert Systems in Libraries. Norwood, New Jersey : Ablex Publishing Corp.
- Schwarz, H. 1986. "Expert Systems and the Future of Cataloging : A Possible Approach," Library Bulletin Vol.26.
- Shaw, Malcolm et al. 1981. Using AACR2 : A Diagrammatic Approach. Phoenix, AZ : Oryx Press.
- Shell, P.S. 1985. Expert SystemA

- Practical Introduction. New York :
Wiley.
- Vizine—Goetz, D. & Weibel, S. &
Oskins, W. 1990. Automating
Descriptive Cataloging. In Rao, A.
and Riggs, D. E. ed. Expert
Systems in Libraries. Norwood,
New Jersey : Ablex Publishing
Corp.
- Waterman, D.A. A Guide to Expert
System. Massachusette :
AddisonWesley.
- Weibel, S. & Handley, J. & Huff, C.
1988. "Automated Document
Architecture Processing and
Tagging." Proceedings of the
Conference on Application of
Scanning Methodologies in
Libraries. Beltsville, Maryland,
National Agricultural Library,
Nov.17—18.
- Weibel, S. & Oskins, M. & Vizine—
Goetz, D. 1989. "Automated Title
Page Cataloging : A Feasibility
Study," IPM Vol.25.
- Weibel, S. 1992. Automated Cataloging :
Implications for Libraries and
patrons, In F.W. Lancaster and
Linda C. Smith ed. Artificial
Intelligence and Expert Systems :
Will they change the Library?
Urbana, IL : University of Illinois.
- Wellisch, H. H. 1980. The cybernetics of
bibliographic control : towards a
theory of document retrieval
systems. In Theory and
Application of Information
Research : Proceedings of the Second
International Research Forum on
Information Science(Copenhagen,
1977), ed. O. Harbo and L.
Kajberg, (London:Mansell, 1980),
pp.82—109.

〈부록 1〉 저작역할어

옮겨엮은이	편집
책임집필	역자
책임편집	공역
옮겨엮음	외역
공저자	감역
원저자	옮김
글쓴이	역해
번역자	감수
펴낸이	그림
지은이	말씀
머리글	추천
편저자	사진
공편자	해설
엮은이	번역
옮긴이	각색
공편저	교열
펴냄	역주
저자	평선
지음	편자
집필	선
원저	저
공저	글
공글	작
등저	편
등편	역
편저	씀
공편	등
엮음	외
주편	
역저	
번역	

〈부록 2〉 저작역할어구분표

- 22(단일원저저작역할어) : 책임집필, 책임편집, 지은이, 머리말, 편저자, 엮은이, 저자, 지음, 집필, 원저, 편저, 주편, 편집, 역해, 감수, 그림, 사진, 말씀, 추천, 사진, 해설, 각색, 교열, 역주, 평선, 편자, 엮음, 선, 저, 글, 작, 편, 씀, 글쓴이, 펴낸 이, 원저자, 펴냄
- 23(복수원저저작역할어) : 공저자, 공편자, 공편저, 공저, 공글, 공편
- 24(다수원저저작역할어) : 등저, 등편, 등, 외
- 35(단일번역저작역할어) : 옮긴이, 옮김, 역, 편역, 감역, 역저, 번역, 옮겨엮음, 역자, 옮겨엮은이, 편역자
- 36(복수번역저작역할어) : 공역
- 37(다수번역저작역할어) : 등역, 외역

〈부록 3〉 세분된 서지 코드표

1-기본기입이 되는 개인명(한글)	(100 1X \$a)
2-기본기입이 되는 개인명(영문)	(100 1X \$a)
3-단체명(전체)(기본기입)	(110 XX \$a)
4-단체명(단체명 기본)(기본기입)	(110 XX \$a)
5-단체명(하위기관명)(기본기입)	(110 XX \$b)
6-회의명(기본기입)	(111 XX \$a)
7-통일서명(기본기입)	(130 XX \$a)
8-통일서명(기본기입이 아님)	(240 XX \$a)
9-관칭	(\$a)
10-본서명(한글)	(\$a)
11-부서명	(\$b)
12-잡제	(\$c)
13-두번째 원저개인명(한글)(저, 편, 역음 등)(\$e)	
14-두번째 원저개인명(영문)(저, 편, 역음 등)(\$e)	
15-세번째 원저개인명(한글)(저, 편, 역음 등)(\$e)	
16-세번째 원저개인명(영문)(저, 편, 역음 등)(\$e)	
17-두번째 원저단체명(저, 편, 역음 등)(\$e)	
18-세번째 원저단체명(저, 편, 역음 등)(\$e)	
19-두번째 원저회의명(저, 편, 역음 등)(\$e)	
20-세번째 원저회의명(저, 편, 역음 등)(\$e)	
21-4명 이상 원저(개인명, 회의명, 단체명)(\$d)	
22-단일원저저작역할어(맨마지막 원저자 다음에 붙임)	
23-복수원저저작역할어	
24-다수원저저작역할어	
25-첫번째 번역개인명(역, 옮김 등)(\$d)	
26-첫번째 번역개인명(역, 옮김 등)(\$e)	
27-세번째 번역개인명(역, 옮김 등)(\$e)	
28-첫번째 번역단체명(역, 옮김 등)(\$d)	
29-두번째 번역단체명(역, 옮김 등)(\$e)	
30-세번째 번역단체명(역, 옮김 등)(\$e)	
31-첫번째 번역회의명(역, 옮김 등)(\$d)	

- 32-두번째 번역회의명(역, 옮김 등)(\$e)
- 33-세번째 번역회의명(역, 옮김 등)(\$e)
- 34-4명 이상 번역(개인명, 회의명, 단체명)(\$d)
- 35-단일번역저자역할어(맨마지막 원저자 다음에 붙임)
- 36-복수번역저자역할어
- 37-다수번역저자역할어
- 38-권차(\$n)
- 39-권차서명(\$p)
- 40-대등서명(\$x)
- 41-판사항(250 XX \$a)
- 42-발행지(260 XX \$a)
- 43-발행사(260 XX \$b)
- 44-발행년(260 XX \$c)
- 45-쪽수(300 XX \$a)
- 46-삽도여부(\$b)
- 47-크기(\$c)
- 48-총서명 또는 하위총서명(440 00 \$a)
- 49-총서번호 또는 하위총서번호(\$v)
- 50-관제(500 20 \$b)
- 51-원서명(\$t)
- 52-별서명
- 53-통일총서명
- 99-알려지지 않는 데이터
- 999-불용어