

DDC 20판의 주기 분석에 근거한 보조표 분류지원 전문가시스템 설계에 관한 연구

A Study on Classification Support Expert System Design
based on Note Analysis for DDC 20 Tables

김 상 미*, 남 태 우
전남대학교 문헌정보학과

Kim Sang-mi*, Nam Tae-woo
Dept. of Library & Information Science Chonnam National University

요 약

DDC 20판에서는 보조표 활용을 위하여 다양한 형태의 주기(Note)가 여러 곳에 마련되어 있다. 이 주기는 새로운 학문들이 이전판의 분류체계와의 중복성을 극복하고, 정확한 문헌분류를 위한 중요한 문법규칙들을 포함하고 있다. 그러나, 기술된 주기의 다양성이 제대로 정리되어 있지 않아서 이 주기의 활용은 미흡한 실정이다. 따라서, 본 연구는 DDC 20판의 보조표 T1(표준세분표: Standard Subdivisions) 및 T2(지리, 시대, 인물 구분표: Geographic Areas, Historical Periods, Persons)에 대한 이용주기를 통계적 빈도수를 고려하여 분석하고, 분석된 주기를 유형별로 분류하여 각 유형별 분류기호 생성 문법을 마련하였으며, 분류기호 생성 문법을 유도트리(Derivation tree)를 활용하여 정확한 분류과정을 예시하고, 이를 자동분류시스템으로 활용할 수 있는 분류지원 전문가시스템 모형을 설계하였다.

1. 서 론

1873년 멜빌 듀이(Melvil Dewey)에 의해 고안된 DDC(Dewey Decimal Classification)는 1989년 코마로미(J. P. Comaromi)등 4인의 편집진에 의해 20판이 출판되어 현재까지 발전하고 있다. DDC 20판은 학문의 학제간적 발전 및 현대분류이론의 특성을 포함하기 위하여, 열거식 분류형식은 제한되었고, 분석합성식 방법이 추가되었다. 따라서 보조표를 이용한 완전한 분류기호는 본표(Schedule)에서 전개된 분류기호에 미리 마련된 보조표(Table)를 사용해서 얻어진 첨가기호를 조합하여 생성된다.

그러나, 보조표를 이용하여 분류기호를 생성하라는 본표의 주기(Note)들은 매우 복잡하고 난해하여, 전문가사서들도 분류업무를 수

행하는데 많은 어려움을 느끼고 있으며 분류오류를 범할 가능성이 크다. 따라서 본 연구는 전문가사서들이 분류업무를 수행하면서 발생가능한 분류오류를 최소화하고, 보다 신속하고 일관성있는 분류결과를 얻을 수 있도록 주기내용을 정확히 분석하여 분류를 효율적으로 지원하는 분류지원 전문가시스템을 설계하였다.

본 연구의 구성은 제 1장에서 연구목적 및 필요성을 기술하였고, 제 2장은 DDC 20판의 보조표 이용주기를 분석하여 주로 T1 및 T2와 관련된 유형을 분류하고, 이를 활용할 수 있는 분류기호 생성 문법 및 유도트리를 마련하였다. 제 3장은 보조표 T1 및 T2 이용주기 생성규칙을 전문가시스템으로 활용할 수 있는 분류지원 전문가시스템을 지식베이스와 추론엔진으로 구분하여 설계하였다.

2. 보조표 T1 및 T2 이동주기의 분석

2.1 보조표 T1 및 T2 이동주기의 유형분석

DDC 20판의 주기에는 분류시 보조표를 활용하라는 주기가 다량으로 나타난다. 이러한 보조표 이용지시 주기는 기호의 구성에 필요한 것으로써 '기본기호에 보조표의 기호를 첨가하라', '이용하라', '적용하라' 등과, '다른곳에 분류하라', 우선순위 및 열기순서에 관한 주기 등이다. 이러한 모든 주기유형 중에서 가장 자주 나타나는 유형은 '기본기호에 보조표의 기호를 첨가하라'는 주기이다.

보조표 이용주기의 유형을 분석하기 위하여 보조표를 활용하여 분류하라는 지시가 나와있는 주기를 본표의 000-999에 걸쳐서 각 류별 및 보조표의 유형별로 조사하여 빈도수를 분석하였다. 분석결과 300대에서 가장 많은 보조표 이용주기가 적용 되었으며, 가장 다양한 유형의 보조표 이용주기 유형이 나타났다.

