

도서분류자동화를 위한 지식베이스의 설계에 관한 연구*

이 경 호**

< 目 次 >

- | | |
|-------------------|---------------------|
| I. 서 론 | III. 분류자동화를 위한 원리유도 |
| 1. 연구목적의 필요성 | 1. 자동화원리유도 개요 |
| 2. 연구방법 | 2. 농학분야 분류자동화 원리 |
| 3. 연구범위 및 한계 | 3. 의학분야 분류자동화 원리 |
| 4. 선행연구의 개요 | 4. 주제의 자동인지 방법 |
| 5. 용어의 정의 | IV. 분류자동화의 결과 및 분석 |
| II. 지식베이스 설계 및 구축 | 1. 분류자동화의 결과 |
| 1. 설계개요 | 2. 주제식별 관점에서 본 분석 |
| 2. 설계원리 | 3. 분류기호에 대한 분석 |
| 3. 설계(안) | 4. 분류자동화상의 고려사항 |
| 4. 용어수집 및 분석 | V. 결 론 |
| 5. 지식베이스의 구축 | |

I. 서 론

1. 연구목적 및 필요성

분류란 어떤 공통적인 성질이나 특성에 따라 유사한 것끼리 모으는 것을 의미한다. 도서관이나 정보센터 등에서 일련고 있는 분류는 그 대상이 문헌이기 때문에, 분류의 과정 또한 문헌의 외형적인 속성에 의해서라기 보다는 내용적

* 이 논문은 1990년도 문교부지원 한국학술진흥재단의 지방대학육성 학술연구조성비에 의하여 연구되었음.

** 대구대학교 문헌정보학과 부교수

인 속성에 따라 구분, 기호화하는 특성을 지니고 있다. 이렇게 문헌의 내용적인 속성을 파악하는 데는 무엇보다도 인간의 많은 지적인 노력을 요구할 뿐만 아니라 컴퓨터에 의한 처리를 어렵게 하는 요인이 되고 있다 하겠다.

그러나 오늘날 도서관이나 정보센터 및 연구소의 기술정보실 등에 있어서는 업무의 상당한 부분을 컴퓨터에 의존하고 있어 이용자에게 보다 신속하고 정확한 정보봉사를 수행하고 있으나, 분류업무만은 아직까지도 전통적인 수작업방식을 탈피하지 못하고 있는 실정이다. 특히 기존의 분류표는 대부분 열거식 분류표로서 분류표의 갱신이 어렵고, 주기적 갱신이 요망될 뿐만 아니라 새로운 주제의 삽입이 어렵다. 또 분류표 갱신에 따른 재분류작업이 요구되고 같은 분류표를 사용하고 있는 도서관이라 하더라도 동일한 분류가 행하여 지지 않는 등의 문제점을 지니고 있다. 이러한 점을 고려하여 볼 때 기존의 분류방법도 어떠한 방법에 의하든 같은 분류표를 사용하는 도서관이라면, 도서관마다 동일한 분류기호를 이끌어 낼 수 있고, 또 인간의 노력을 최소화할 수 있는 측면에서 보다 나은 연구가 수행되어야 하리라고 본다.

따라서 본 연구는 기존의 분류체계가 지니고 있는 단점을 란가나단(S. R. Ranganathan)이 고안한 콜론분류법의 원리인 분석합성식의 원리를 이용하여 분류자동화가 가능한 지식베이스를 구축하는데 있다. 그리고 이 지식베이스에 의하여 분류자동화가 가능하겠끔 분류자동화의 원리를 유도하여 내며, 나아가 문헌의 표제나 키워드의 입력을 통하여, 분류하고자 하는 문헌이 어떤 주제의 문헌인지 자동적으로 파악하도록 함과 동시에 서가상에 배가할 수 있는 분류기호를 자동적으로 만들어 내하고자 하는데 있다.

이렇게 컴퓨터에 의하여 분류가 자동적으로 이루어 지게 되면, 도서관이나 정보센터 등에 있어서 정보처리의 대부분이 완전자동화를 이룰 수 있게 되어 토탈시스템의 자동화에 크게 기여할 수 있게 된다. 뿐만 아니라 지금 현재 대부분의 도서관에서 독자적으로 분류하는 방식을 하나의 단일 프로그램 소프트웨어로서 업무처리가 가능하게 되어 업무중복을 피할 수 있고, 또 모든 도서관은 동일한 자료에 대해서는 동일한 분류번호를 가질 수 있어 정보유통면에 있

어서도 기여하는 바가 크다 하겠다. 그리고 이용자들의 입장에서 보면, 기존의 열거식 분류방식에서는 이용자 자신이 찾고자하는 주제가 어떻게 분류되어 있는지 검색하기가 거의 불가능한 반면, 이용자 스스로가 원하는 문헌의 분류기호를 생성, 확인할 수 있어 정보검색의 측면에서도 키워드검색에 버금가는 효과가 있다 하겠다.

2. 연구 방법

본 연구는 분류자동화가 가능한 지식베이스를 설계하는데 있으나, 이 지식베이스의 설계가 타당성이 있는지의 여부를 알아보기 위해 실제 실험분야인 의학과 농학 분야를 대상으로 용어를 수집 분석하여 지식베이스를 구축한다. 그리고 문헌의 표제(표제가 불명확할시는 키워드)를 입력하여 주제를 식별하여내고, 분류기호를 만들어 낼 수 있도록 상기 2개의 주제분야에 대해 분류자동화가 가능한 분류자동화 원리를 유도하여 내어 실제 프로그램을 작성한 후 실험을 행한다. 연구 방법은 다음과 같다.

- (1) 분류자동화가 가능한 지식베이스를 설계한다.
- (2) 연구대상 분야인 농학과 의학분야를 대상으로 용어를 수집, 분석한다.
- (3) 수집한 용어는 란가나단이 제시하고 있는 주제분야별 카테고리로 분석한다.
- (4) 분석한 용어를 지식베이스에 입력하여, 분류용 지식베이스를 구축한다.
- (5) 문헌의 표제(표제에 의해 주제인식이 어렵거나 분류가 불가능할시는 인위적으로 키워드 입력)에 의해 주제식별과 분류가 가능하도록 분류자동화 원리를 유도하여 플로우차트화 한다.
- (6) 플로우차트를 근거로 프로그램을 행하여 농학분야 283권의 문헌, 의학분야 196권의 문헌을 대상으로 실제 데이터를 입력하여 실험을 행한다.
- (7) 향후의 분류자동화를 위한 지식베이스 설계방법을 제안한다.

3. 연구범위 및 한계

연구 범위 및 한계는 다음과 같다.

- (1) 본 연구는 용어의 망라적인 수집에 의한 완전한 지식베이스를 구축하기보다는 어디까지나 분류자동화가 가능한 분류용 지식베이스 즉, 기계가독형 분류표를 설계하고자 하는데 있으며 실제 완전한 적용이 가능한 기계가독형 분류표를 만드는 것은 연구의 범위에서 제외한다.
- (2) 수집된 용어는 농학 564개, 의학 792개로서 총 1,356개나, 이 가운데 81개의 용어는 두 주제분야에 공통적으로 수집, 분석된 관계로 실제 용어 수는 총 1,275 이다.
- (3) 지식베이스에 의해 분류자동화를 가능케 하는 분류자동화 원리는 란가나 단의 CC의 원리를 적용하되 CC의 분류원리를 상세히 그대로 적용하지않고 다소 수정하여 사용한다.
- (4) 지식베이스 구축과 분류 자동화원리에 의한 자동화 가능성 여부의 실험은 농학과 의학문헌으로 국한한다.
- (5) 실험대상문헌은 영남대학교 의과대학 도서관 소장 의학관계 문헌을 무작위 추출한 196권과 대구대학교 중앙도서관 소장 농학관계문헌(1991년 4월 30일 현재 소장본) 283권을 대상으로 하였다.
- (6) 실험대상문헌은 지식베이스 자체의 용어가 영문으로된 단어만을 대상으로 하였기 때문에 영문자료로 한정하였다.
- (7) 실험대상분야인 농학과 의학분야를 대상으로 분류자동화 원리를 유도하고 이들 분야의 용어를 수집, 분석, 지식베이스를 구축한다.
- (8) 실제 분류는 내용분석에 의해 추출된 키워드에 의한 분류도 가능하나 이와 같은 분류방식은 분류자에 따라 동일한 분류기호를 생성할 수 없기 때문에 문헌의 표제에 의한 분류를 원칙으로 하고, 표제에 의한 주제인식이나 분류가 행하여 지지않는 경우에는 분류자가 주제분야를 지정한 후 표제에 의해 분류하게 한다. 그리고 이 방법에 의해서도 분류가 되지 않

는 경우에는 궁극적으로 주제와 키워드를 동시에 입력하여 분류하도록 한다.

- (9) 본 연구는 컴퓨터가 문헌의 표제를 근거로 주제분야별 주제인식과 더불어 분류기호를 자동적으로 생성시킬 수 있는지의 여부를 알아보고자 하는 하나의 실험적 연구임을 밝혀두고자 한다.

4. 선행연구의 개요

분류자동화에 관한 연구는 1960년대에 접어들면서 부터 시작되었으며, 크게 정보검색을 위한 문헌분류(document classification)의 자동화와 기존의 열거식이든 합성식이든 분류표를 통한 분류자동화 즉 도서분류(book classification)로 대별할 수 있다. 이에 대한 주요 연구들을 살펴보면 다음과 같다.

가. 분류표적 측면에서의 연구

분류표적 측면에서의 연구는 주로 분류기호를 부여하는 도서분류 즉, 코드분류의 형태로서 이에대한 연구를 연대순으로 살펴보면 다음과 같다.

1) 1950년대 이전

1937년 프랑스 파리에서 개최된 세계도큐멘테이션회의(world conference of documentation)에서 그로리어(Grolier)가 일반분류표(comprehensive classification)를 위한 일반적 원리 및 이론을 진단하여 보도록 강조한 사실¹⁾은 하나의 학문으로서 분류에 관한 연구의 시발이라할 수 있다. 또 1948년 영국학사원의 과학정보회의(Royal society's scientific information conference)²⁾에서 토의된 주제의 하나가 분류에 관한 것이었고 이 회의의 결과로 분류연구를 위한 위원회가 발족되고, 베르날(J. B. Bernal)교수가 이 위원회의 업무를 담당하여 연구하던

1) P. S. G. Kumer, "International Study Conference on Classification Research (3)," *Herald of Library Science*, V.14, No.1(Jan. 1975), p. 8.

2) A.C. Foskett. *The Subject Approach to Information*. 3rd ed. (London: Clive Bingley, c1977), p. 51.

시기이다.

2) 1950년대

1950년대에 접어들면서 분류에 대한 관심은 점차 고조 된다. FID에 대한 란가나단의 끈질긴 노력과, 당시 FID의 사무국장 뒤비스(Donker Duyvis)의 구상으로 1950년 FID/CA(Committee on General Theory of Classification)가 발족되고 란가나단이 의장이 된다.³⁾ 이로 인하여 분류도 하나의 과학(science)으로서 국제적인 차원에서 연구하기 시작한다.

한편 베르날 교수가 담당하여 오던 분류위원회의 업무는 1951년 비커리(B.C. Vickery)교수가 인계받아 1952년 CRG(Classification Research Group)⁴⁾라는 명칭으로 발족 시킨다.

이 CRG는 1952년 웰즈(A. J. Wells), 비커리, 코우즈(E. J. Coates), 페라던(J. Farradane), 포스켓(D.J. Foskett), 밀즈(J. Mills) 및 팔머(B. E. Palmer)⁵⁾등을 구성원으로 하여 최초의 회의를 개최하게 된다. 이들은 이 회의를 통하여 기존의 분류표에 대한 불만을 나타냄과 동시에 독특한 분류방법을 모색하려고 하였다. 이에 1953년 파세트분류에 대한 자신들의 견해를 밝힘과 동시에, 1955년에는 “정보검색의 모든 방법의 기초는 파세트분류”라고 하는 비망록을 간행하였다. 이 비망록에는 도서분류의 기초로서 파세트분류, 페라던의 카테고리간의 관계에 관한 이론, 기호법(notation)의 사용 세가지 기본관념을 강조하고 있다.

1957년 제1회 정보검색을 위한 분류연구회의(International Study Conference on Classification for Information Retrieval=ISCCR-1)가 CRG와 런던대학 도서관학교의 후원으로 Aslib에 의하여 도킹(Dorking)⁶⁾에서 개최되었는 바, 이 회의의 결과로 도서관인과 정보전문가가 한자리에 모여 분류에 관한 새로운 견해 및

3) P. S. G. Kumer, op. cit., p.9.

4) A. C. Foskett, op. cit., p.51.

5) Krishan Kumer. *Theory of Classification*(New Delhi: Vikas Publishing House, c1979), p. 466.

6) E. J. Coates, “Classification in Information Retrieval; the Twenty Years following Years following Dorking,” *Journal of Documentation*, V.34, N.4(Dec. 1978), p.289.

방법론을 토의할 전기를 마련함과 동시에 파세트 분류가 모든 정보검색의 기초가 된다고 선언 하였다.

3) 1960년대

1950년에 발족한 FID/CA는 1961년 FID/CR(=Committee on Classification Research)로 재편성 한뒤, 1964년 덴마크의 엘시노어(Elsinore)에서 제2회 국제분류연구회의를 개최하였고,⁷⁾ 이 회의의 주요의제가 분류의 일반이론과 자동화 분류였다. 특히 FID/CR은 62개국의 FID멤버로 구성되어 있었기 때문에 국제적 차원에서 가장 영향력 있는 집단이었으며, 개개 멤버들도 FID/CR의 업무에 적극적인 자세로 임하였다.

