

문헌분류 전문가시스템의 설계에 관한 연구

김 정 현*

<목 차>

- | | |
|-----------------------|--------------|
| I. 서 론 | 2. 지식베이스의 설계 |
| II. 문헌분류 전문가시스템의 연구동향 | 3. 추론기제 |
| III. 문헌분류 전문가시스템의 설계 | 4. 지식획득시스템 |
| 1. 시스템의 개관 | 5. 설명시스템 |
| | 6. 이용자인터페이스 |
| | IV. 결 론 |
| | 참고문헌 |

1. 서 론

1. 연구의 목적

정보화시대를 맞이하여 각종 정보와 문헌의 생산량이 급증함에 따라 방대한 정보 자료들을 체계적으로 수집 정리하여 이용자에게 효과적으로 제공하는 정보관리업무의 중요성이 더욱 강조되고 있다. 그런데 이러한 정보관리업무를 효율적으로 수행하기 위해서는 모든 주제의 자료들을 일관성있고 체계적으로 분류 하여 조직화 하는 일이 선행되어야 한다. 그러나 실제에 있어서 문헌분류는 고도의 지적인 판단과 경험을 요구하는 전문적 업무로서 분류전문가가 동일한 자료를 분류하더라도 분류자의 관점에 따라 달리 분류될 수도 있다. 이와 같이 문헌의 핵심주제에 대한 판단

* 경북대학교 도서관

이 분류자의 관점에 따라 다를 수 있으므로, 컴퓨터의 일관성 있는 논리적 처리의 도움을 필요로 하는 것이다.

일찌기 1960년대 부터 자동분류에 관한 연구가 이루어져 왔지만, 이는 자료를 서가상에 배열하기 위한 분류기호의 자동생성이 아니라 주로 정보검색 측면에서 용어분류에 치우쳐 왔던 것이다.

그후 1980년대에 들어오면서 인공지능분야의 하나인 전문가시스템(expert system)이 각종 전문적 경험적 지식분야에 응용되기 시작하면서 분류업무에도 새로운 가능성이 나타났지만 분류 전문가시스템의 설계에 있어서 가장 핵심이 되고 있는 지식베이스의 구축과 규칙 생성을 위한 전문가의 지식 및 경험을 종합하여 이를 규칙화하는 문제는 문헌분류업무를 시스템화 하는데 있어 아직 정형화되지 못하고 있는 부분이다. 또한 현재까지의 연구도 아직 일반적이며 단편적이고 초보적인 연구에 그치고 있어 실제업무에는 거의 활용되지 않고 있는 실정이다. 그리고 현재까지의 연구는 대부분 키워드나 주제개념이 서명이나 목차 등에서 쉽게 추출될 수 있는 자연과학이나 응용과학분야의 문헌을 대상으로 하고 있고, 이들 분야에 비해 더욱 추상적이고 주제개념보다는 형식성이 두드러지며 보조표적용이 매우 복잡한 문학류의 분류시스템에 대한 연구는 거의 이루어지지 않고 있다.

이상과 같은 관점에서 본 연구의 목적은 전통적인 수작업 분류방식을 탈피하고 컴퓨터에 의해 분류번호가 자동으로 생성되도록 함으로써 분류의 일관성과 정확성을 기하고, 문헌분류를 위한 인간의 지적노력을 최소화하는 전문가시스템을 설계하는데 있다.

2. 연구의 방법 및 범위

본 연구는 아직 초보적 단계지만 각종 문헌분류 전문가시스템의 설계에 대한 국내외 선행연구를 분석한 후, 이를 바탕으로 전문가시스템의 일반적인 구축원리에

따라 문헌분류, 그 중에서도 DDC 20 문학류(800) 만을 대상으로 문헌분류 전문가 시스템을 설계하고자 한다. DDC 20 문학류를 연구대상으로 한 이유는 분류업무중에서도 다른 분야에 비해 표제의 추상성이 두드러져 주제개념 추출이 용이하지 않으며, 독특하고 복잡한 형식요소로 구성되어 있어서 차후 다른 주제분야의 자동분류문제 해결이 보다 용이해질 것이라고 판단되기 때문이다.

II. 문헌분류 전문가시스템의 연구동향

전문가시스템의 응용연구는 다양하기 때문에 지금까지 개발된 여러 응용분야의 전문가시스템에 관한 연구를 모두 검토하기는 사실상 불가능하므로 본 연구에서는 문헌분류 업무분야의 전문가시스템에 관한 연구에 한해서만 살펴보고자 한다.

1. 국외의 연구동향

1980년대에 들어 오면서 인공지능분야의 하나인 전문가시스템에 대한 연구가 활발히 이루어 지면서 의료진단이나 분류업무에도 적용되기 시작하였는데 그 중에서 유명한 것이 스텐포드 대학교에서 개발한 의료진단용 시스템인 MYCIN이다. 이는 균혈증과 수막염의 진단 및 치료분야에 의사를 도와주는 전문가시스템인데, 전문가인 의사와 비슷한 수준으로 환자의 특성이나 병의 증상 등을 단계별로 분석하여 진단 및 처방을 하며, 추론행위에 대해 이용자에게 설명을 해주는 최초의 대형시스템이다¹⁾. 그런데 MYCIN의 경우처럼 환자의 특성이나 병의 증상 분석을 위한 지식이 명확하고도 적절한 형태로 표현되어 이용될 수 있다면 기본적으로 문헌분류업무에도 전문가시스템을 응용할 수 있다는 생각을 갖게 되었다. 최초로 이러한 가능

1) Donald A. Waterman, *A Guide to Expert Systems*. London, Addison-Wesley, 1986. pp.33-35.

성을 시사한 사람은 Clarke와 Cronin이다.