다음 <표1>은 300대에서 보조표 T1과 T2만을 적용하여 분류하라는 주기가 있을 경우 나타난 '유형 20가지의 유형번호와 그 유형이 갖는 특성을 요약한 것이다. 이 표에서는 T1을 먼저 적용시키는 경우와 T2를 먼저 적용시키는 경우로 분리한 다음 빈도수 순으로 정리하였다.

2.2 보조표 T1 및 T2 이동주기의 분류기호 생성규칙

300대의 T1 및 T2 이동주기에 나타난 분류기호 생성방법은 크게 기본기호에 보조표의 기호를 첨가하는 것, 본표의 기호를 첨가하는 것, 본표와 보조표의 기호의 일부분을 첨가하는 것 등의 3종류가 있는데, 본 연구에서는 이를 좀 더 세분하여 유형 20가지에 대한 분류기호 생성 문법을 트리 문법(Tree Grammar) G로 표현하고, 분류 예를 통하여 유도트리(Derivation Tree)를 생성하였다.

트리문법 G는 다음과 같은 4개의 요소로 구성된다.

<표1> 300대의 보조표 이용주기 유형

유형번호	유 형	유 형 특 성
유형1	B + T1	기본기호 + T1
*유형2	B(특정주제기호) + T1-T2	특정주제기호 + T1 혹은 T2
유형3	B + 0 + T1	기본기호 + 0추가 + T1
*유형4	B + T1P	기본기호 + T1기호의 일부분
유형5	B + 00 + T1	기본기호 + 00추가 + T1
유형6	B + T1 + T2 (B + 09 + T2)	기본기호 + T1 + T2 (기본기호 + T1의 기호-09 + T2)
유형7	B + 0 + MST1	기본기호 + 0부가 + 본표에 추가도 전개된 T1
유형8	B + 0 + MST1 + T2	기본기호 + 0부가 + 본표에 추가도 전개된 T1 + T2
유형9	B + T1 + T1 => B + T1	기본기호를 선정하여 부가 T1은 우선순위에 선행
*유형10	B + T2	기본기호 + T2
*유형11	B + T2 + 0 + MSP	기본기호 + T2 + 0부가 + 본표일부분
*유형12	B + T2 + MST1	기본기호 + T2 + 본표에 전개된 T1
*유형13	B + T2 + 0 + T2	기본기호 + T2 + 0부가 + T2
유형14	B + T2P	기본기호 + T2기호의 일부분
유형15	B + T2 + 0 + T1	기본기호 + T2 + 0부가 + T1
유형16	B + T2 + MST1(04)	기본기호 + T2 + 본표에 특별히 전개된 T1(04: special topics)
유형17	B + T2 + MSP(930-990)	기본기호 + T2 + 본표의 930-990의 기호의 일부분
유형18	B + T2P + 0 + MSP	기본기호 + T2의 일부분 + 0부가 + 본표의 일부분
유형19	B + T2 + 00 + MST1	기본기호 + T2 + 00부가 + 본표에 추가도 전개된 T1
유형20	B + T2 + T2 => B + T2	두나라 정부가 비교될 경우에는 T2에 선행되는 한 나라의 기호를 부가

* 적용빈도가 10회 이상인 유형
 <범례> B : Base Number(기본기호 또는 분류번호)
 T1P- T7P : 보조표1-7 기호의 일부분(Part)
 MSP : Main Schedule에 전개되어 있는 기호의 일부분
 MST1 : Main Schedule에 추가로 전개된 T1(표준새본)

$$G = (Vt, Vn, P, S).$$

- Vt : 종단기호(Terminal Symbol)
- Vn : 비종단기호(Non-terminal Symbol)
- P : 생성규칙(Production Rule)
- S : 시작기호(Starting Symbol)

종단기호 Vt는 분류과정에서 마지막에 생성되는 세부분류기호를 의미하고, Vn은 T1과 T2 보조표 및 본표의 일부에서 계속 분류가 필요하다는 것을 나타낸다. P는 분류기호 생성규칙으로서, Vt와 Vn의 조합으로 구성된다.

본 연구에서는 다양한 주기유형을 적용하여 반복되는 분류과정을 표현하고자 할 때, P를 다음과 같은 형식으로 사용한다. P는 "Ti → TjTk"같은 형태로 표현하고, 그 의미는 Tj 또는 Tk가 Ti에 의해서 유도된다는 뜻이다. 즉, Ti는 Tj 또는 Tk로 대체될 수 있다는 의미이다. 여기서 '→'는 Backus 정규형식(Backus Normal Form: BNF)[7]에서 '::'와 같고 '|'는 또는(or)의 뜻이다. S는 시작기호로서 비종단기호의 부분집합이다. 일반적으로 Vt는 영문자 소문자 또는 숫자로 표현하고, Vn은 영문자 대문자로 표현하며, Vt ∩ Vn = ∅ 이고, S ⊂ Vn이다.