한편 1960년에 접어들면서 CRG는 특수분류표와 일반분류표 사이의 관계 및 파세트원리에 의한 새로운 일반분류표(new general classification)의 편찬에 관심의 초점을 두었으며, 1962년에는 새로운 일반분류표 개발을 위한 NATO⁸⁾ 기금이 영국도서관 협회에 제공되자, 협회측은 이 기금을 다시 CRG에게 위임하였다. 이에 CRG의 톰린슨(H. Tomlinson)이 1964~1968년 까지, 오스틴(Derek Austin)이 1968~1969년 까지 이 계획에 참여 하였는데, 이 때 CRG는 개념의 카테고리화를 위한 원리결정, 카테고리내에서의 개념간의 순서, 개념사이의 관계 등에 연구를 치중하였는 바, 이 기간중에 연구되어진 CRG의 도서관학분류표(Classification of Library Science)⁹⁾는 파세트원리에 의거한 대표적인 분류표라고 하겠다.

4) 1970년대 이후

1970년대에 접어들면서 CRG는 서가상의 배열 및 정보검색의 기능을 함께 지닐 수 있는 호환성있는 분류표(compatible scheme)의 고안을 목표로 하였었다. 그러나 이 두가지 기능의 양립은 불가능하다는 사실이 입증¹⁰⁾되자 자동검색

7) P. S. G. Kumer, op. cit., p.9.

8) Krishan Kumer, op. cit., p.467.

9) Classification Research Group, "bulletin no. 9," *Journal of Documentation*, V.24, N.4(Dec. 1968), pp.290~292.

시스템을 위한 분류표의 개발로 그 방향을 전환하였다. 이러한 상황 하에서 파 세트분류에 입각한 자동화시스템을 위한 최초의 소산물인 바로 PRECIS(PRESERVED Context-Indexing System)이다.

그리고 제3회 국제분류연구회의(ISCCR-3)가 1975년 1월 6일에서 11일까지 인도의 뭄바이에서 개최되었는데, 이 회의 주제가 전세계적 정보망을 위한 분류시스템¹¹⁾(Ordering System for Global Information Networks)이었다. 그러나 이것은 당초 UNISIST(World Science Information System)가 정보의 유통을 원활하게 하기 위해 표준참조코드(Standard Reference Code)라 불리는 일종의 코드화에 의한 분류인 표준망라분류표(Standard Roof Classification)가 개발되어야 한다는 결정에 따라 FID의 로이드(G. A. Lloyd) 및 코우즈 등이 이를 위탁받아 새로운 분류표의 개발을 시도한 것인데 이 SRC가 후에 BSO(Broad System of Ordering)으로 합병 되기에 이른다.

나. 정보검색적 측면에서의 연구

정보검색적 측면에서의 분류자동화에 관한 연구는 주로 용어분류(keyword classification)와 문헌분류(document classification)로 대별 된다.¹²⁾ 용어분류는 정보검색시 이용자의 질문과 문헌의 내용을 나타내는 용어가 서로 일치하는 범위를 넓히기 위하여 개개의 용어를 유사어군으로 분류하는 방법이고, 문헌분류는 서로 관련성있는 문헌을 군집화시켜 검색시 한 문헌군만을 탐색하도록 제한시켜 검색의 신속성을 기하기 위한 방법으로 이와같은 분류방법은 도서분류와는 거리가 멀다 하겠다.

5. 용어의 정의

본 연구에서 사용하고 있는 용어를 정의하면 다음과 같다.

12) G. Salton. *Dynamic Information and Library Processing* (Englewood Cliffs, NJ: Prentice-Hall, c1975), p.223.

(1) 도서분류(book classification) : 분류(classification)란 어떤 공통된 성질이나 특성에 따라 유사한 것끼리 모으는 것을 의미한다.¹³⁾ 도서분류(book classification)란 도서형태의 자료를 서가상에 배가하기 위해 일정한 기호를 부여하는 것을 의미한다.

따라서 도서분류는 정보검색시 이용자의 질문과 문헌의 내용을 나타내는 용어가 일치하는 범위를 넓히기 위하여 개개 용어를 유사어군으로 분류하는 용어분류(keyword classification)나 서로 관련성있는 문헌을 군집화시켜 관련문헌만을 탐색하도록 제한 하는 문헌분류(document classification)와는 구별할 필요가 있다.

(2) 지식베이스(knowledge base) : 지식베이스시스템이란 문제해결에 전문지식을 적용하는 컴퓨터시스템으로, 일반적으로 지식에 기반을 둔 컴퓨터시스템을 일컫는다.¹⁴⁾ 일상적인 컴퓨터 정보처리는 데이터와 프로그램에 의존하고 있으나, 지식베이스시스템은 데이터와 지식베이스 및 제어(control)의 세가지¹⁵⁾ 측면에서 지식을 처리하고 있다. 따라서 본 연구에서의 지식베이스란 특정한 문제영역의 문제를 (여기서는 분류문제) 해결하기 위해 전문영역의 각종정보를 수록하여 두고, 컴퓨터프로그램에 의해 처리 가능하도록 한 지식베이스시스템의 일부분을 의미한다.

Ⅲ. 지식베이스 설계 및 구축

1. 설계개요

정보시스템에서 사용하고 있는 분류표는 열거식 분류표가 가장 많이 사용되

13) Eric J. Hunter. *Classification Made Simple* (Aldershot: Gower, c1988), p.3.

14) 정영미, "문헌정보학 영역 지식기반시스템에서의 지식표현," *情報管理學會誌*, 第7卷 第2號 (1990), pp.35~36.

15) N. Shahla Yaghmai and Jacqueline A. Maxin, "Expert systems: A Tutorial," *JASIS*, V. 35, N.5(Sept. 1984), p.297.

고 있다. 이 중에서도 DDC가 가장 널리 사용되고 있다. 이의 주된 이유는 오랜 기간 동안 사용상에서 오는 탄력적인 점도 있겠으나 무엇보다도 장기적인 개정 작업과 목록중심의 사용, 그리고 이를 대치할 만한 우수한 일반분류표가 없었기 때문으로 풀이된다.¹⁶⁾ 그러나 열거식 분류표는 자동화 시스템에 적용할 수 없을 뿐만 아니라 여러성분으로 되어 있는 정보를 정보의 손실이 없이 일직선상에 기호로 표시할 수 없는 단점이 있다.¹⁷⁾

이러한 이유로 인하여 자동화 분류용 지식베이스의 설계는 기존의 수작업용의 분류표와는 다른 차원에서 설계하여야 한다. 이 설계과정에 포함시켜야 할 요소로서는 컴퓨터가 표제내지는 키워드를 인식하여 주제를 인식하여 내고, 이를 근거로 분류기호도 만들 수 있어야 한다는 점이다.

예컨대, 오늘날 정보시스템에서 가장 많이 사용하고 있는 용어통제표는 시소오러스(thesaurus)와 주제명표목표(subject headings)이다. 그러나 이들 통제표는 특정주제 분야내에서의 용어간의 관계를 묘사하고 있을 뿐, 학문분야간의 관계는 묘사하지 못하고 있다. 그러나 지식베이스상에서는 개개 용어마다 각 주제분야에서의 속성과 위치를 특정지워 주어야한다. 한 예로서, 우리는 boys, girls, teachers, reading 등의 용어가 나타나는 문헌의 주제는 교육학으로 추측할 수 있다. 문제는 인간의 사고과정이 어떻게 이것을 추론하여 낼 수 있느냐 하는 점이다. 이점을 토대로 지식베이스의 원리를 생각해 낼 수 있다. 이는 이러한 용어들의 속성이 모두 교육학분야에서 연구대상으로 하고 있는 용어라는 사실을 추론가능 하기 때문에 교육학관계 문헌으로 인식이 가능한 것이다.

따라서 분류용지식베이스 설계는 기존의 주제분야별 시소오러스를 전부 통합한 성격을 지니면서, 각 주제분야별로 개개용어마다 속성을 특징지워 줌으로써 주제식별이 가능한 형태를 설계하여야 할 것으로 생각된다.

16) Peter G. B. Enser. *Automatic Classification of Book Material Represented by Back-of-the-book Index*. Ph. D. Thesis. University of Sheffield. 1983. p.15.

17) *Ibid.*, pp.15~16.

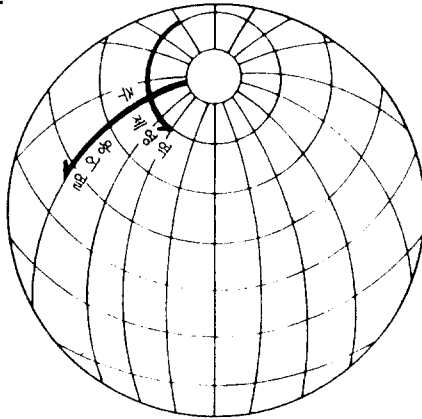
2. 설계원리

우리가 일상생활에서 사용하고 있는 모든 용어를 한곳에 모아, 체계적으로 편찬한 것이 사전이다. 이 사전으로 우리는 의미는 알 수 있으나, 학문분야간의 관계나 속성을 파악할 수는 없다. 즉 사전상의 단어의 의미는 표현상의 의미일 뿐이며, 주제구분의 의미는 없다. 그러나 하나의 단어는 주제분야에 따라 다른 속성, 다른 의미, 다른 위치를 가진다. 이러한 모든 요소들이 분류용 지식베이스상에 묘사될때만이 분류자동화가 가능할 것으로 보인다.

이러한 지식베이스를 설계함에 있어서는 다음과 같은 두가지 경우를 생각해 볼 수 있다.

(1) 지구의(globe) 원리

우리가 일상생활에서 사용하고 있는 모든 키워드가 하나의 주머니나 둥근 공의 핵심부위, 즉 지구의의 중심부위에 다음 <그림 1>과 같이 모두 들어있다고 생각해 볼 수 있다.



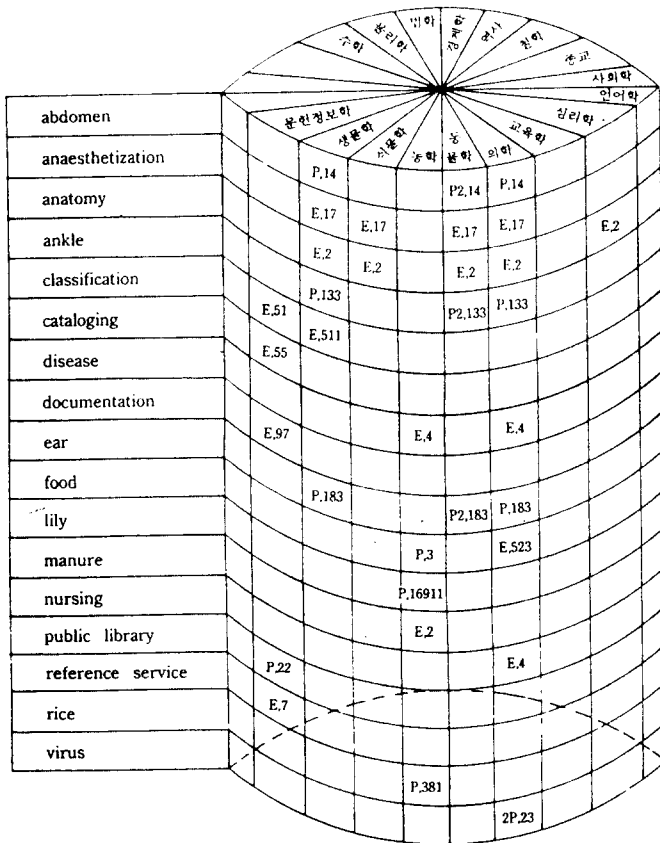
<그림 1> 지구의의 원리에 의한 설계원리

그리고 지구의의 위도는 용어열이고, 경도는 각 학문영역으로 볼 때, 하나의 용어는 어떤 학문분야에서든지 그 속성을 지닐 수 있게 된다. 실제로 용어는 핵심부위에 있기 때문에 언제든지 방향만 전환하면, 다른 주제분야의 용어로

그 성격을 변환할 수 있다. 그리고 개개 주제분야는 그를 주제 분야에서 연구 대상으로 하고 있는 용어만을 선택, 코드화 시킬 수 있어 각 주제분야 내에서의 특징을 묘사할 수 있게 된다.

(2) 원통형의 원리

상기 지구의에 의한 원리는 가장 이상적이거나, 좀 더 이해하기 쉽게 도식화하여 보면 다음 <그림 2>와 같이 원통형으로 설명할 수 있다.



<그림 2> 원통형의 원리에 의한 설계원리

이 원통형의 가장 중심부위 상하로 키이워드가 집결(이해를 돕기위해 알파벳 순으로)되어 있다고 가정하고, 원통의 중심점을 기점으로 여러 조각으로 구분하고, 각 조각 부분을 그림에서와 같이 하나의 학문영역으로 생각한다. 또 이해를 돕기 위해 중심부위에 상하로 배열되어있는 용어를 그림과 같이 좌측부위에 나열 시켜볼 수 있다. 이제 각 용어에 대하여 란가나단이 주장하고 있는 방법으로 분석하여 해당주제에 그 용어가 지니고 있는 속성이 다르게 나타날 수 있어 주제분야별로 특정의 고유값을 가지게 되며, 때에 따라서는 전 주제분야에 동일한 하나의 기호값을 부여할 수도 있다.