Clarke와 Cronin은 1983년 UDC분류규칙에 대해 다음과 같은 기본적인 지식표현규칙을 제시하면서 전문가시스템이 문헌을 분류하는 데에도 응용될 수 있다고 주장하였다²⁾.

IF 의학 THEN 61

IF 61 AND 생리학 THEN 612

IF 612 AND 위장 THEN 612.32

그러나 이들은 기본적인 발상만 제시하였을 뿐, 실제 시스템에 대한 실험은 실시하지 않았다.

그후 이러한 구상은 1985년 Burton의 연구에서 DDC 19를 대상으로 실제로 분류 전문가시스템이 구현되었는데, 이는 전문가시스템 개발도구인 ESP / Advisor와 Expert-Ease가 이용되었다³⁾. 그는 분류 전문가시스템으로 말미암아 어떤 사람이 분류를 하건 일관성이 있고 통일된 분류를 제공할 수 있을 것이라고 주장하면서, 분류대상자료의 분석은 분류자의 특권이며 해당번호를 찾기 위해 분류표를 찾는 단순작업의 측면을 줄이는 것이 그 목적이라고 하였다. 그러나 이 시스템이 실제 가동될 때 어휘통제의 문제, 탐색어(주제어) 확장의 문제들을 해결하기 위한 절차가 필요한데 여기에 대한 실제적인 방안을 제시하지는 못하였다.

1988년 이시카와는 자동화된 도서관시스템에서 보조기능으로서의 인간과 컴퓨터의 인터페이스 측면을 강조한 분류번호 자동부여시스템을 개발하였는데 이는 두개의 서브시스템 즉, 일본십진분류법의 내용을 데이터베이스로 구축하여 이를 이용해 분류번호를 검색하는 분류번호 검색시스템과 일본십진분류법의 분류번호 합성규칙에 의해 최종적인 분류번호를 출력해 주는 분류번호 자동부여 서브시스템으로 구성

2) A. Clarke and B. Cronin, "Expert Systems and Library / Information Work," *Journal of Librarianship*, Vol. 15, No. 4 (1983), pp. 277-292.

3) P.P. Burton, "Expert System in Classification," *Expert Systems in Libraries* / edited by F. Gibb, London, Taylon Graham, 1986, pp.50-66.

되어 있다⁴⁾. 이 시스템은 기계가독화된 일본십진분류표 데이터베이스를 기본전제로 삼고 있지만 개발 당시에는 임시로 필요한 주제분야만을 뽑아서 색인파일을 구성하였으며, 추론기능이나 설명기능 등이 없이 단지 분류표의 기능을 일부 자동화한데 불과하다.

한편 1988년 호소노 등은 NDC의 보험분야를 실험대상으로 분류전문가시스템을 개발하였다⁵⁾. 이 시스템에서는 IBM에서 개발한 Expert System Environment (ESE) 라는 전문가시스템 셸을 이용하여 지식표현은 생성규칙을, 추론기제는 후방향 방식을 적용하고 있는데, ESE에는 일본어가 지원되지 않아 키워드의 표기는 로마자로 하고 있다.

또한 1988년 나카무라와 이시카와는 기계가독화된 국제십진분류표의 파일을 이용하여 임의의 개념에 대한 UDC번호의 자동부여 전문가시스템을 개발하였다⁶⁾. 이 시스템은 앞서 논급한 이시카와의 경우처럼 기계가독화된 분류표를 응용하여 지식베이스에 필요한 색인파일들을 자동생성한 점이 특징이다.

1988년 Sharif는 분류전문가시스템 개발의 타당성분석을 위해 Stratclyde프로젝트를 수행하였다. 이 프로젝트에서 그는 마이크로컴퓨터상에서 운영될 수 있는 세 가지 종류의 셸, 즉 Expert-Ease, ESP / Advisor와 X I / X I Plus들이 단행본의 주제분류를 위한 전문가시스템의 개발에 적합한지 여부를 ① 산출된 분류번호의 정확성, ② 시스템의 응답시간, ③ 이용의 용이성, ④ 지식베이스(분류표)의 수정 가능성, ⑤ 시스템의 정보요구와 결론에 대한 이유를 설명하는 기능 등 다섯

4) Ishikawa T, "The Man-Machine Interface Aspect of an Automatic Classification Numbering System in a Computerized Library System," *Journal of Information Processing*, Vol. 11, No. 3 (1988), pp. 199-205.

石川徹也, 篠原きさ子, "日本十進分類法(NDC)を対象とする分類番號附與支援システム," *情報科學技術研究集會發表論文集*, 第22回(1985), pp. 201-209.

5) 細野公男, 田村俊作, 原田隼史, "圖書分類エキスパートシステム," *情報學基礎*, 8-4(1988), pp. 1-7.

6) Nakamura Y. and Ishikawa T, "Expert System for Automatic UDC Numbering Assignment," *Proceedings of the 44th Fid Conference and Congress part I (Helsinki, 1988)*, pp. 274-283.

가지 기준으로 조사하였다⁷⁾. 그리고 그는 DDC 19판의 농학분야로 분류지식베이스를 구성하였는데, 분석결과 이러한 셸들을 이용하면 시스템의 개발시간은 절약할 수 있겠으나 여러가지 제한점이 많았기 때문에 오히려 Prolog나 C, LISP같은 언어로 분류 전문가시스템의 개발을 주장하였다.

1989년 Endres-Niggemeyer와 Schmidt는 실험용으로 Herba Medica라는 의약품 및 관련분헌용 분류 전문가시스템을 개발하였다⁸⁾. 이 시스템은 지식베이스의 정형화에 패싯분류가 가장 적합할 것이라 가정하고, 실제로 약품분야의 모의 패싯 분류표를 만들어 Prolog에 의해 구축되었는데, 다른 주제분야로 확장하기 위해서는 관련지식의 패싯조직화가 선행되어야 하는 어려움이 있다.