20개의 주기유형으로 부터 유도되는 분류기호 생성문법은 다음과 같이 정의된다.

$$G = (Vt, Vn, P, S), \text{ 여기서}$$

$$Vt = \{ 0, 00, 01, 06, 065, \dots, \text{base-no} \}$$

$$Vn = \{ S, B, T, T1, T2, TIP, \dots, MST1 \}$$

$$P = \{ S \rightarrow B + T ;$$

$$B \rightarrow BNO ; BNO + 0 ; BNO + 00 ;$$

$$BNO \rightarrow \text{base-no} ;$$

$$T \rightarrow T + T ; T + 0 + T ; T1 ; T2 ;$$

$$TIP ; \dots ; MST1 ; MSP ;$$

$$MST1 \rightarrow MS04 ; MS06 ; MS07 ;$$

$$MS06 \rightarrow MS061 ; MS065 ; \dots ; 06 ;$$

$$MS065 \rightarrow 065 ; \}$$

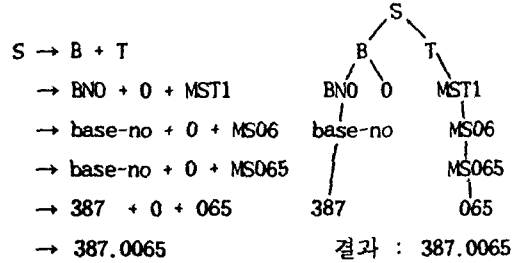
$$S = \text{시작기호}$$

보조표 이용주기에 의한 분류기호 생성문법에 의하여, 분류하고자 하는 문헌의 분류기호는 다음과 같은 유도트리에 의하여 생성된다. 이 각각의 유도트리는 전체 분류기호 생성문법트리의 부분트리들이다. 전체분류기호 생성문법트리는 분류하고자 하는 주기의 유형크기에 따라 결정된다. 다음 <그림1>에서 적용된 예는 유형별 분류기호 생성문법에 근거한 유도트리이다.

생성문법 적용

유도트리

유형 7 예 : Watertransportation Business.

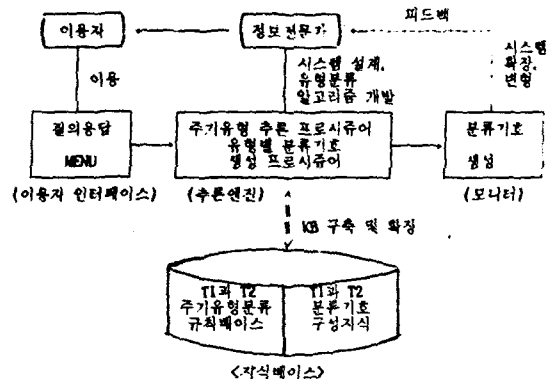


<그림1> 분류기호 생성 유도트리 예

3. 주기분석에 근거한 분류지원 전문가시스템의 설계

3.1 시스템 구성

DDC 20판의 보조표 이용주기를 활용하기 위한 분류지원 전문가시스템은 메뉴방식의 사용자 인터페이스, T1과 T2의 분류기호 구성 지식베이스, 주기유형 분류 규칙베이스, 주기유형을 분류할 수 있는 추론엔진 및 유형별 분류기호 생성 프로시저어들로 구성된다. 다음 <그림2>은 분류지원 전문가시스템의 시스템 구성도이다. 여기에서 정보전문가는 지식공학과 문헌정보에 전문성을 가진 사람을 의미하며, 이용자는 분류사서이거나 정보관리자이다.



<그림2> 분류지원 전문가시스템 구성도

3.2 지식베이스 설계

분류지원 전문가 시스템에 저장될 지식베이스 내용은 크게 T1 및 T2 분류기호 구성지식과 주기유형 결정을 위한 규칙들로 구성된다. T1 및 T2의 분류기호 구성 지식은 DDC 20판의 보조표를 그대로 활용한다. T1 및 T2의 분류기호 구성 지식은 T1과 T2 요약표에 나타나는 내용을 계층적으로 지식베이스의 사실로 저장하고 이와 관련된 질의어가 나타나면 자동적으로 화면에서 나타나서 이용자가 한 주제를 메뉴방식으로 선택한다. 따라서, 세부 주제 내용은 질의어와 쌍으로 지식베이스에 저장될 수 있도록 설계한다.