란가나단은 Colon Classification (6th ed)의 색인부분¹⁸⁾에서 다음과 같은 형식으로 용어간의 특징을 묘사하고 있다.

abdomen	G[P], K[P2], L[P], 14
anaesthetization	G, I, K, L[E], 17
anatomy	G, I, K, L, S[E], 2
ankle	G[P], K[P2], L[P], 133

상기에서 abdomen의 경우는 생물학(G)에 있어 [P], 동물학[K]에 있어 [P2], 의학(L)에 있어 [P]의 속성을 지니되, 그 값은 14라는 의미를 나타내며, anaesthetization의 경우는 생물학, 식물학, 동물학, 의학에서 다같이 [E]의 속성으로 17개의 값을 지닌다는 의미이다. 란가나단의 이와같은 색인방식은 하나의 용어에 대하여 각 주제분야별 속성과 자리매김을 분명히 하여 줌으로써 권말 색인이 통합시소오러스의 형태를 지니고 있는 예이다. 이들 용어에 대하여 기호화하여 보면 <그림 2>와 같다.

이 그림<그림 2>상에 표기된 용어를 설명하여 보면, anatomy는 생물학, 식물학, 동물학, 의학, 심리학에서 연구대상으로 하며, 용어의 속성은 [E]의 속성을 지니면서, 기호값이 2로서 동일한 값을 지니고 있다. 또 abdomen은 생물학과 동물학, 의학에서 연구대상으로 하며, 생물학에서는 [P]의 속성, 동물학에서는

18) S. R. Ranganathan. *Colon Classification*. 6th ed. (New York: Asia Publishing House, 1960), pp.2.126~2.129.

[P2]의 속성, 의학에서는 [P]의 속성을 지니면서 그 값은 다같이 14로 나타난다.

이상과 같은 분석 및 설계원리에 의거하여 우리가 일상생활에서 사용하고 있는 모든 용어를 분석하고, 데이터베이스를 구축한다면, 키워드나 표제를 컴퓨터에 입력하여 자동적으로 주제를 인식케하고, 이에 따라 분류기호도 자동적으로 만들어 낼 수 있을 것으로 본다. 따라서 이와 같은 원리에 입각하여 분류자동화가 가능한 지식베이스를 설계하고자 한다.

3. 설계(안)

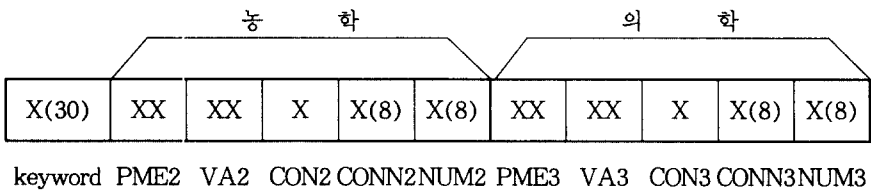
상기에서 언급한 원통형의 원리에 입각하되 농학과 의학만을 대상으로 컴퓨터 처리가 용이한 포맷으로 분류용 지식베이스를 설계하여 보면 다음 <표 1>과 같다.

여기에서 제1주제의 문헌정보학과 제4주제는 아직 미 사용 상태이며, 각 주제분야 내의 제어코드1과 제어코드2는 가능한 한 CC원리에 의거하여 분류가 될 수 있도록 분류원리를 제어하는 기능을 수행케 하는데 그 목적이 있다. 그리고 이들 필드별 요소를 설명하여 보면 다음과 같고 그림으로 나타내면 <그림 3>과 같다.

- 키워드 : 지식베이스상의 용어
- 키워드속성 : 용어의 속성을 기술한다.
- 배열값 : 한 주제내의 용어만을 출력하되 분류기호 순으로 배열 하고자 할 때 기호의 배열 값을 나타낸다.
- 제어코드 1, 2: 분류원리에 의해 분류되도록 제어하는 키
- 기호값 : 분류시에 조합되는 분류기호

〈표 1〉 지식베이스의 구성요소

필드항목	필드명	필드길이
· 키워드	Keyword	X(30)
· 미사용 여백		X(10)
· 제1주제 영역(문헌정보학) : 미사용		
키워드 속성	PME1	X(2)
배열값	VA1	X(2)
제어코드 1	CON1	X
제어코드 2	CONN1	X(8)
기호값	NUM1	X(8)
· 제2주제 영역(농학)		
키워드 속성	PME2	X(2)
배열값	VA2	X(2)
제어코드1	CON2	X
제어코드2	CONN2	X(8)
기호값	NUM2	X(8)
· 제3주제 영역(의학)		
키워드 속성	PME3	X(2)
배열값	VA3	X(2)
제어코드1	CON3	X
제어코드 2	CONN3	X(8)
기호값	NUM3	X(8)
· 제4주제영역 : 미사용		



〈그림 3〉 지식베이스의 구조

4. 용어수집 및 분석

가. 용어수집

용어수집의 목적은 분류를 자동화 하는데 있어서 컴퓨터로 하여금 문헌의 표

제나 키워드에 의한 탐색으로 자동적인 주제인식에 이어 분류원리에 입각하여 분류기호를 생성 시킬 수 있도록 분류용 파일인 지식베이스를 만들기 위한 것이다.

기존의 분류체계에 의한 분류는 모든 분류가 분류표를 근거로하여 이루어 지거나 자동화된 시스템의 경우에는 이 분류용 지식베이스를 근거로 하여 분류번호가 만들어진다. 따라서 용어의 수집은 기존의 분류방식에 있어서는 분류표가 필수적인 것과 마찬가지로 분류를 자동화하는데 있어서는 필수적인 과정이다.

지식베이스 구축을 위한 용어의 수집은 그 분야에서 사용 되어질 수 있는 대부분의 용어를 수집하되 가능한 한 동의어나 유사어 등을 포함한 광범위한 수집이 바람직 하다. 그러나 본 연구가 어디까지나 분류자동화를 위한 지식베이스의 설계에 있는 만큼 용어의 수집은 연구 대상분야인 농학과 의학분야의 용어를 망라적으로 수집 하였다기 보다는 아주 기본적인 정도의 용어만을 수집, 분석 하였다. 무엇보다도 해당주제분야의 용어수집은 주제배경이 없는 연구자 자신으로서는 용어의 망라적인 수집이 거의 불가능하기 때문에, 향후 실제 적용을 위해서는 보다 망라적인 차원에서 용어를 수집, 분석 하여야 하리라고 본다.

본 연구에서 수집한 용어는 농학분야 564개, 의학분야 792개 로서 총 1,356이다. 용어수집의 방법은 Colon Classification의 제6판과 7판¹⁹⁾에 근거하되, DDC 19판과 MESH 등을 주로 참고하였으며, 실제분류과정에서 표제상의 주요용어가 매치되지 않을 시는 이들 용어를 분석, 삽입하여 지식베이스를 갱신하여 분류하는 방법으로 하였다.

나. 용어의 분석

수집한 용어의 분석은 용어자체가 지니고 있는 속성, 즉 용어의 성질을 분석하고, 기호화함으로써 그 용어에 대한 속성치를 기호로 나타내어 지식베이스를

19) S.R. Ranganathan. *Colon Classification*. (Bangalore: Sarada Ranganathan Endowment for Library science, 1989), pp.216~219. 229~234.

구축하기 위함이다.

용어의 분석은 동의어나 유사어가 아닌 경우에는 가능한 한 고유번호를 부여하는 것이 좋다. 그 이유는 고유번호를 부여하면 그 만큼 고유의 분류번호가 만들어 질 수 있으므로 분류기호에 의한 배열뿐만 아니라 검색시에도 그만큼, 정도율을 향상시킬 수 있기 때문이다.

그러나 너무 고유번호에 치중하다보면, 용어의 수가 방대하여짐에 따라 분석된 용어에 대한 코드의 길이가 길어지게 되는 단점이 있고, 이것은 결국 분류번호의 길이가 길어지는 결과를 초래하게 된다. 따라서 분석된 용어의 코드화는 가능한한 개개용어마다 고유번호를 부여하도록 노력하되, 분류번호에 의한 배열 및 문헌검색적 측면을 고려하여 동의어 및 유사어 등을 다소 그룹핑하여 코드화 하였다. 또 코드의 길이를 줄일수 있는 방법으로서 십진수에 의한 코드화만 적용할것이 아니라, 비십진수의 알파벳트를 혼용하여 사용하는 것도 바람직하다 하겠다.

수집한 용어의 분석방법은 다음과 같다.

- (1) 두개의 명사와 결합하여 하나의 개념을 형성하는 경우에는 분리하지 않는다.
- (2) 동의어 및 유사어는 가능한 한 같은 기호를 부여한다.
- (3) 형용사+명사로 이루어져 하나의 개념을 형성하고 있는 경우에는 분석하지 않는다.
- (4) 형용사+복합명사의 경우에는 형용사와 복합명사를 분리하여 분석한다.
- (5) 복수의 형태는 단수로 취급한다. 단 반드시 복수로 사용되는 경우는 복수를 그대로 분석한다.
- (6) 지식베이스의 개념배열상의 이해를 돕기 위하여 CC상에 표기된 용어로 그대로 분석 코드화한 경우도 있으며, 이는 단지 표의 이해를 돕기위한 것이다.
- (7) 분석한 용어의 개념카테고리는 [P], [E], [S], [T]로 분석 하였으며, 의학분야의 경우는 특수질병(specific disease)을 나타내는 별도의 카테고리[D]를 생성시켜 분류하도록 하였다.

(8) 카테고리 결정후의 코드화는 CC의 6판과 7판 및 DDC 19판을, 주로 참고 하였으며, 기호로는 숫자 및 알페벳문자를 혼용하여 사용하였다. <표 2 참조>

(9) 하나의 개념의 최대 문자수는 30자로 제한 하였다.

5. 지식베이스의 구축

용어수집 및 분석결과에 따른 지식베이스의 구축은 DBASE III +를 이용하여 구축한 후, COBOL 프로그램에 의해 분류가 되도록 하였다. 구축된 농학과 의 학분예의 지식베이스는 <표 2>와 같다. <표 2>상의 지식베이스는 앞서 언급한 바 있는 원통형의 설계원리에 의거하여 컴퓨터가 입력된 용어를 인식하여 특정 주제의 인식과 더불어 분류기호를 합성하여 낼 수 있도록 설계되어 있다.

<표 2> 지식베이스의 구축 예

dermatopathology				p		87
dermotology				p		87
development	e		5			
development ontogeny				e		7
dew berry	p		1814			
diabetes				d	293	46
diagnosis				2e		3
diagnostic method				2e		34
diaphragm				p	25	1493
diarrhoea				d		452
diet regulation				2e		8
diet therapy				2e		81
digestive system				p		2
dilatation				e		3571
dioscorea	p		123			
diphtheria				d	41	4241
disease	e	4	4	e		4
diseases of foetus				e		324
disorder				p		70
dissection				2e		71

dividivi	p		872			
drainage	e		52	e		5718
draught				2p		951
drug	p	1	6			
drupaceous fruits	p		372			
dry farming	p		9d			
ductless gland				p		6
due to foetus				e		3542
due to passage				e		3544
dura mater				p		7211
dye	p	1	8			
dysentery				d	25	474
ear				p		183
east	s		9b	s		9B
eyelid				p		18511
fabric	p	1	7			
face				p		181
face and brow				e		3512
fallopian tube				p		552
farm	2p	2	2			
farm machine	2e	7	90			
farm machinery	2e	7	90			
farmer	2p	2	7			
fast cure				2e		6426
fasting				e		5736
fauces				p		215
favoured country	s		3	s		3
feed	p	1	2			
female genital organ				p		55
fertilizer	2p	2	4			
fertilizing	2p	2	4			
fever				2p		14
fibroma				2p		72463
fibrosarcoma				2p		72563
field crop	p		380			
fig	p		3748			
figs	p		3739			
filarial elephantiasis				d	39	4371
filberts	p		3894			
finger				p		168
fire				2p		91
first aid				2e		97
first molar				p		21436
fixation				e		195

flax	p		742		
floor hygiene				e	5711
floor of the mouth				p	2121
floriculture	p		16		
flower	p	0 6	16		
flower(drug)	p		66		
flower(food)	p		36		
flower(oil)	p		56		
fluid				2p	85
foetus				e	32
foliage	p		15		
food	p	1	3	e	573
food control				e	523

Ⅲ. 분류자동화를 위한 원리유도

1. 자동화원리유도 개요

도서관이나 정보센터 등에서 서가상의 배열이나 검색이 가능한 코드화를 위한 분류의 자동화란 컴퓨터의 도움으로, 체계적으로 편성된 분류용파일에서 한 문서의 내용, 주제 또는 형식에 일치하거나 유사한 분류번호를 탐색하여 그 문서에 자동적으로 분류기호를 배정하는 행위를 뜻한다.

그러나 분류자동화를 위한 가장 중요한 부분은 분류용파일을 만들기 전에 분류자동화가 가능한 원리, 즉 분류가 행하여지는 과정을 하나의 표인 플로우차트로 도식화 하여 내는 일이다. 이 플로우차트로 분류과정을 도식화하기 위해 서는 먼저 다음과 같은 전제조건이 충족되어야 한다.

- (1) 분류의 방법은 열거식이 아닌 조합식이 바람직하다.
- (2) 각 용어마다 개개 주제분야 내에서 성격과 위치가 분명하여야 한다.
- (3) 조합 하고자 하는 각 개념은 일정한 성격을 표현할 수 있는 기호가 있어야 한다.(CC의 P.M.E.S.T등)

- (4) 개념의 조합에는 일정한 원리가 있어야 한다.(주제 분야마다 동일할 필요는 없다)
- (5) 각 학문 분야마다 학문의 특징을 가능한 한 표현할 수 있는 고유의 분류 공식(formula)이 있어야 한다.
- (6) 여러개의 개념은 조합원리에 따라 일직선상에 표현할 수 있어야 한다.
- (7) 개념의 속성에 따라 주제의 인식과 분류가 함께 이루어질 수 있어야 한다.
- (8) 각 학문마다의 고유한 분류공식은 일정한 원리에 따라 조합가능 하여야 한다.
- (9) 각 주제분야의 분류공식은 하나의 일관성있는 도표로서 나타낼 수 있어야 한다.