1991년 핀란드의 Valkonen과 Nykanen은 Macintosh PC와 Hypercard소프트웨어를 이용하여 국제특허분류(IPC) 전문가시스템을 개발하였다⁹⁾. 이는 특허분류의 계층구조를 적절하게 활용하여 해당 분류번호가 단계적으로 적용될 수 있도록 지식베이스를 구성하였으며, 지식베이스와 추론엔진을 언어독립적으로 설계하여 이용자 인터페이스에 의한 자국어 접근이 가능하다.

전술한 이외에도 1990년 Gilroy는 Crystal이라는 셸을 이용하여 DDC에 대한 전문적 지식이 없는 초보자들도 쉽게 이용할 수 있는 분류전문가 시스템을 개발하였으며¹⁰⁾, 또한 Cosgrove는 n-Cube라는 셸을 이용하여 UDC를 대상으로 아이템

7) Carolyn A. Sharif, *Developing an Expert System for Classification of Books Using Micro-Based Expert System Shells*, (British Library Research Paper 32). Yorkshire: British Library Research and Development Department, 1988.

8) B. Endres-Niggemeyer & Bettina Schmidt, "Knowledge Based Classification Systems: Basic Issues, a Toy System and Further Prospects," *International Classification*, Vol. 16, No. 3 (1989), pp. 146-156.

9) P. Valkonen & O. Nykanen, "An Expert System for Patent Classification," *World Patent Information*, Vol. 13, No. 3 (1991), pp. 143-148.

10) Joyce Gilroy, "An Expert System for Classification," *Its News*, No. 22 (August 1990), pp. 39-40.

분류를 지원할 수 있는 시스템을 개발하였는데¹¹⁾, 이들은 앞서 논급한 사례들과 거의 유사하다.

2. 국내의 연구동향

국내에서 분류업무와 관련된 전문가시스템에 관한 연구는 1987년 정영미의 논문에서 비롯된다. 이 논문에서 정영미는 “전문가시스템으로서의 분류시스템은분류기호의 선택에 사용되는 분류표의 내용을 지식베이스에 담고 있어서 일련의 규칙에 의해 어떤 경우에 어떤 분류기호를 선택할 것인지를 결정하거나 도와주는 시스템이 될 것”¹²⁾이라고 주장하면서 다음과 같은 예 즉, ‘도서선택’이나 ‘수서’ 등의 단어가 특정한 수식 어구가 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 제시하고 있다. 이는 비록 구체적으로 지식베이스를 설계하고 시스템을 구현한 것은 아니지만 앞으로 분류 전문가시스템의 설계시 기본방향을 제시한 점에서 그 의의가 크다.

1990년 이용민은 DDC 20에 새로 추가 확장된 전산과학분야 본표(004-006) 및 관련분야와 이와 관련된 상관색인 일부를 대상으로 Turbo Prolog를 이용하여 프레임구조로 표현되는 지식베이스를 구축, 분류표 지식베이스와 문헌검색의 결합시스템을 실험적으로 개발하였다¹³⁾. 그런데 이 시스템은 지식베이스 자체가 일관성이

11) S. J. Cosgrove, "Item Classification Using the n-Cube's Hierarchical Knowledge Representation Schema," *Libraries and Expert Systems*/edited by C. McDonald and J. Weckert (London: Taylor Graham, 1990), pp. 88-98.

S. J. Cosgrove and J. M. Weiman, "Expert System Technology Applied to Item Classification," *Library Hi Tech*, Vol. 10, No. (1992), pp. 33-40

12) 정영미, "도서관·정보학에서의 인공지능의 응용에 관한 고찰," *圖書館學*, 제14집(1987), pp. 87-88.

13) 이용민, "온라인목록 주제접근점 확장을 위한 분류표 지식베이스설계에 관한 연구," 석사학위논문, 연세대학교 대학원, 1990.

없는 여러 분류항목을 계층적으로 분석하는 일반적인 시스템으로 설계되지 않았으며, 일부 주제분야에 한정된 모의시스템으로 설계되었을 뿐 DDC분류표 본표와 보조표의 조합과정은 시스템에서 제외되었다.

1990년 최광은 INIS(International Nuclear Information System) 데이터베이스의 분류표인 *INIS Subject Categories and Scope Descriptions*의 A14.1(plasma)을 대상으로 분류표지식베이스와 분류기호 부여규칙을 이용한 자동분류시스템을 실험적으로 구축하였다¹⁴⁾. 그런데 이 시스템은 대상분류표가 각 분야의 주제를 총망라하는 일반분류표가 아니라 원자력분야중심의 특수분류표이고 지식베이스내부에 축적한 정보의 한글화가 이루어지지 못하였기 때문에 불완전한 것이었다.

1992년 이경호는 콜론분류법(CC)의 패시방식과 기호법을 약간 수정적용하여 전문도서관에서 실용화할 수 있는 자동분류(AutoBC) 시스템을 개발하였다¹⁵⁾. 이 시스템은 분류자동화가 가능한 분류데이터베이스를 설계하여 구축하고, 도서의 표제나 키워드를 컴퓨터에 입력함으로써 주제를 자동적으로 인지시키고 표제속에 있는 키워드를 CC의 기호 조합방식으로 처리하여 분류기호를 자동적으로 생성시킬 수 있는 시스템이다. 여기서 가장 핵심되는 부분이 분류데이터베이스라고 할 수 있는데 분류데이터베이스 구축을 위해 미리 입력된 키워드에 한정되어 자동분류가 가능하며, 또한 이 시스템은 표본으로 선정한 농학이나 의학분야처럼 표제에 주제가 포함되어 있을 가능성이 높은 경우에는 효과적이지만 문학류와 같이 표제가 추상적이거나 애매한 내용이 많은 경우에는 적용에 무리한 점이 많다.