이상과 같은 조건만 충족되어진다면 각 학문분야 마다 필요한 모든 용어를 수집하고 분석하여 분류용 과일을 만든다면 코드화 분류도 얼마든지 가능할 수 있다 하겠다.

그러나 지금까지의 많은 분류표 가운데에서도 오로지 CC의 경우에 있어서는 학문분야마다 분류의 공식이 다르기는 할지라도 사용하는 개념들의 성격이 비교적 명확하고 또 조합의 원리가 일정하게 명시하고 있어 향후의 분류 자동화의 가능성이 어떤 분류방식 보다도 높다고 하겠다.

2. 농학분야 분류자동화 원리

가. CC의 농학분야분류의 특징

농학분야의 분류표는 CC6판의 경우 p.2.64에서 p.2.66에 이르기까지 3페이지에 걸쳐 나온다. 분류공식은 J[P]:[E][2P]:[2E]로 나타나 있으나 [S]와 [T] 다같이 조합된다. 타분야에 비하여 좀 더 특이한 점은 [E]파셋트의 결합에 있어 많은 차이점을 발견할 수 있다. Second round energy facet가 존재한다는 점과 우선 표상에 나타난 점을 살펴보면 식물의 쓰임새(utility)에 따라 [P] 파세

트를 구분하되 이를 다시 Foci in Part Array of [P]를 사용하여 Foci in [P]하에서 다음과 같이 상세하게 세분하고 있다.

	Foci in Utility Array of [P]		Foci in Part Array of [P]
1	Decoration	1	Sap
2	Feed	2	Buld
3	Food	3	Root
4	Stimulant	4	Stem
5	Oil	5	Leaf
6	Drug	6	Flower
7	Fabric	7	Fruit
8	Dye, Tan	8	Seed
91	Adhesive	97	Whole plant
92	Manure		
93	Vegetable		
94	Sugar producing		

이와같이 농학은 쓰임새에 따라 [P]파세트를 대별하는 Utility Array 부분과, 이들 파세트의 여러부분을 이루는 Part Array가 먼저 나타나고, 이어서 [P]파셋트가 나온다.

[P]파셋트 후에는 [E]파셋트가 나타난다. [E]파셋트의 분류공식은 :[E] [2p]로 나타나고 있으나, 표상에는 [E]파셋트의 속성에 따라 조합방식을 달리하고 있는점이 특이하다. 특히 농학의 경우는 : [E] cum [2P] 이후 별도의 [2P]가 존재하고 있어 [E]파셋트 후에 나타날 수 있는 [2E]와 [2P]를 별도로 [E]파셋트 후에 열거하고 있다.

즉 표상의 Foic in [E]하에서 1 soil, 2 Manure, 3 Propagation, 4 Disease... 등은 이들의 성격에 따라 [2E]가 올수도 있고 [2P]가 올수도 있다는 점이다. 같은 [E]파셋트 이지만 1 soil의 경우는 다음에 [2E]가 결합되고, 2 Manure의 경우는 [2P]가 결합된다.

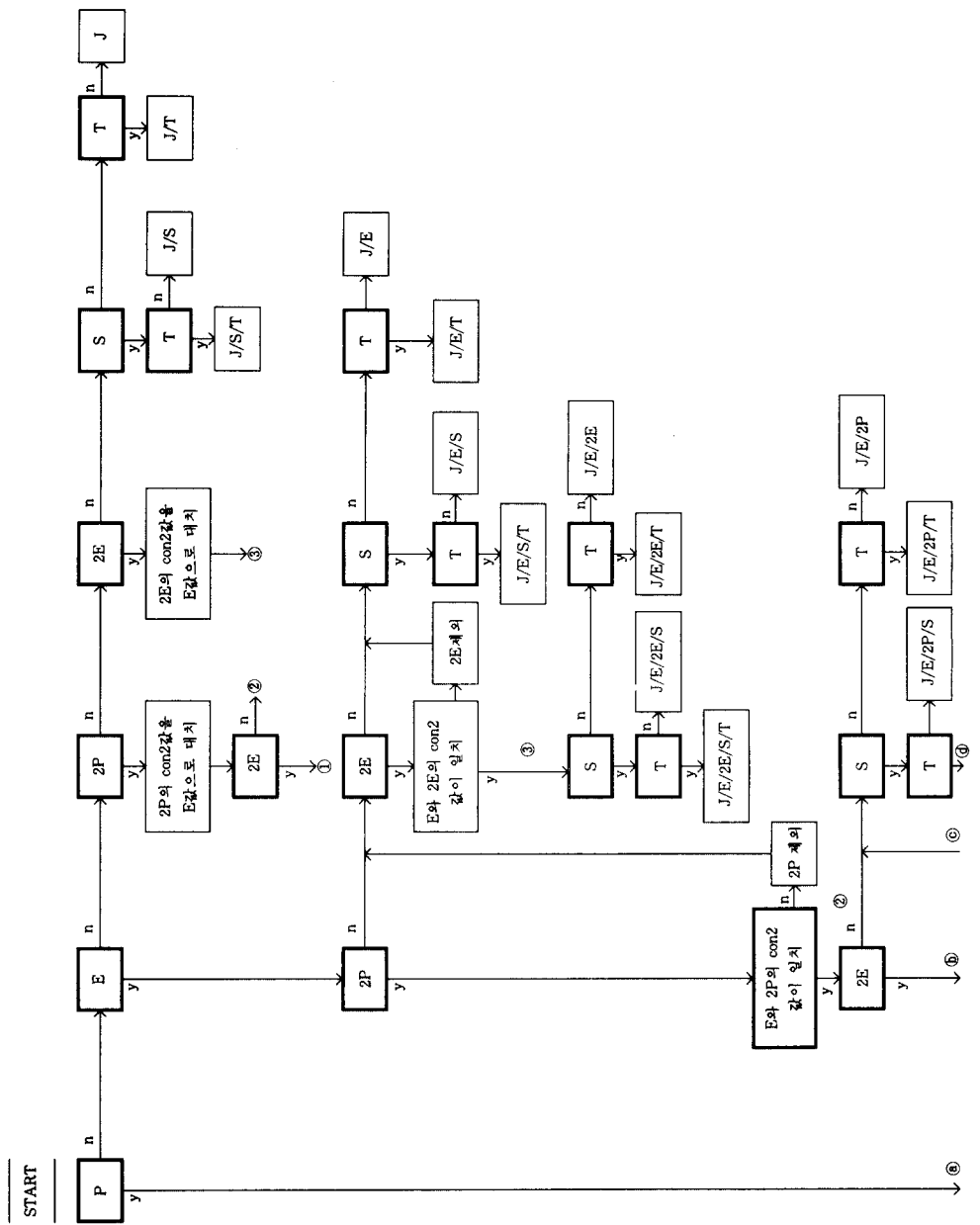
이 과정을 도표로서 나타내어 보면 다음과 같다.

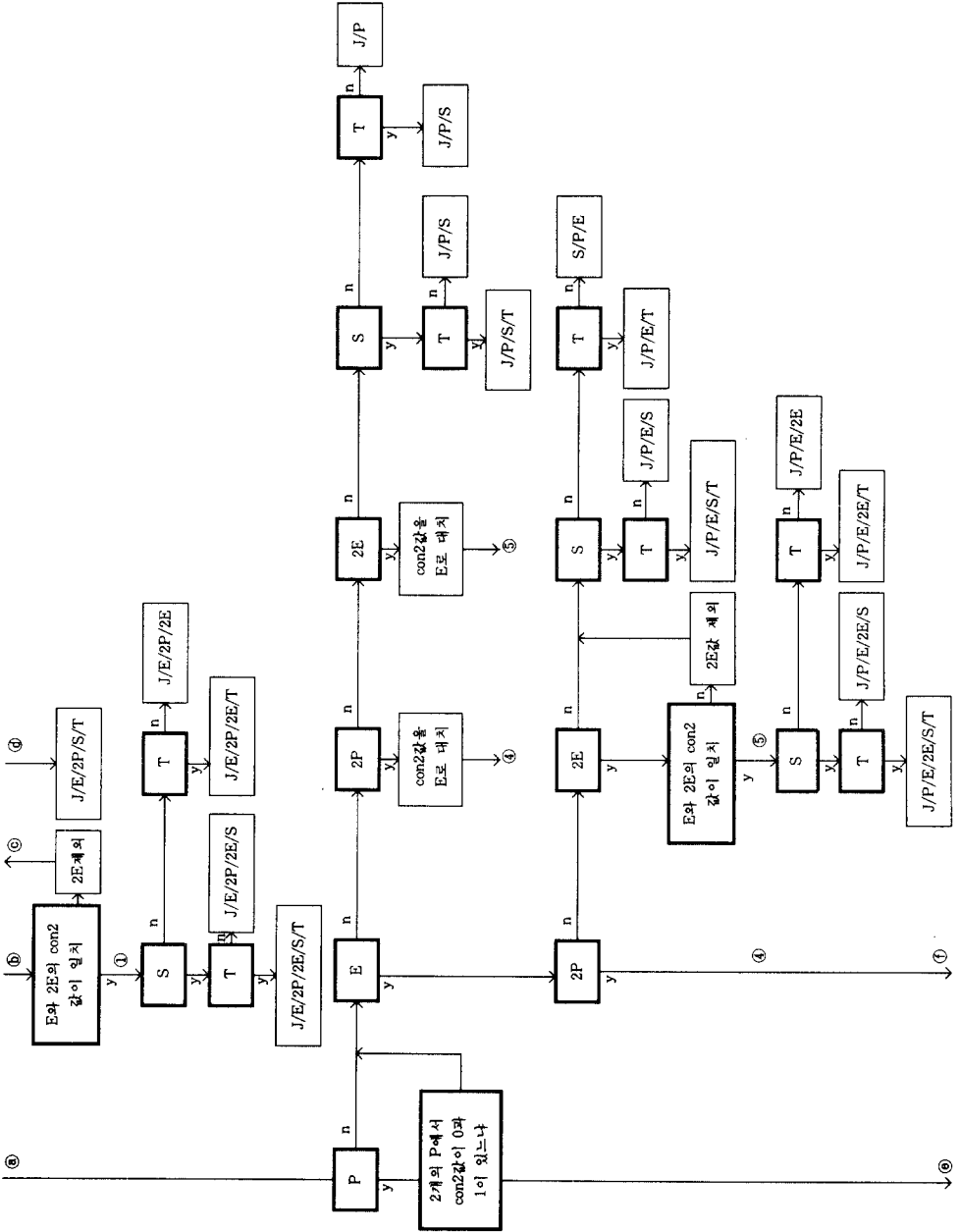
J[P] : [E]	조합될 공식
1 soil	J[P] : [E] : [2E]
	2 Tillage
[2E]	4 Damage
	6 Improvement
	7 Conservation
2 Manure	J[P] : [E] : [2P] : [2E]
	1 Green
[2P]	2 Farm
	3 municipal refuse
	4 Fertilizer
	5 Organic
[2E]	2 Collecting
	3 Application
	6 Compositing
	8 Storing
3 Propagation	J[P] : [E][2P] : [2E]
[2P]	Part Array of [P]를 사용
[2E]	3 Sowing
	5 Transplanting
	8 Stroing

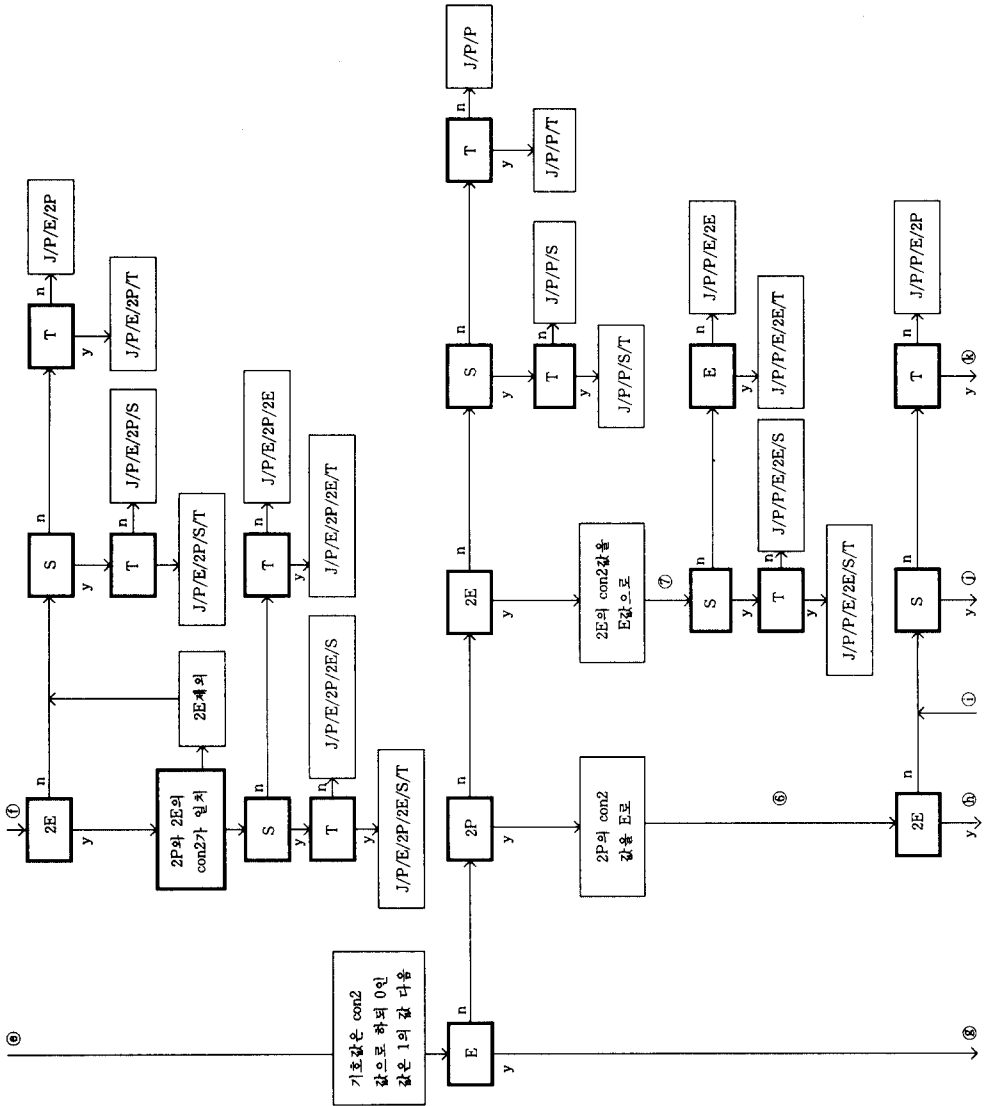
나. 농학(업)분야의 분류자동화 원리유도

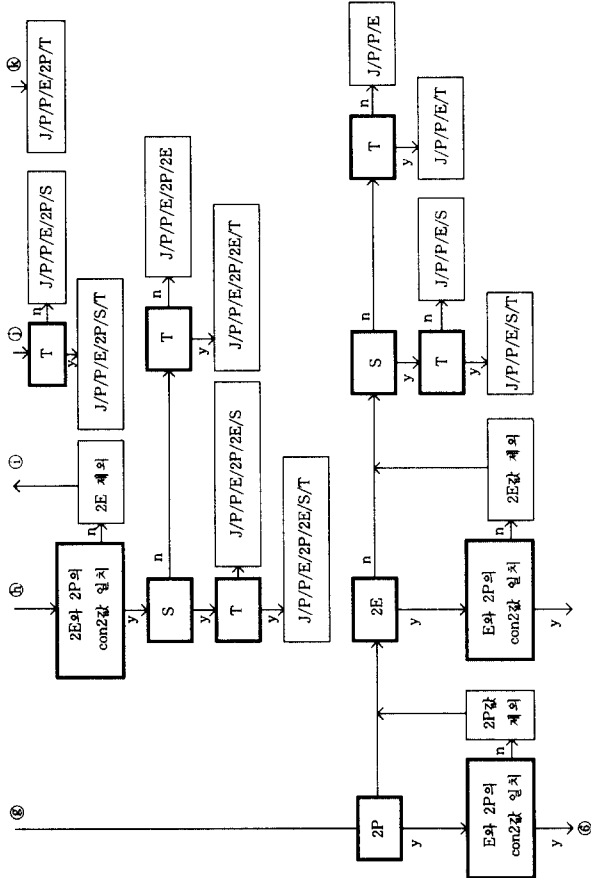
농학분야의 분류공식은 CC의 6판의 경우 J[P]:[E][2P] : [2E]로 표현하고 있다. 그러나 [E]파세트하에서 [2P]를 허락하는 경우도 있고, 허락하지 않는 경우도 있는 등 변칙적이다. 또 [E]4의 질병(disease)의 경우는 의학(L)의 질병 isolate들을 [2P]로 사용하는 등 비교적 복잡하다. 그러나 이러한 모든 사항들을 반영시킨 분류원리의 플로우차트는 다음 (그림 4)와 같이 도식화 할 수 있다.

(그림 4)에 나타난 분류시의 조합의 경우의 수는 모두 56가지로 나타난다. 이것도 물론 일반적인 조합의 경우의 수이다. 이 공식상에서 제외된 조합의 예로서는 [P]나 [E]의 개념이 2개이상 사용시는 조합을 하여 로테이션 시키는 경우이다. 그러나 이러한 요소들로 플로우차트를 상세하게 묘사하면 하나의 표









(그림 4) 능회분야 분류자동화 원리

상을 다 함께 묘사되어질 수 있는 성질이다. 이렇게 분류의 과정을 하나의 도표로서 묘사할 수 있다는 점은 바로 기계적인 수작업방식의 분류시스템에서 탈피될 수 있는 가능성이 있다고 하겠다.

이 <그림 4>상에 나타난 것을 중심으로 실제 문헌이 분류된다. 이 그림상에 나타난 조합의 경우의 수는 56개밖에 되지 않으나 실제 분류상에서는 각 카테고리마다 많은 isolate들이 있기 때문에 이의 조합의 경우의 수는 천문학적인 수가 된다.

3. 의학분야 분류자동화 원리

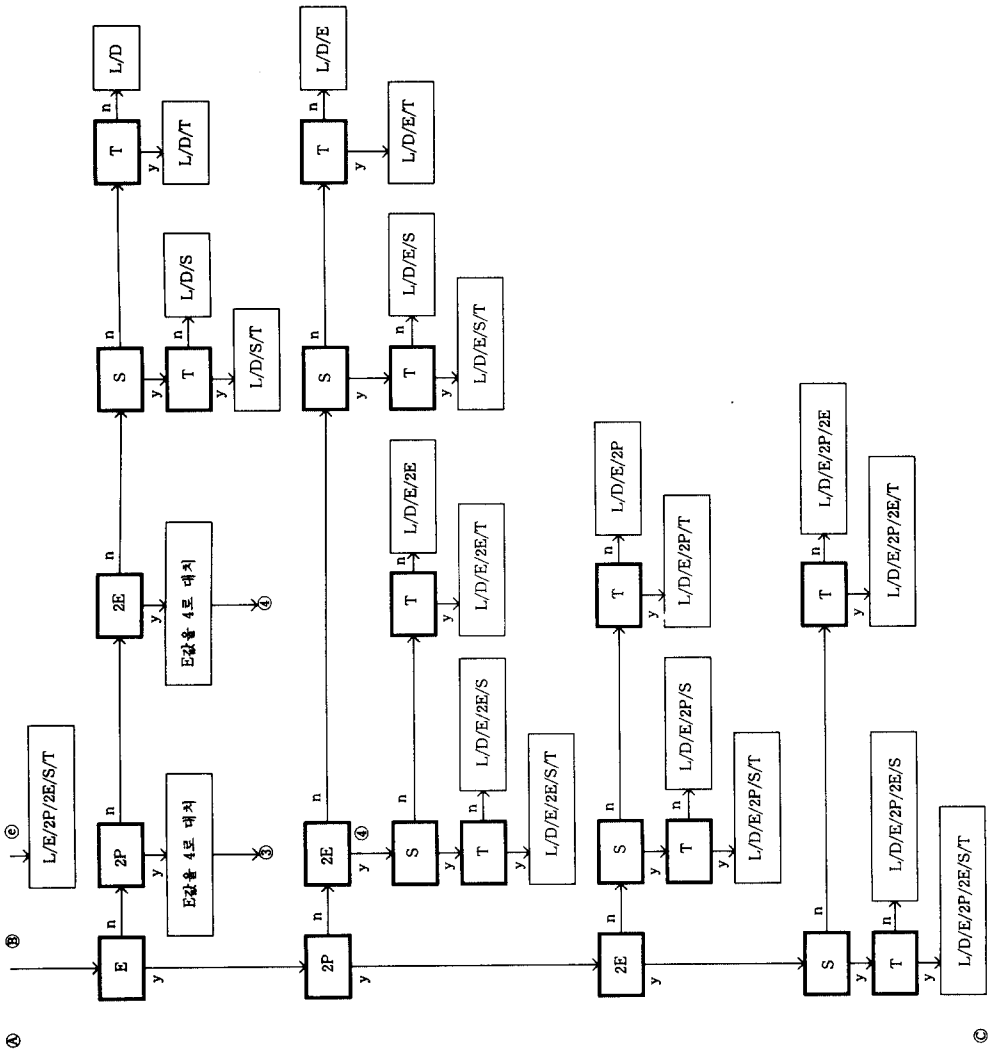
가. CC의 의학분야 분류의 특징

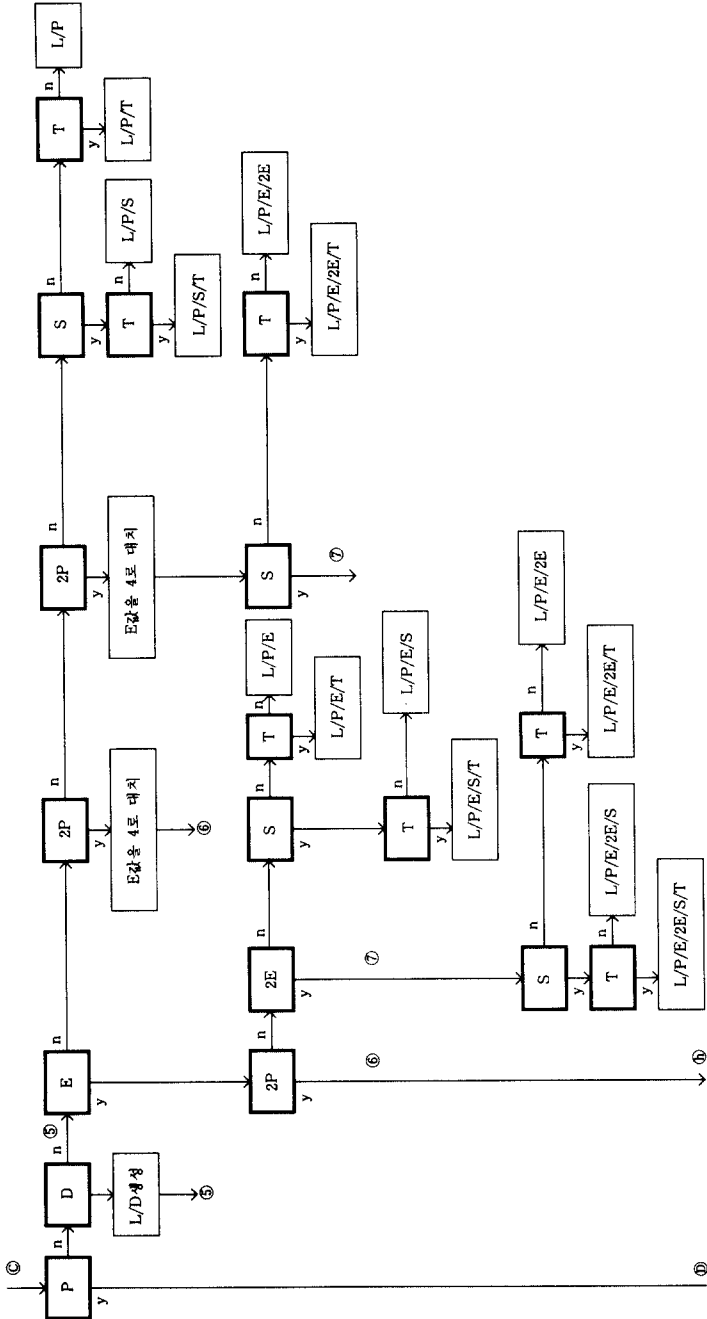
CC 6판에 의하면 의학분야 분류표는 p. 2.80에서 p. 2.86에 걸쳐 나타난다. 분류의 공식은 $L[P] : [E][2P]$ 이다. 의학분야 분류에서 특이한 점은 특정질병에 대하여 이미 조합된 분류번호를 배정하고 있는 점이다. $[P]$ 와 $[E]$ 의 개념들은 이미 열거되어져 있으며, $[2P]$ 는 $[E]$ 파셋트 중 4 disease 바로뒤에 결합하도록 규정하고 있는 파셋트로서 이것도 이미 열거되어 있다.

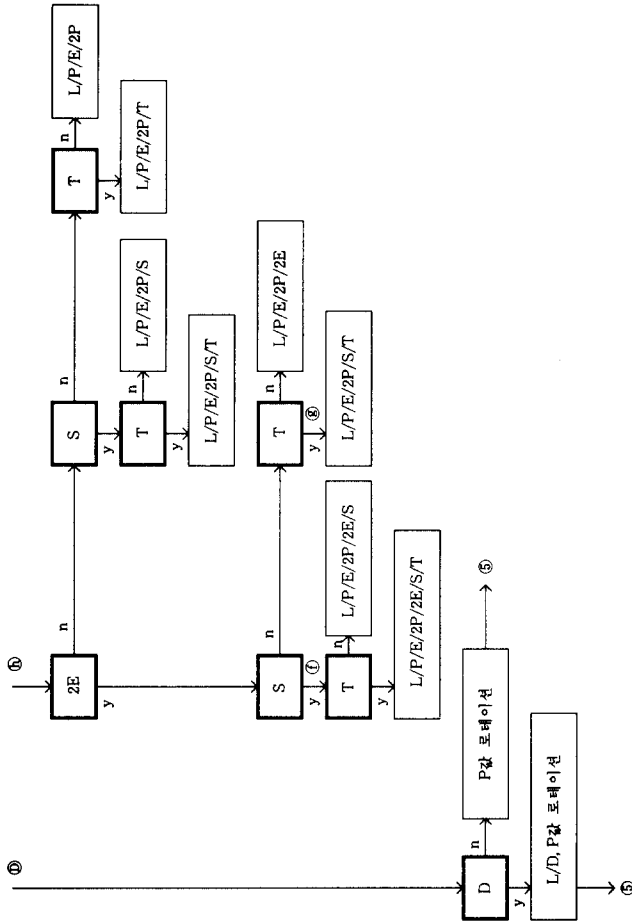
특히 특정질병의 분류번호의 구성을 보면, L:423 pox, 와 같이 분류하여 둔 경우와 L12:36 obesity와 같이 분류하여둔 2가지 경우로 대별된다. 위의 pox의 경우 분류번호 L:423은 개념의 조합이 L/E/2P로서 medicime(L)/ disease(E)/virus(2P)인 경우와 같게 나타나며, L12:26 obesity는 조합이 L/P/E로서 medicine (L)/tissue(P)/morphology(E))와 같은 형태이다. 그러나 단지 특정 질병의 경우는 이들 개념을 조합하여 분류하는 형식을 취하지 않고 바로 표상에 열거하고 있는 점이 특이하다. 이에 본 논문에서는 이들 특정질병에 대해서는 별도로 $[D]$ 의 카테고리 하에서 조합이 되도록 하였다.