이상에서 국내외의 연구들을 분석한 바, 이들은 대부분 지식베이스내에 들어갈 분류지식의 축적방법, 이를 탐색하여 분류기호를 부여하기 위한 분류규칙, 분류자의 정보처리 행태에 관한 연구와 사용중인 분류표를 부분적으로 시스템화하여 실험

14) 최광, "분류표지식베이스를 이용한 자동분류에 관한 연구," 석사학위논문, 연세대학교 대학원, 1990.

15) 이경호, "콜론분류법에 바탕한 자동분류시스템의 개발에 관한 연구: 농학 및 의학 전문도서관을 사례로," 박사학위논문, 성균관대학교 대학원, 1992.

한 것이다. 그러나 이러한 시스템들은 전문가시스템 본래의 기능과 문헌분류의 특성이 제대로 반영되어 있지 못한 것이다. 분류 전문가시스템도 일반 전문가시스템과 마찬가지로 지식베이스를 정점으로 한 추론기제, 지식획득시스템, 이용자 인터페이스, 설명기능 등이 필요하며, 지식내용과 구조화방법에 있어서 문헌분류의 특성을 고려해야 하는데 이러한 요소들이 결여되어 있기 때문이다.

본 연구는 이러한 선행연구의 분석에서 드러난 문제점들을 개선하고 시스템의 성능을 보완하여, 새로운 문헌분류 전문가시스템을 설계하고자 시도한 것이다.

Ⅲ. 문헌분류 전문가시스템의 설계

1. 시스템의 개관

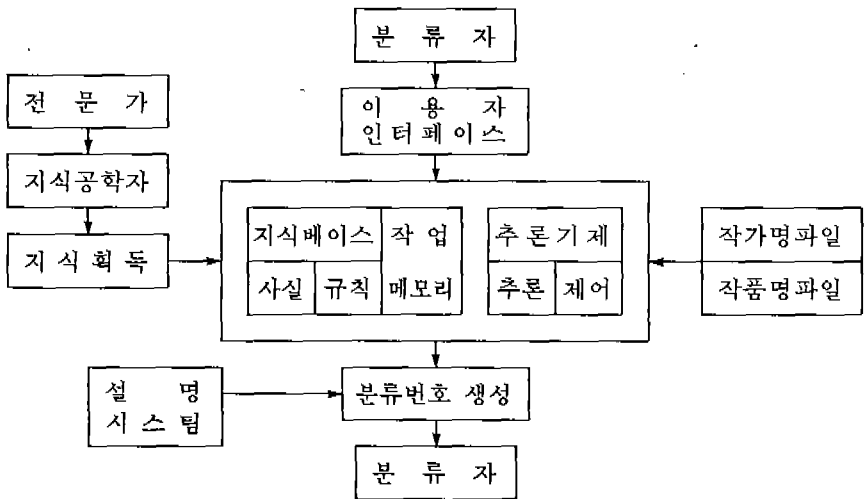
본 연구를 통해 설계하고자 하는 시스템은 DDC 20을 바탕으로 한 문학류의 문헌을 분류하기 위한 전문가시스템으로 <그림 1>과 같이 크게 지식베이스, 추론기제, 작업메모리, 지식획득시스템, 설명시스템, 사용자인터페이스, 그리고 작가명과 작품명파일 등으로 구성하였다.

문학은 다른 주제분야와는 달리 표제가 추상적이거나 애매한 경우가 많으므로 일정한 주제개념을 나타내는 키워드를 추출하기가 쉽지 않다. 특히 문학작품일 경우는 표제나 목차, 본문 모두가 대부분 추상적인 내용이므로 컴퓨터에 의한 주제개념의 자동추출은 매우 힘들다. 그러므로 문학류는 주제어의 자동인지에 의한 자동분류보다는 인간의 판단기능을 최대한 지원해 주는 형식의 전문가시스템이 바람직하다. 즉, 문학류의 분류과정을 철저히 분석하여 분류지식베이스와 추론기제를 설계하고, 각각의 분류단계에서 필요한 분류요소들을 메뉴화면에 체계적으로 제시함으로써 분류자로 하여금 주제의 추상성이나 애매함으로 부터 벗어나 필요한 분류항목을 쉽게 생각해 낼 수 있고, 분류번호 조합을 일관성있게 할 수 있도록 최대한

지원해 주면 되는 것이다.

따라서 본 시스템은 서명이나 키워드 등을 거의 입력할 필요없이 모두 대화형 메뉴방식으로 시스템을 설계하여 분류번호를 생성할 수 있도록 하였으며, 화면의 선택메뉴를 한글과 영문으로 함께 나타내었으므로 언어적인 문제에 구애받을 필요가 없고, 또한 책자형 분류표대신에 소위 온라인 분류표로도 사용할 수 있도록 하였다. 그리고 DDC 20 문학류의 분류규칙, 분류요소, 분류과정들을 지식베이스화 하여 추론기제를 통해 분류기호가 조합될 수 있도록 하였으며, 특히 전문가시스템의 일반적인 구성요소에는 없지만 문학류의 분류특성을 살려 작가명과 작품명 파일을 별도로 설계하였다.

한편 본 시스템의 구현은 개인용 컴퓨터를 이용하였고, 개발언어로는 마이크로소프트사의 C언어를, 한글처리를 위해서는 2바이트 조합형(도깨비 한글)을 사용하였다. 그리고 C언어의 컴파일러는 Micro Soft C 7.0버전을 사용하였고, 이를 위한 O/S는 마이크로소프트사의 MS-DOS 6.0을 이용하였다



〈그림 1〉 문헌분류 전문가시스템의 구성

2. 지식베이스의 설계

본 시스템에서의 지식베이스는 DDC 20 문학류의 분류에 대한 요소들과 규칙들을 저장하고 있다. 문학류의 각 분류요소에 대한 조합규칙은 생성규칙으로, 분류요소에 대한 세부항목은 프레임으로 표현하였다.