나. 의학분야 분류자동화 원리유도

의학분야에서의 분류자동화의 원리는 $[P]$, $[D]$, $[E]$, $[2P]$, $[2E]$, 의 5개의







(그림 5) 의화본야 분류자동화 원리

학문적 성격의 카테고리과 공통적 성격의 [S]와 [T]카테고리로 조합이 이루어지게 된다. 단지 플로우차트상의 조합의 예는 개념의 조합에 따를 경우의 수를 도표화 한 것이며, 이것에 의해 바로 분류기호화 되지는 않는다. 이의 대표적인 예가 [D] 카테고리가 결합된 경우이다. 한 예로서 개념의 결합이 L/D/E/2P인 경우는 실제로는 L/D와 L/E/2P로 2개의 분류기호가 생성되도록 하였다. 의학 문헌에 대한 분류자동화 원리를 유도하여 보면 <그림 5>와 같다.

4. 주제의 자동인지 방법

가. 표제입력과 탐색

분류를 행하기 위해서는 먼저 문헌이 표제를 PC상의 키보드에 입력한다. 지식베이스상의 모든 용어는 소문자로 사용하는 것을 원칙으로 하였기 때문에 약어가 아닌 이상 전부 소문자로 입력하는 것이 좋다. 그러나 첫자를 대문자로 입력한 경우는 대문자를 소문자로 변환시켜 탐색한다. 입력상의 요망사항으로서 복수형의 단어는 단수로 입력하고, 반드시 복수로 사용되는 단어는 예외로 한다. 최종적으로 나타나는 마침점(.)은 입력하지 않는다.

나. 지식베이스에서의 탐색

키보드에서 입력한 표제는 프로그램에 의해 지식베이스상의 용어를 탐색한다. 문헌의 표제가 다음과 같이 A study of rice harvesting and storage in korea인 경우를 예를 들어보면,

K(1)	A study of
K(2)	A study
K(3)	A

문헌의 표제상에서 최대 30글자 수의 범위내에서 3단어까지 읽어 K(1)에, 2단어 까지 읽어 K(2)에, 1단어를 읽어 K(3)에 옮겨 놓는다. 탐색은 K(1)에서 K(3)

순으로 탐색하여, 어느 것이든지 먼저 탐색이 되면 지식베이스상에 기호를 메모리상에 기억시키고, 탐색된 용어의 길이만큼 문헌의 표제를 앞으로 이동시킨다. 그리고 K(1), K(2), K(3)에서 어떠한 탐색도 이루어지지 않으면 최종적인 K(3)만큼 앞으로 이동시켜 다시 반복한다. 따라서 위의 경우는 탐색이 되지 않으므로 K(1), K(2), K(3)에 다음 A와 같이 데이터를 옮겨 놓는다.

A	<ul style="list-style-type: none"> — K(1) — K(2) — K(3) 	Study of rice
		Study of
		Study

이때도 역시 탐색이 매치되지 않기 때문에 위의 과정을 반복한다. 탐색이 되는 시점은 B에서와 같은 상태에서 K(3)에서 최초로 탐색이 이루어지게 된다.

B	<ul style="list-style-type: none"> — K(1) — K(2) — K(3) 	rice harvesting and
		rice harvesting
		rice

이와 같은 과정을 반복하여 최종적인 탐색결과는 다음과 같이 나타난다.

Keyword	agriculture		Medicine
rice	[P]	381	
harvesting	[E]	7	7
storage	[2E]	7	8
korea	[S]	2	[S] 2

이제 이 탐색결과를 근거로 어느 주제에 용어가 많이 매치되었는지 계산하여 주제를 결정한다. 상기의 예에서는 너무 농학에 치우치고 있어 쉽게 주제를 결정할 수 있겠으나, 주제의 결정이 되지않는 경우에는 주제를 사서가 입력하도록 컴퓨터가 요구하는 때도 있다. 이 경우에는 반드시 주제를 입력하여야만 분

류가 가능하게 된다. 이렇게 주제인식이 되고나면 분류는 매치된 이들 용어를 근거로 해당주제분야의 분류조합공식인 자동화원리에 의해 조합이 되게 된다. 상기 문헌은 주제가 J로 결정되므로써 J/P/E/2E/S로 결합되어 분류는 J381:7:8.2로 조합된다.

IV. 분류자동화의 결과 및 분석

1. 분류자동화의 결과

이미 언급한 바 있는 지식베이스 설계원리에 따라 지식베이스를 구축하고, 분류자동화 원리에 의거하여 프로그램을 작성한 후 농학과 의학문헌을 대상으로 실제 문헌의 표제를 입력하여 분류를 행하였다. 농학관계 문헌은 총 283권이었으나, 이들 문헌중 조경학 원예학, 임업, 축산업 등 CC 6판에서 농학으로 취급하지 않고 별도의 학문으로 취급하고 있는 문헌을 제외한 143권의 도서를 분류한 결과는 <표 3>과 같다. <표 3>상의 분류기호는 자동분류를 행한 후 이들 기호에 의하여 재배열한 형태이다.

의학관계 문헌은 총 196권으로서 이 가운데, 생물학에 관련된 30권을 제외하였으므로, 실제 분류대상 문헌은 166권이다. 의학관계 문헌의 분류결과는 <표 4>와 같다.

분류기호의 배열에 있어서는 분류기호상에 포함된 기호의 컴퓨터 코드값에 따라 배열이 다르게 나타난다. 실제 분류를 행한 후 분류기호를 배열하게 되면, 표상의 배열과는 다른 형태가 된다. 그 이유는 [P]파셋트 앞에 부여하는 콤마(,), [M]파셋트 앞에 부여하는 세미콜론(;), [E]파셋트 앞에 부여하는 슬래쉬(/)의 값이 일정하지 않다. 즉 콜론과 세미콜론은 숫자보다 큰 값이며, 코머와 마침점 및 슬래쉬는 숫자보다 작은 값으로 나타난다. 이로인하여 컴퓨터 상의 배열을 그대로 적용하면 관련주제의 인접이 제대로 되지않는다. 따라서 <표 3>과 <표 4>상의 분류기호의 배열은 콤마, 세미콜론, 콜론, 마침점 및 슬래쉬의

값을 모두 숫자의 최저치(0)보다 작은 값을 부여하되, , < ; < : < . < / 의 순으로 값을 부여 한 후 배열한 것이며, 이러한 배열방식은 관련주제의 인접을 위하여 절대적으로 필요하다.

〈표 3〉 농학문헌의 분류의 예

J	Practical lessons in agriculture.
J	Agriculture in our lives.
J	Yearbook of agriculture.
J	Economics of agriculture.
J	Linear programming applications agriculture.
J	Behavioral change in agriculture.
J:1	The Lock of the land.
J:1	Fundamentals of soil science.
J:1	Land finance law.
J:1	Design on the land.
J:1:3	Introduction to soil physics.
J:1:3	Soil physics.
J:1:51	Soil erosion and conservation.
J:1:76	Social geography and land use.
J:1:76	Land use and Landscape planning.
J:1:76	Land into landscape.
J:1:76	Land-Use controls.
J:1:761.1	Administration of land reform in Korea.
J:1:77	Comparison of forest water and energy exchange models.
J:1:77	Water in landscape architecture.
J:1:771	Soil and water conservation.
J:1:771	Soil and water conservation engineering.
J:1:8	Soil survey manual.
J:1:85	Methods of soil analysis.
J:22	Foundations for managing the farm business.
J:22	Malabar farm.
J:22	The real book about farms
J:220	Bacteriology for dairy students.
J:221.9G	Dairy farming in the south.
J:222.413	Agricultural change and the peasant economy of south China.
J:24	Fertilizing for maximum yield.

- J:241 Nitrogen in agricultural soils.
 J:27 Farmers in business: studies in cooperative enterprise.
 J:438 Fundamentals of applied entomology.
 J:438 Practical insect pest management.
 J:52.1 A world geography of irrigation.
 J:55 Solar application in agriculture.
 J:56 Weed control handbook.
 J:7:90 Machines for the farm, ranch, and plantation.
 J:7:90 Principles of farm machinery.
 J:7:90 Farm machinery and equipment.
 J:7:90 Farm machinery and equipment, 2nd ed.
 J:7:90 Farm machinery and equipment.
 J:7:901 Tractors and crawlers.
 J:7:L Processing agricultural and municipal wastes.
 J. 1 Economic crises in world agriculture.
 J/M15 The farmer's age: agriculture 1815-1960.
 J01,02 Agricultural economics and agribusiness: and introduction.
 J015 Principles of agronomy.
 J015 Introduction to agronomy.
 J015 Agronomy journal.
 J016 Rural development and human fertility.
 J016 The Innovator's situation: upper-middle class conservatism
 J016:1:76 A Theoretical approach to rural land-use patterns.
 J016. 7 The face of rural america.
 J017 Energy in the transition from rural subsistence.
 J02 Agricultural economic analysis.
 J02 Agricultural economics.
 J025 Agricultural product prices.
 J025 Marketing of agricultueal products.
 J025 Agricultural product prices.
 J025:7:L1 Processing equipment for agricultural products.
 J026 Agricultural marketing.
 J0261, 0262 Livestock and meat marketing.
 J03 Handbook of agricultural occupations.
 J04.4 Asian agricultural survey.
 J041.5 Agrarian conditions in modern European history.
 J07 Summaries of studies in agricultural education.
 J073:5 Economic development and programming of rural education.

- J09 Agricultural development and economic growth.
 J09 The economics of agricultural development.
 J09 The administration of agricultural development.
 J09 .1 Agricultural development: an international perspective.
 J09 .74 Agricultural development in the mexican tropics.
 J095 Economic analysis of agricultural projects.
 J098 Agricultural geography.
 J1 Principles of horticulture.
 J1 Horticultural science.
 J1 Pollution in horticulture.
 J1 Cloning agricultural plants via in vitro techniques.
 J1 Fundamentals of horticulture.
 J1, 380.1 Plant science:an introduction to world crops.
 J1:4 Pathogens, vectors, and plant diseases
 J1:42 Herbicide resistance in plants.
 J1:423 Handbook of plant virus infections:comparative diagnosis.
 J1:424 Fundamentals of plant pest control.
 J1:6 Breeding plants for less favorable environments.
 J1:6 Progress in plant breeding.
 J1:6 Breeding plants resistant to insects.
 J10 The vegetable encyclopedia and gardener's guide.
 J10, 16 Vegetable gardening.
 J10, 16 Better Vegetables for your home Garden.
 J10, 17 Recent advances in the biochemistry of fruits and vegetable.
 J10, 17:7:L Handling, transportation and storage of fruits and Vegetable.
 J10:7:L Commercial vegetable processing.
 J10. 41 Chinese vegetables.
 J108:7 Vegetable seed production.
 J12 How to grow bulbs.
 J121 Potatoes.
 J15,16 Foliage house plants.
 J16 Introduction to floriculture.
 J16 Low maintenance gardening.
 J16 The Total book of house plants.
 J16 Basic gardening.
 J16 Garden trees.
 J16 Complete home gardening.
 J16 1,000 beautiful house plants and how to grow them.
 J16 The how and why of better gardening.
 J16 Gifts from the garden.

J16:1615	Organic Gardening.
J1671	Better Lawn for your home.
J1671	Sunset lawn and ground cover book.
J17	The first book of fruits.
J17	Orchards and their cultivation.
J17:24	Mineral nutrition of fruit trees.
J17:6	Methods in fruit breeding.
J17:7:L	Commercial fruit processing.
J1754.1	World soybean research conference II : abstracts.
J18	Bibliography of seeds.
J18:7	Seed production.
J251	Grass.
J3:7:1	Professional food preparation.
J3:7:1.7361	Made in Washington: food policy and the political expedient
J3:7:11.1	Global food interdependence.
J3.22	House hold consumption patterns of food grains in Suwon.
J38	Seed to civilization: the story of man's food.
J380,3:7	Crop science and food production.
J380:1:77	Soil, water and crop production.
J380:424	Pests of field crops.
J380:6	Beeding field crops.
J380:7	Crop production.
J380:7	Crop production: principles and practice.
J380:7	Principles of field crop production.
J381:7	Rice: production and utilization.
J382:7	Wheat: production and utilization.
J383:6	Oats: a guide to breeding and showing.
J385:7	Modern corn production.
J385:7:L	Corn: culture, processing, products.
J39:7:7	Tree nuts: production processing products.
J6	The flower encyclopedia and gardeners guide.
J662,10	Leafy salad vegetables.
J9d1	The green house environment.

〈표 4〉 의학문헌의 분류의 예

L	The principles and practice of medicine.
L	Ancient medicine and other treatises.
L	Medicine and partient care: a guide to important principles.
L	The medical & health sciences word book.
L:14	Cost containment in hospitals.

- L:14 Management of hospitals.
 L:14 Employee benefits: a guide for hospitals.
 L:17 Basic physics and measurement in anaesthesia.
 L:2 Anatomy: for students and teachers of physical education.
 L:2 Roentgenologic anatomy: an introduction.
 L:2 Grant's method of anatomy: by regions descriptive and
 L:2 Dynamic anatomy and physiology.
 L:3 The physiology of oral reconstruction.
 L:31 The complete pregnancy exercise program.
 L:3512:7 Oral and maxillofacial surgery.
 L;4 The hazard from dangerous exotic diseases.
 L:4:1 Promoting wellness in nursing practice.
 L:4:1 Legal accountability in the nursing process Irene Murchison.
 L:4:1 Using nursing research: discovery, analysis, and interpretation.
 L:4:1 The educational process in critical care nursing.
 L:4:3 Diagnosis of viral infections: the clinical laboratory.
 L:4:3 Current surgical diagnosis & treatment.
 L:4:311 The clinical measurement package.
 L:4:511 An introduction to epidemiology.
 L:4:56 Introduction to molecular immunology.
 L:4:56 Methods in immunology: a laboratory text for instruction
 L:4:56 Immunology III.
 L:4:56 Textbook of immunology.
 L:4:56 Clinical concepts of immunology.
 L:4:56 Immunology.
 L:4:625 Computers in radiotherapy: clinical aspects.
 L:4:625 Textbook of radiotherapy.
 L:4:625 A textbook of radiology and imaging.
 L:4:625 Physical foundations of radiology.
 L:4:6253 Illustrated guide to X-ray techniques.
 L:4:7 Biomaterials in reconstructive surgery.
 L:4:7 Fundamental skills in surgery.
 L:4:73 An introduction to neurosurgery.
 L:4:74 Orthopaedics: principles and their application.
 L:4:74 Manual of orthopedics.
 L:4:74 Campbell's operative orthopaedics.
 L:4:75 Atlas of limb prosthetics: surgical and prosthetic
 L:42 Infectious diseases and medical microbiology.
 L:4230 Essentials of medical virology.
 L:4230 Handbook of medical virology.
 L:430:34 Diagnostic methods in clinical virology.
 L:45373:3 Differential diagnosis and treatment of pediatric allergy.

- L: 4537: 56 Manual of allergy and immunology: diagnosis and therapy.
 L: 46 Living nutrition.
 L: 5 Introduction to public health.
 L: 549721 Tuberculosis and its prevention.
 L10: 4 Lymphatic system metastasis.
 L12: 26 The treatment of obesity.
 L12: 4 Connective tissue diseases.
 L130: 3 Human physiology the mechanisms of body function.
 L130: 47 Structure and function of the body.
 L18, 17 Reconstruction of the head and neck.
 L18, 17: 2 A colour atlas of head and neck anatomy.
 L183 The middle ear.
 L183, 41: 4 Boies's fundamentals of otolaryngology: a text of ear, nose
 L1830 Interpreting hearing aid technology.
 L185 Manual of ophthalmic terminology.
 L185 An atlas of ophthalmic surgery.
 L185: 44 Micorsurgery of the glaucomas.
 L1851: 4: 7 Ophthalmic lasers: photocoagulation, photoradiation, and
 L2: 3 Digestive system physiology.
 L20 Manual of clinical problems in internal medicine: annotated
 L20 Principles of internal medicine.
 L20 Problems in internal medicine: selected topics with annota-
 L21451: 3 Wheeler's dental anatomy, physiology, and occlusion.
 L21456 Dental management of the medically compromised patient.
 L252 Debates in nephrology.
 L252 Fundamentals of nephrology.
 L2521 Pdiatric nephrology.
 L291 The hepatic coma syndromes and lactulose.
 L291, 2918 Hepatic, biliary & pancreatic.
 L291: 4 Hepatology: a textbook of liver disease.
 L293: 46 Diabetes and other endocrine disorders during pregnancy and
 L2941 Renal medicine.
 L32: 321: 4 Heart disease: a textbook of cardiovascular medicine.
 L32: 37 The heart, arteries and veins.
 L32, 371: 4: 3 Nomenclature and criteria for diagnosis of diseases of the
 L32, 9a11: 4 Office management of medical and surgical heart disease:
 L320 Clinical cardiology.
 L321 Cardiovascular system.
 L322 Cardiac rehabilitation: implications for the nurse and
 L323, 32 Cardiac roentgenology: shadows of the heart.
 L3251 Principles of clinical electrocardiography.

- L326 Cardiac anatomy: an integrated text and colour atlas.
 L35, 361 Blood: atlas and sourcebook of hematology.
 L35: 4 Infectious blood diseases of man and animals: diseases
 L35: 4115 The radiology of childhood leukemia and its therapy.
 L351, 361 Color and fluorescein angiographic atlas of renal vascular
 L352 Clinical hematology.
 L358 New aspects of the structure, function, and synthesis of
 L361 Laboratory medicine: hematology.
 L361, 82: 4721 Ischemia and necroses of bone.
 L4, 45 Respiratory function of the lung and its control.
 L40 Introduction to respiratory care.
 L401 Respiratory Physiology.
 L401 Introduction to respiratory physiology.
 L401 Respiratory physiology.
 L41, 177: 4 Diseases of the nose, throat and ear.
 L45: 4 The Lung: structure, function and disease.
 L45: 5713 Automatic ventilation of the lunge.
 L5150 General urology.
 L5150 Basic clinical urology.
 L5150 Endoscopic opearative urology.
 L54: 491 Syphilis and other venereal diseases.
 L540 The year book of endocrinology.
 L540 Clinical endocrinology: a survey of current practice.
 L540 Hormones and cancer.
 L540 Textbook of indocrinology.
 L540, 1661 Biochemical endocrinology of the vertebrates.
 L540: 3 Reproductive endoclinology: physiology, pathophysiology and
 L541: 4: 3 Laboratory tests in the diagnosis and investigation of
 L5551, 554: 4: 7 Surgery of the vulva and vagina: a practical guide.
 L556 The breast: an atlas of reconstruction.
 L568 The mammalian testis.
 L568: 472 Testis tumors.
 L7 The human nervous system: an anatomic viewpoint.
 L7: 46 Nutrition and the developing nervous system.
 L70 Neurological anatomy: in relation to clinical medicine.
 L70 Principles of neurology.
 L70 Medical neurology.
 L70, 132 Disorders of the foot.
 L701 Diagnostic and statistical manual of mental disorders.
 L71 Atlas of the central nervous system in man.
 L72 The understanding of the brain.
 L72 Serial dissections of the human brain.

L72: 2	The human brain: an introduction to its functional anatomy.
L826, 116: 4	Bone ischaemia and infarction in sickle cell disease.
L83, 742	Clinical atlas of muscle and musclocutaneous flaps.
L83: 46	Drug treatment of the rheumatic diseases.
L86: 4: 7	Arthritis surgery.
L87	Symposium on the treatment of burns.
L87	Illustrated encyclopaedia of dermatology.
L87	Clinical dermatology.
L87, 70: 4: 3	Dermatology: diagnosis and treatment of cutaneous disorders.
L87: 4: 3	Differential diagnosis in dermatopathology.
L87: 4: 41	Histopathology of the skin.
L870	Cicatrix optima: techniques for ideal wound healing.
L96	Orban's oral histology and embryology.
L9a	Preparing to enter medical school.
L9a1	An introduction to profession of medical technology.
L9a11	Physician's handbook.
L9a11: 4: 74	Orthopadic physician's assistant techniques.
L9a4	The year book of nuclear medicine.
L9a4	Nuclear medicine.
L9a7: 5	Public health and the law: issues and trends.
L9a7: 5	Preventive medicine and public health.
L9a8	Biological foundationsl of psychiatry.
L9a8	The powers of psychiatry.
L9a8	An outline of psychiatry.
L9a8	Psychiatry.
L9a8	Psychiatry: essentials of clinical practice: with examina-
L9a8	Modern Clinical psychiatry.
L9c, 1830	Hearing in children.
L9c, 32: 4: 7	Infant and child care in heart surgery.
L9c: 42	Infectious disease of children.
L9c: 42: 3	Common bacterial infections in infancy and childhood:
L9c: 4537	Allergy in children: a guide to practical management.
L9f1	To Linde's operative gynecology.
L9f2	Color atlas of gynecological surgery.
L9f3	Gynawcological cancer.

2. 주제식별의 관점에서 본 분석.

컴퓨터가 문헌의 표제를 근거로 주제를 자동인지 하기란 쉬운일이 아니다. 어떠한 방법으로 주제를 인식하느냐 하는 방법론상의 문제이기는 하나, 본 연

구에서의 지식베이스 설계원리는 단지 몇개의 용어만으로도 주제인식이 가능하다. 그러나 문헌의 표제가의 주제성이 희박하거나 여러학문 분야의 용어를 고루 사용하고 있을 시는 주제인지가 어렵게 된다. 이 때는 컴퓨터에서 주제를 입력하라는 메시지를 제공한다. 분류자가 해당주제(main class)의 기호를 입력하면 분류를 계속한다.

분류자가 컴퓨터로 하여금 자동적으로 주제를 인식케 하는 방법으로서는 문헌의 표제를 입력하는 가장 기본적인 방법이 있다. 이 방법은 가장 기초적인 방법이며, 이것에 의하여 주제인지 및 분류가 되지 않을 때는 문헌의 목차에서 주제인식이 가능할 때까지 주요 키워드를 입력하는 방법 및 본문속의 주요 키워드를 입력하는 방법이 있다. 주제결정 후의 분류기호는 가능한 한 표제상의 키워드에 의하여 분류하는 것이 좋다. 목차나 내용상의 키워드에 의하여 분류를 하게되면, 동일한 문헌 이지만, 도서관에 따라 서로 다른 분류기호가 생겨나기 때문이다.

그리고 주제식별의 정확성을 향상하기 위해서는 무엇보다도 주제분야별 망라적인 용어수집과 철저한 분석에 기인한 지식베이스의 구축이 선결과제 하고 하겠다.

3. 분류기호에 대한 분석

문헌의 표제(표제만으로 주제인지가 되지 않을 시는 목차나 본문상에서 입력한 키워드)에 의해 주제가 인지 되고 나면, 다음은 분류단계로 접어든다. 분류는 지식베이스(분류용파일)상에서 탐색된 용어를 분류자동화의 원리에 의거하여 일정한 순서로 조합 함으로써 이루어진다. 이러한 분류기호의 조합상에 고려해야할 사항으로서는 다음과 같다.

가. 동일한 파셋트가 연속 출현하는 경우,

용어의 조합시에 나타날 수 있는 사항으로서는 같은 파셋트를 가진 용어가 2회이상 출현하는 경우이다. 농학의 경우 표제가 apple and orange의 개념의 조합이 J/P/P로 나타나 실제 분류시는 J371,372이나, 이때는 J372,371로 부출지시가

반드시 필요하며, 이와같은 경우는 이하의 파세트에서도 다같이 적용될 수있다. 그러나 동일한 파세트가 너무 반복하여 나타나면, 분류기호는 길어질 뿐만 아니라 혼란이 야기 되기 때문에 두번이상 출현시는 생략하는 방법도 고려하여 볼 수 있다 하겠다.

나. 분류기호에 영향을 미치는 용어 수의 제한

분석합성식 분류는 개념조합의 수에 따라 분류기호가 길어진다. 이 때 하나의 문헌을 기계적인 방법으로 자동처리 하기위해서는 분류기호에 영향을 미치는 전체 개념 수의 제한 및 동일개념을 어느정도까지 제한하여 줄 필요가 있다. 예를 들어 한편의 도서를 분류하는데 있어 전체 개념의 수를 10개 이내로 하되, 동일 파세트는 두개를 넘지 못하게 하는 등의 제약이 요구된다. 본 연구에서는 동일개념이 두개 이상인 경우는 분류상에서 제외하고, 사용된 동일개념 중 [P]파세트는 로테이션을 시키는 방법을 취하였다. 그러나 실제 적용시에는 [E]파세트도 로테이션이 고려될 수 있다. 이러한 문제는 방법론상의 문제인 만큼 도서관의 분류정책 여부에 따라 결정하여야 할 사항이라 하겠다.

4. 분류자동화상의 고려사항.

가. 표제(title)의 표기

문헌의 표제는 그 문헌이 무엇에 관하여 기록하고 있는지 가장 간단 명료하게 표현하고 있다. 따라서 표제상에 사용된 용어에 근거하여 분류를 하게되면 모든 도서관이 동일한 분류기호를 생성시킬수 있는 장점이 있다. 표제상의 사용 용어가 애매 모호하거나 이들 용어만으로 주제식별 및 분류가 불가능한 경우도 상당 수에 이를 것으로 사료된다.

따라서 향후의 분류자동화를 위해서는 문헌의 표제표기에 있어 상당히 신중을 기하여야 할 것으로 생각된다. 표제의 표기를 가능한 한 특정성이 있는 학술적인 것으로 기술하도록 권장하되, 저자나 출판관계자 모두가 관심을 가져야 할 사항이다. 단 하나의 대안으로서는 문헌 출판시에 ISBN(International Standard Book Number)를 부여하는 방식과 같이 서명에 이어 문헌의 내용을 나타

내는 몇개의 키워드를 동시에 표기하도록 하는 방법이다. 이러한 문제점들이 해결되면 도서관 및 정보센터 등에 있어 자료조직 특히 분류문제에 있어서는 크나큰 변혁을 가져올 것으로 보이며, 나아가 BSO(Broad System of Ordering)와 같은 스위칭언어(Switching language)를 대신 할 수 수 있어 양자의 기능을 겸비한 진일보한 시스템이 될 수 있다.