1) 생성규칙

문학류의 분류요소는 특정국어, 문학형식과 형식유형, 시대, 표준세구분, 특징/테마 등이며, 이들 요소들의 유무에 따라 분류번호가 조합된다. 본 시스템에서는 사용자가 쉽게 접근할 수 있도록 하기 위해 이들 각 요소들이 분류규칙에 따라 조합순서대로 화면에 나타나도록 설계하였다.

특정국어의 경우 DDC의 본표에 「미국문학이면 810에 분류한다」라고 되어있다. 이 규칙을 생성규칙으로 표현하면 「IF 미국문학 THEN 분류번호 = 810」이 된다.

이와 같이 특정국어에 따라 분류번호를 나타내는 규칙을 생성규칙으로 나타낸 예는 다음과 같다.

【규칙】

① IF 미국문학 THEN 분류번호 = 810

② IF 영국문학 THEN 분류번호 = 820

:

이러한 언어에 대한 적용이 끝나면 곧바로 문학형식이 적용된다. 영국문학을 대상으로 특정언어와 문학의 조합에 대해 생성규칙으로 나타낸 예는 다음과 같다.

【규칙】

① IF 영국문학(820) AND 시(1) THEN 분류번호 = 821

② IF 영국문학(820) AND 희곡(2) THEN 분류번호 = 822

:

이와 같은 방식으로 「영국문학 + 문학형식(시) + 형식유형」, 「영국문학 + 문학형식(시) + 시대구분」, 「영국문학 + 문학형식(시) + 시대구분(19세기) + 표준세구분」, 「영국문학 + 문학형식(시) + 시대구분(19세기) + 표준세구분(전집) + 특징/테마」 등을 각각의 생성규칙으로 나타낸 예는 다음과 같다.

【규칙】

- ① IF 영국시(821) AND 극시(02) THEN 분류번호 = 821.02
- ② IF 영국시(821) AND 설화시(03) THEN 분류번호 = 821.03
- :

【규칙】

- ① IF 영국시(821) AND 시대 1066-1400(1) THEN 분류번호 = 821.1
- ② IF 영국시(821) AND 시대 1400-1558(2) THEN 분류번호 = 821.2
- :

【규칙】

- ① IF 19세기 영국시(821.8) AND 전집(08)
THEN 분류번호 = 821.808
- ② IF 19세기 영국시(821.8) AND 평론(09)
THEN 분류번호 = 821.809
- :

【규칙】

- ① IF 19세기 영국시 전집(821.808) AND 특징/테마(사랑 : 354)
THEN 분류번호 = 821.8080354
- ② IF 19세기 영국시 전집(821.808) AND 특징/테마(역사 및 정치 : 358)
THEN 분류번호 = 821.8080358

:

한편 특정국어의 문학이 아니면서 문학사나 해설, 평론 등에 관한 저작은 809에 분류를 하는데 여기에 문학형식과 형식유형, 특징/테마, 시대 등을 조합하는 규칙을 생성규칙으로 나타낸 예는 다음과 같다.

【규칙】

- ① IF 809 AND 소설(3) THEN 분류번호 = 809.3
- ② IF 809 AND 소설(3) AND 형식유형(탐정, 미스터리, 서스펜스, 스파이, 고딕소설 : 0872) THEN 분류번호 = 809.3872

:

특정국어의 문학이 아니면서 전집 및 선집에 관한 저작은 808.8에 분류를 하는데 여기에 문학형식과 형식유형, 특징/테마, 시대 등을 조합하는 규칙을 생성규칙으로 나타낸 예는 다음과 같다.

【규칙】

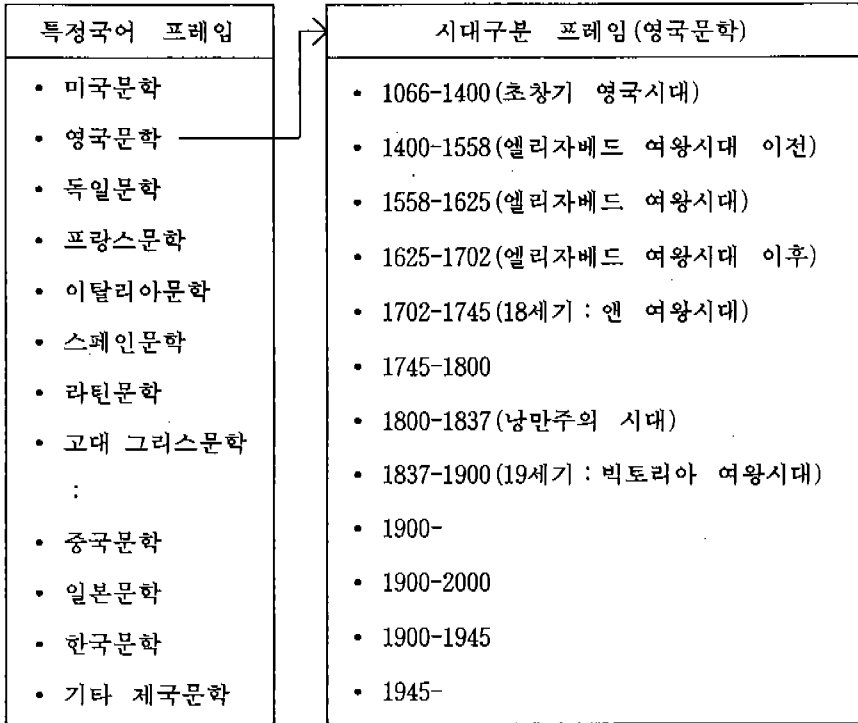
- ① IF 808.8 AND 소설(3) THEN 분류번호 = 808.83
- ② IF 808.8 AND 소설(3) AND 형식유형(탐정소설 : 0872) THEN 분류번호 = 808.83872

:

본 연구에서 시스템화하여 사용한 기본규칙의 수는 <표 1>과 같으며, 이들 규칙들이 분류규칙에 의해 자동조합되도록 설계하여 문학류의 주제를 대부분 나타낼 수 있도록 하였다.