나. 각 학문분야별 분류자동화원리 정립

분류자동화는 문헌의 표제나 키워드 등을 컴퓨터에 입력하고, 컴퓨터내에서 이들 데이터를 근거로 주제를 인식함으로써 시작되며, 이 때 분류되는 조합방식도 상이하지만, 각 개념의 속성이 분야마다 일치하지 않는다. 한 예로서 대부분의 주제분야 내에서 지리적인 개념은 [S]파셋트를 사용하고 있으나, 역사학(V)에서는 분류요이 조합방법이, V[1P1];[MP1]:[E]:[2MP1]으로 [P]의 속성을 지닌 [1P1]가 지리구분에 해당하는 국가에 해당한다.

이에 개개 학문분야별 분류자동화 원리의 정립은 그 학문의 특성에 따라 별개의 분류원리를 정립하여야 한다. 이의 원리의 정립에는 란가나단이 제시하고 있는 개개 학문분야별 분류공식(formula)을 참고할 수 있겠으나 CC의 분류공식도 향후의 자동화분류를 위해서는 상당한 부부를 수정, 보완하여야 하리라고 본다.

다. 종합지식베이스(분류용파일)의 구축.

지식베이스(분류용파일)의 구축은 해당 학문의 특성에 따라 분류자동화가 가능한 분류자동화원리를 정립하고 난 후 가능하다. 본 연구에서와 같이 두 개의 주제분야가 아닌 다수의 학문분야를 대상으로 분류자동화를 실현하기 위해서는 자동화 원리의 정립에 이어 광범위한 용어를 전문적으로 수집, 분석하여 지식베이스를 구축하는 작업이 필요하다 하겠다.

라. 분류번호의 길이

분석합성식에 의한 분류는 기존의 열거식 분류방법에 비하여 분류번호가 길어지는 경우가 많다. 그 이유는 여러개의 개념결합으로 분류번호가 형성되기 때문이다. 이러한 단순한 이유 때문에 분석합성식의 분류방식에 익숙하지 못한

사람들은 분류번호가 길기 때문에 서가상에 배가하기에는 부적합하다고 생각하고 있다. 그러나 대부분의 문헌은 표제상에 두 세개 이상의 개념이 출현하는 경우가 더물기 때문에 서가상에 배가하기에 곤란한 정도의 분류기호는 그렇게 많지 않을 것으로 생각된다. 또 긴 분류번호라 하더라도 기존의 열거식 분류표상의 분류기호와 같이 분류번호가 연속된 상태는 아니고, 여러 부분으로 조합되어 있기 때문에 서배(書背)상에 계층적으로 표기할 수 있는 이점도 있어 해결가능성은 얼마든지 있다. 이미 란가나단의 마드라스대학(Madras University Library)도서관의 30,000권의 장서를 분류한 예²⁰⁾에서도 이를 증명해 주고 있다 하겠다.

마. 보다 복잡한 분류규정의 적용.

이 연구에서는 개념의 기본 카테고리에 의거하여 분류하는 방법을 취하였으나, CC상에서 언급하고 있는 공통구분기호(common isolate)나 열거식 분류상에 나오는 형식구분기호 등과 같은 규정을 적용할 수도 있다. 이러한 규칙을 적용하기 위해서는 적용규칙이 증가함에 따라 컴퓨터 자체의 프로그램적인 처리에만 의존하기에는 어려운 점이 수반된다. 이러한 규칙을 적용함에 있어서는 다소의 인위적인 입력조작이 필요할 것으로 사료된다. 한 예로서 입력 키이워드에 encyclopedia란 용어가 입력된다고 하여 이것이 사전인지 아니면 사전에 관하여 기술한 문헌인지 컴퓨터측에서 판단하기에는 어려운 점이 많다.

따라서 이렇게 복잡한 규칙을 현 시점에서 적용하고자 하면 문헌의 표제나 키이워드를 입력할 때 태그(tag)를 입력하는 방법을 이용할 수도 있다. 단지 이러한 방법을 적용하면 이러한 분류시스템은 완전 자동화라기 보다는 반 자동화적인 성격을 띠게 되고, 대신 분류는 그만큼 상세하게 분류할 수 있는 이점이 있다. 향후의 연구는 이러한 측면에서의 연구도 필요하리라 본다.

20) S. R. Ranganathan. *Colon Classification*. Susan Artandi ed. (New Jersey: The Rutgers Univ. Press, c1965), p.15.

V. 결 론

도서관업무 가운데 분류만은 아직까지도 수작업방식을 탈피하지 못하고 있다. 그러나 해를 거듭할수록 문헌의 출판량이 기하급수적으로 증가 하고 있고, 이용자들의 정보요구 또한 신속, 정확성과 더불어 깊이있고, 그들의 요구에 맞는 특정의 미시적 정보만을 요구하는 경향으로 나아가고 있어, 더욱더 어려운 현실을 직면할 것으로 보아진다.

그러나 지금 현재 대부분의 도서관에서 사용하고 있는 열거식 분류표로서는 이러한 상황과 요구를 능동적으로 대처해 나아갈 수 없을 뿐만 아니라, 향후의 정보화 사회에서 이러한 수작업방식의 분류시스템은 어떠한 방법에 의해서라도 기계적인 처리가 가능한 측면에서의 연구가 필연적이라고 하겠다.

이에 본 연구에서는 기존의 열거식분류 방법론이 아닌 분석합성식의 분류법을 응용하여 분류자동화의 가능성 여부를 진단하여 보고자 농학과 의학분야를 대상으로 총 1,275개의 용어를 수집하여 지식베이스를 구축하고, 실제 데이터를 입력하여, 컴퓨터에 의한 주제인식과 분류기호의 생성의 가능성에 대하여 진단하여 보았다.

본 연구의 결과를 요약하면 다음과 같다.

1. 본 연구는 도서분류 자동화가 가능하도록 분류용 지식베이스를 설계하였다.
2. 지식베이스의 설계는 농학과 의학으로 국한 하였다.
3. 지식베이스의 설계는 컴퓨터가 문헌의 표제나 키워드의 입력으로 부터 주제인식과 더불어 분류기호를 자동적으로 생성시킬 수 있도록 하였다.
4. 지식베이스상의 각 용어는 특정 주제분야 내에서의 위치를 기호로써 나타내었다.
5. 컴퓨터의 주제인식에 이어 분류는 개개 주제 분야별 분류자동화 원리에 의거하여 조합이 되도록 하였다.

6. 분류자동화의 원리는 란가나단의 CC의 원리를 응용하여 용어가 조합이 되도록 하였다.
7. 지식베이스는 농학분야 564개의 용어, 의학분야 792개의 용어를 수집, 분석하여 구축하였다.
8. 지식베이스 설계후의 실제분류는 농학관계 143권의 도서, 의학관계 166권의 도서를 대상으로 실험하였다.
9. 농학과 의학문헌을 대상으로 실제 분류하여 본 결과 분류자동화가 가능함을 입증 하였다.
10. 향후의 도서분류자동화를 위해서는 해당 주제분야의 망라적인 용어수집과 분석을 통하여 분류용 지식베이스의 구축과, 분류 자동화 원리를 유도하여 내는 일이 선결과제라 하겠다.

또한 분류자동화가 실현될 경우에는 다음과 같은 효과를 기대하여 볼 수 있다 하겠다.

1. 신속 정확한 분류를 기대할 수 있다.
2. 여러개의 개념으로 되어있는 정보를 정보의 손실없이 일직선상에 기호로써 표현할 수 있다.
3. 표제상의 개념이 한정적이고 명확할시는 도서관마다 동일한 분류기호를 만들어 낼 수 있다.
4. 이미 분류되어 있는 도서를 타자동화분류 시스템으로 변경하고자 할 때는 일시에 전체도서를 분류할 수 있다.
5. 분류표의 개정에 따른 재분류가 거의 필요없으며, 필요할 시는 단시간내에 기계적인 처리로서 재분류가 가능하다.
6. 분류표의 개정이 간편하다.
7. 동일자료에 대해서는 시스템마다 동일한 분류기호를 갖기 때문에 정보유통에 기여할 수 있다.
8. 이용자는 자기가 찾고자 하는 문헌의 분류번호를 사전에 주제나 키워드의 입력을 통하여 탐지할 수 있다.

9. 도서관의 토털시스템의 자동화에 기여할 수 있다.
10. 분류표를 항상 개정(update)된 상태로 유지할 수 있다.

참 고 문 헌

- Batty, C.D. *Introduction to Colon Classification*. Hamden: Archon Books, 1966.
- Classification Research Group, "Bulletin no. 11," *Journal of Documentation*, V.34, N.1 (March 1978), pp.21-50.
- Classification Research Group, "Bulletin no.9," *Journal of Documentation*, v.24, N.4 (Dec. 1968), pp.273-291.
- Coates, E.J., "Classification in Information Retrieval: the twenty years following Dorking," *Journal of Documentation*, V.34, N.4(Dec. 1978), pp.288-299.
- Dahlberg, Ingetraut, "Major Developments in Classification," *Advances in Librarianship*, V.7(1977), pp.41-103.
- Enser, peter G.B. *Automatic Classification of Book Material Represented by back-of-the book Index*. Ph.D. Thesis. University of Sheffield, 1983.
- Foskett, A.C. *The Subject Approach to Information*. 3rd ed. London: Clive Bingley, c1977.
- Gopinath, M.A.; C. Harishanlar, "Compound Subject in Biology and Agriculture: a quantitative analysis of facets," *Lib. Sci. Slant. Doc.*, N.2(June 1985), pp.113-119.
- Hunter, Eric J. *Classification made Simple*. Adershot: Gower, c1988.
- Kumer, Krishan. *The Theory of Classification*. New Delhi: Vikas Publishing House, c1979.
- Kumer, P.S.G., "International Study Conference on Classification Research (3)," *Her-*

- ald of Library Science*, V.14, N.1(Jan. 1975), pp.8—26.
- Nation Agricultural Library. *AGRICORA Subject Category Codes with Scope Notes*.
Beltsville: U.S. Dept. of Agricultural Library, 1990.
- National Library of Medicine. *Medical Subject Heading; supplement to Index Medicus*
Volume 30, 1989. Maryland: National Library of medicine, 1989.
- Navalani, K.; gidwani, N.N. *A Practical Guide to Colon Classification* New Delhi: Ox-
ford & IBH Publishing Co., 1981.
- Ranganathan, S.R. *Colon Classification*. 6th ed. New York: Asia Publishing House,
1960.
- Ranganathan, S.R. *Colon Classification*. 7th ed. Bangalore: Sarada Ranganathan En-
dowment for Library Science, 1989.
- Ranganathan, S.R. *Colon Classification*. Susan Srtandi ed. New Jersey: The Rutgers
University Press, c1965.
- Richmond, P.A., “The Future of Classification,” *Drexell Library Quarterly*, V.10, N.4
(1974()), pp.105—117.
- Salton G. *Dynamic Information and Library Precessing*. Englewood Cliff, N.J.: Prentice
-Hall, c1975.
- Satiya, M.P. *Manual of Practical Colon Classification*. New Delhi: Sterling Publishers
Private Limited, c1984.
- Vashista, Rama N., “Automatic Classification: some latest development, *Indian*
Librn, v.32,N.2(Sept. 1977), pp.51—53.
- Vickery, B.C. *Faceted Classification: a guide to construction and use of Special Schemes*.
London: Aslib, 1960.
- Yahmai, N. Shahla; Maxin, Jacqueline A., “Expert Systems: A tutorial, *JASIS*, V.35,
N.5(Sept. 1984), pp.297—305.
- 이경호, 심의순, “도서분류자동화 원리유도,” *도서관학* 논집, 제11집 (1984), pp.
32—66.

이경호, “CC의 구조적 분석을 통한 분류자동화 원리 유도,” 도서관학논집 제15집(1988), pp.113-151.

정영미, “문헌정보학연역 지식기반시스템에서의 지식표현,” 정보관리학회지, 제7권, 제2호 (1990), pp.35~57.

A Study on the Design of Knowledge Base for Automatic Book Classification

Lee Kyung-ho

〈Abstract〉

Though the computer has become deeply entrenched as the major tool in information processing(library works), it may be obvious that automatic book classification techniques are still under experimentation, and the techniques have not yet been tested against the criterion of usefulness.

The purpose of this study is to design of knowledge base for automatic book classification which can be put to use in library operation, and to present a methodology of application of the automatic classification into the library.

Since the enumerative classification schemes which are existing are manual systems, it cannot be applied to the automatic classification, the principle of faceted classification based on concept analysis is brought in and studied.

The result of this study are summarized as follows:

1. The design of knowledge base confined the field of agriculture and medicine.
2. If title is entered by the computer keyboard, it will be searched in knowledge base, and then be classified by the principle of automatic classification.
3. Program flowcharts are designed as a bases of classification procedures for automatic subject recongnition and classification.
4. 283 books in agriculture, 196 books in medicine were drawn at random from Taegu University Library and Young-Nal Medical Center Library respectively.
5. The expriment of automatic classification is performed 143 books in agriculture, 166 books in medicine except for other subject books.

6. It was proved that automatic book classification is possible by design of knowledge base.

In addition, the expected values from design of knowledge base for automatic book classification are as follows:

1. The prompt and accurate process of classification is possible.
2. Though some title is classified in any library, it can be classified the some classification number by a program.
3. The user can retrieve the classification codes of books for which he or she wants to search through the computer.
4. Since the concept coordination method is employed, the representing of a multi-subject concept is made simple.
5. By performing automatic book classification, the automation of total system can be achieved.
6. The efficient international information transfer will be advanced since all the institution maintain unified classification number.