이와같은 형식의 생성시스템으로 지식을 표현함으로써 우수한 점은 생성규칙 즉, 조건-행위와 같은 단편적인 지식의 단위가 많은 전문지식의 기술에 아주 적당하다는 점과 규칙의 형태를 인간이 쉽게 이해할 수 있다는 점, 그리고 지식을 모듈화할 수 있다는 점이다.

레이에 시대구분 프레임이 연결된 예이다.



〈그림2〉 특정국어 프레임

〈그림 3〉 시대구분 프레임 (영국문학)

〈그림 4〉는 형식구분 프레임이며, 〈그림 5〉는 형식유형 프레임인데 각 형식구분 프레임값마다 형식유형 프레임을 연결하였다. 〈그림 4〉와 〈그림 5〉는 시 프레임에 형식유형 프레임이 연결된 예이다. 그런데 여기서 〈그림 4〉의 형식구분 프레임은 〈그림2〉의 특정국어 프레임과도 연결되도록 하여 분류전개상 어려움이 없도록 설계하였다.

이러한 프레임의 장점으로는 복잡한 개념의 구조를 자유로이 기술할 수 있고 당면한 문제에 대한 추론제어를 실현할 수 있는 유연성도 지니고 있으며, 일관성과

함께 계층적인 구조를 가지고 있다는 점이다. 문학류의 분류요소에 대한 세부항목의 적용규칙도 이 범주에 속한다.

본 연구에서 설계한 위의 프레임구조들은 DDC 20 문학류에 적용되고 있는 대표적인 분류요소들이며, 분류번호 조합순서와 거의 동일하다. 예를 들면 먼저 특정국어 프레임중에서 분류할 자료의 유형이 영국문학일 경우는 영국문학 프레임을 불러낸다. 이어서 형식구분, 시대구분, 표준세구분, 특징/테마의 프레임을 불러내어 해당 항목들을 결정할 수 있도록 설계하였다.

형식구분 프레임
<ul style="list-style-type: none"> • Poetry (시) • drama (희곡) • Fiction (소설) • Essays (수필) • Speeches (연설) • Letters (서간문) • Satire and humor (풍자 및 유우머) • Miscellaneous writings (잡문)

〈그림 4〉문학의 형식구분 프레임

형식 유형 프레임 (시)
<ul style="list-style-type: none"> • Dramatic poetry (극시) • Narrative poetry (설화시) • Epic poetry (서사시) • Lyric and balladic poetry (서정시 및 민요시) • Sonnets (소네트) • Odes (송시) • Ballads (민요시) • Didactic poetry (교훈시) • Satirical and humorous poetry (풍자시 및 유머시) • Limericks (5행극시) • Light and ephemeral verse (경시 및 단일성시)

〈그림 5〉문학의 형식유형 프레임(시)

3) 작가명, 작품명파일

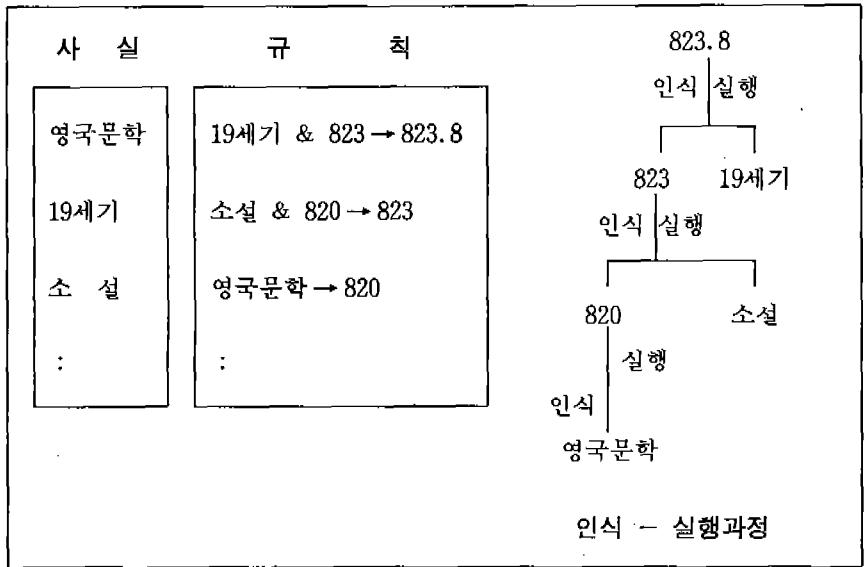
유명한 작가나 작품에 대해서는 계속 연구, 출판될 가능성이 높으며, 따라서 같은 작가나 작품에 대하여 반복 분류작업을 해야 할 가능성도 그만큼 높다. 그러므로 본 연구에서는 유명한 작가나 작품(학원출판사의 世界文學大事典을 참고하여 선정하였음)에 대하여 미리 분류작업을 하여 데이터베이스를 구축함으로써 분류의 일관성을 유지할 수 있고, 시간과 노력을 줄일 수 있도록 별도의 파일을 구성하였다.

3. 추론기제

추론기제는 전문가시스템의 지식베이스에 있는 사실과 규칙을 어떻게 적용시킬 것인가에 대한 제어구조를 나타낸다. 즉, 추론과 제어를 담당하는 부분으로 현 상태를 목표상태로 만들어 가는 것이다.

본 연구에서는 전방향 추론방식을 채택하였으며, 이는 인식-실행의 순서로 진행되는데 그 과정은 <그림 8>과 같다. 예를 들어 작업메모리에 「영국문학, 소설, 19세기」라는 사실이 있을 경우, 새로운 사실 「823.8」이라는 분류번호를 어떻게 추론하느냐 하면 <그림 8>의 규칙에서 「영국문학 ——> 820」이므로 「소설 & 820 ——> 823」의 규칙에서 「823」이 만족되고, 「19세기 & 823 ——> 823.8」이라는 규칙에서 「823.8」이라는 사실(분류번호)이 추론되었다. <그림 8>의 오른쪽 부분은 이러한 추론과정을 체인으로 표현한 것인데 인식-실행과정을 반복하여 분류번호가 생성된다는 것을 나타낸 것이다.

이러한 추론과정은 단말기의 선택메뉴 가운데 이용자가 해당 분류항목을 선택하면 자동적으로 이루어지며, 추론기제에서 지식베이스의 분류번호 생성규칙과 대조하여 적절한 분류번호를 자동으로 생성할 수 있도록 설계하였다.



〈그림 8〉 추론기제의 구조

4. 지식획득시스템

전문가시스템의 성능은 시스템이 가지고 있는 지식베이스 내용의 폭과 깊이에 좌우되며, 전문가시스템을 개발하는데 있어서 그 분야의 지식을 얻는 것이 가장 기본적인 동시에 가장 중요한 절차이다.

본 연구에서는 〈Dewey Decimal Classification and Relative Index(20th ed.)〉, Osborn의 〈Dewey Decimal Classification, 19th Edition: a Study Manual〉, Downing의 〈Introduction to Cataloging and Classification(5th ed.)〉, 정필모의 〈文獻分類論〉, 김명옥의 〈자료분류법〉, 학원출판사의 〈世界文學大事典〉 등의 문헌을 통한 공적인 지식과 분류업무 담당자의 실무전문가를 통한 사적인 지식을 수집하여 지식베이스를 구축하였다.

5. 설명시스템

전문가시스템의 특징 가운데 하나가 추론결과에 대한 설명이 있다는 점이다. 즉, 본 시스템에서도 설명 시스템이라는 별도의 모듈을 구성하여 이용자의 요청시 분류번호 생성단계에서 분류번호의 생성과정과 조합요소들을 보여줌으로써 분류번호의 생성결과에 대한 설명기능을 수행하도록 하였다. 이렇게 함으로써 추론과정을 유추하여 볼 수 있고, 생성된 분류번호를 다시 검토할 수 있도록 설계하였다.

〈그림 9〉는 〈Critical appraisal of 20c American short stories about children〉에 대한 분류번호 813.01099282가 생성된 결과를 설명한 예이다.

분 류 번 호 → 813.01099282	
위의 분류번호는	
다음의 분류요소들이 조합되어 생성된 것입니다.	
810	American literatures (미국문학)
3	Fiction (소설)
01	Short stories (단편소설)
09	History, critical appraisal (문학사, 평론)
9282	Children (어린이)

〈그림 9〉 설명시스템의 구성

6. 이용자인터페이스

본 연구에서는 대화형 메뉴방식을 채택하였는데 이는 화면상에 표시되는 메뉴선택에 따라 시스템과 이용자의 대화가 이루어지며, 추론이 실행되어 분류번호가 결정될 수 있도록 설계하였다. 즉, 가능한 어떤 명령어나 용어를 입력하지 않고서도

분류번호를 결정할 수 있도록 분류에 필요한 모든 사항들을 메뉴화면에 최대한 나타내어 초보자로부터 숙련된 분류전문가에 이르기까지 다양하게 접근할 수 있도록 하였다.

IV. 결 론

본 연구의 결과를 요약하면 다음과 같다.

본 연구에서 설계한 시스템은 시스템의 구성상 크게 두부분, 즉 전문가시스템과 작가명·작품명파일로 이루어져 있으며, 논리적으로는 작가명·작품명파일이 전문가시스템의 하부구조이다. 이 시스템은 문학류의 각종 분류요소를 조합하여 분류번호를 생성할 수 있도록 설계되어 있으며, 지식베이스, 추론기제, 지식획득시스템, 설명시스템, 이용자인터페이스로 구성되어 있다. 작가명·작품명파일은 유명한 작가나 작품명을 미리 수작업으로 분류하여 알파벳순으로 검색할 수 있도록 하였다. 지식베이스는 생성규칙과 프레임으로 표현하였고, 추론기제는 전방향 추론방식을 채택하였다.

본 연구에서 설계한 시스템은 다음과 같은 특징을 갖고 있다.

(1) 대화형 메뉴방식을 채택하여 명령어 입력방식에 비해 입력시간이나 오류발생을 줄일 수 있고, 초보자라도 쉽게 접근할 수 있는 이용자 지향적인 인터페이스로 설계하였다.

(2) 의사결정 지원형식의 시스템으로 설계하였으므로, 표제나 키워드에 의한 자동분류형식보다는 표제가 추상적이거나 애매한 문학류와 같이 주제개념보다 형식요소에 의해 분류번호를 조합할 때는 훨씬 효과적인 결과를 얻을 수 있다.

(3) 분류번호조합에 사용된 모든 요소들을 순차적으로 축적한 후, 이것을 분류번호 생성결과의 설명어로 사용하고 있어 조합과정을 보다 정확하게 추적, 검토할

수 있도록 하였다.

(4) 문학류의 분류특성을 살려 유명한 작가나 작품들을 미리 분류한 후 별도의 파일을 구축하여 둬으로써, 분류의 일관성을 유지하는 동시에 시간과 노력을 줄일 수 있도록 하였다.

(5) 메뉴화면에 한글과 영문을 함께 나타내었으므로 언어적인 문제를 쉽게 해결하도록 하였다.

(6) 지식베이스와 작가명·작품명파일에 소장된 지식을 모듈화하여 구성하였으므로 지식의 추가, 수정, 삭제가 용이하도록 하였다.

이상의 연구결과를 종합할 때 본 시스템은 다음과 같은 효과를 기대할 수 있다.

(1) 복잡하고 까다로운 분류규칙, 특히 문학류의 각종 형식요소 및 부가요소 적용에서 초래되기 쉬운 어려움을 쉽게 해결할 수 있을 것이다.

(2) DDC 20에 대한 기본적인 지식만 가지고 있는 초보자라도 본 시스템에 쉽게 접근할 수 있을 뿐만 아니라, 분류전담사서가 아니더라도 분류업무에 곧바로 적용할 수 있을 것이다.

(3) 책자형 분류표의 사용을 필요로 하지 않는, 소위 온라인 분류표로 사용될 수 있을 것이다.

(4) 분류 및 도서관자동화에 대한 교육강화 도구로도 이용될 수 있을 것이다.

(5) DDC 20의 문학류뿐만 아니라 나머지류에도 지식베이스를 확장함으로써 충분히 적용할 수 있을 것이다.

〈참 고 문 헌〉

김성혁. 주식투자 정보상담을 위한 전문가시스템에 관한 연구. 박사학위논문, 연세대학교 대학원, 1992.

이경호. 콜론분류법에 바탕한 자동분류시스템의 개발에 관한 연구; 농학 및 의학전

- 문도서관을 사례로. 박사학위논문, 성균관대학교 대학원, 1992.
- 이용민. 온라인목록 주제접근점 확장을 위한 분류표 지식베이스설계에 관한 연구. 석사학위논문, 연세대학교 대학원, 1990.
- 정영미. 도서관·정보학에서의 인공지능의 응용에 관한 고찰. 圖書館學, 제14집 (1987), pp. 67-92.
- 최 광. 분류표지식베이스를 이용한 자동분류에 관한 연구. 석사학위논문, 연세대학교 대학원, 1990.
- 細野公男, 田村俊作, 原田陵史. 圖書分類エキスパートシステム, 情報學基礎, 8-4(1988), pp. 1-7
- 石川徹也, 條原きさ子, 日本十進分類法(NDC)を對象とする分類番號附與支援システム. 情報科學技術研究集會發表論文集, 第22回(1985), pp. 201-209.
- Clarke, A and Cronin, B. "Expert Systems and Library/Information Work," *Journal of Librarianship*, Vol. 15, No. 4(1983), pp. 277-292.
- Cosgrove, S.J and Weiman, J.M. "Expert System Technology Applied to Item Classification," *Library Hi Tech*, Vol. 10, No. 1-2(1992), pp. 33-40.
- Dewey, Melvil. *Dewey Decimal Classification and Relative Index*. 20th ed. Albany, N. Y. : Forest Press, 1989. 4 vols.
- Endres-Niggemeyer, Brigitte and Schmidt, Bettina. "Knowledge Based Classification Systems : Basic Issues, a Toy System and Further Prospects," *International Classification*, Vol. 16, No. 3(1989), pp. 146-156.
- Ford, Nigel. *Expert Systems and Artificial Intelligence: an Information Manager's Guide*. London : LA, 1991.
- Gibb, Forbes ed. *Expert Systems in Libraries*. London : Taylor Graham, 1986.

- Gilroy, Joyce. "An Expert System for Classification," *Its News*, No.22 (August 1990), pp.39-40.
- Hayes-Roth, Fredrick etc. ed. *Building Expert Systems*. London : Addison-Wesley, 1983.
- Ishikawa T. "The Man-Machine Interface Aspect of an Automatic Classification Numbering System in a Computerized Library System," *Journal of Information Processing*, Vol.11, No.3(1988), pp.199-205.
- Morris, Anne ed. *The Application of Expert Systes in Libraries and Information Centres*. London : Bowker-Saur, 1992.
- Nakamura Y. and Ishikawa T. "Expert System for Automatic UDC Numbering Assignment," *Proceedings of the 44th FID Conference and Congress part 1(Helsingki, 1988)*, pp.274-283.
- Sharif, Carolyn A. *Developing an Expert System for Classification of Books Using Micro-Based Expert System Shells. (British Library Research Paper 32)*. Yorkshire : British Library Research and Development Department, 1988.
- Valkonen, Pekka and Nykanen, Olli. "An Expert System for Patent Classification," *World Patent Information*, Vol.13, No.3(1991), pp.143-148.
- Waterman, Donald A. *A Guide to Expert Systems*. London : Addison- Wesley, 1986.

Designing an Expert System for Library Classification

Kim Jeong Hyen*

〈Abstract〉

The purpose of the study is to design and implement a prototype expert system for library classification in the literature field of the DDC 20.

The system was largely consisted of a knowledge base, an inference engine, a knowledge acquisition facility, an explanation facility and an user interface facility. The knowledge base was represented by inference rules and frames. The name file for authors and titles was designed separately. The forward chaining technique was chosen for the inference engine and the menu-driven dialog technique was also taken for the user interface.

The conclusions of the study can be summarized as follows :

- 1) The difficulty of document classification work is due to the complex and stringent classification rules. Such problems can be considerably alleviated by using the present system.
- 2) Even the novice with a knowledge about the DDC 20 can easily access the system. And also librarian other than the professional classifier can easily be accustomed to the classification work.

* The Library of Kyungpook National University

- 3) The system can be used as an online classification scheme.
- 4) By adding any local language other than English or Hangeul on the menu screen, the language problem relating classification can be overcome.
- 5) The system can be employed as the intensification tool for the education of classification as well as library automation.