T1 및 T2 활용을 위한 보조표 이용주기 유형 분류규칙베이스는 IF-THEN 형식으로 구성된다. 보조표 이용주기 유형 분류를 위한 정보는 이용자가 이용자 인터페이스를 통해 시스템에게 전달해야 한다. 위의 모든 규칙들은 이용자 인터페이스를 통하여 이용자가 분류시 필요한 사항을 제시할때 한 규칙이 결정된다. 따라서, 추론엔진은 이러한 생성규칙과 이용자의 질의응답내용을 효과적으로 처리하는 정합(matching) 알고리즘을 설계하여야 한다. 즉, 추론엔진은 이용자 질의응답 결과에 따라 주기유형이 결정되도록 한다.

3.3 이용자 인터페이스 및 추론엔진 설계

분류지원 전문가시스템의 이용자 인터페이스는 메뉴방식을 선택하여 시스템이 질문과 선택사항을 제시하면, 이용자는 마우스 또는 키보드를 이용하여 한 항목을 선택하도록 한다. 이용자 인터페이스를 통하여 메뉴에 나타나는 주요질문사항은 주기유형 결정을 위한 질문들이다.

추론엔진은 이용자 인터페이스를 통한 질의응답과정을 제어하고, 지식베이스 검색을 통하여 보조표 이용주기의 유형을 결정한다. 주기의 한 유형이 결정되면 추론엔진은 그 유형에 맞는 분류기호 생성 프로시저를 호출하고 기본기호에 새로 생성된 세부분류기호를 첨가하여 화면에 출력하도록 한다. 만일 이용자가 분류기호 생성과정 또는 추론과정을 확인하고 싶을 때는 후진추적기법

(Back tracking)을 사용하여 추론과정을 화면에 제시하고, 그 결과를 검증하도록 한다.

4. 결 론

본 연구에서는 DDC 20판의 보조표 이용주기를 활용한 분류업무의 신속성과 일관성을 향상시키기 위해서 분류지원 전문가시스템을 설계하였다. 연구내용은 첫째, 300대에서 보조표 T1 및 T2만을 적용하여 분류하라는 이용주기를 분석한 결과 20가지의 주기유형이 나타났다. 둘째, 300대에 나타난 20가지의 유형이 갖는 특성을 요약 정리하여, T1 및 T2 이용주기의 생성규칙을 트리문법(Tree Grammar)으로 표현하였고, 유도트리에 의하여 분류기호 생성과정을 예시하였다. 셋째, 보조표 T1 및 T2 이용주기 생성규칙을 전문가시스템으로 활용할 수 있는 분류지원 전문가시스템을 메뉴방식의 이용자 인터페이스, T1 및 T2의 분류기호 구성 지식베이스, 주기유형 분류 규칙베이스, 주기유형을 분류할 수 있는 추론엔진 및 유형별 분류기호 생성 프로시저들로 설계하였다.

본 연구와 관련된 다음 연구는 DDC 20판 주기의 다양성을 더욱 정확하게 분석하고 지식베이스 내용을 확장하여 실제 분류업무에 활용할 수 있는 시스템을 구현하는 것이다.

<참고 문헌>

- [1]김정현, "분류전문가 시스템 개발을 위한 분류업무 분석: DDC 20판 문학류를 중심으로," 도서관, vol.48, no.6(1993), pp.40-61.
- [2]남태우, "DDC 20판의 특성에 관한 연구," 현대사회과학연구 1, 전남대학교 사회과학연구소, 1990, pp.73-105.
- [3]Burton, P.F., "Expert Systems in Classification," In: *Expert Systems in Libraries* /ed. by F. Gibb, London, Taylor-Graham, 1986, pp.50-66.
- [4]Chen, L.M. *Cataloging and Classification An Introduction*. New York, McGraw-Hill Book Company (McGraw-Hill series in Library Education), 1985.
- [5]Clarke, A. and Cronin, B., "Expert Systems and Library/Information work," *Journal of Librarianship*, vol.15, no.4(1983), pp.277-292.
- [6]Dewey, M., *Dewey Decimal Classification, 20th ed.*, edited by J. P. Comaromi. Albany, Forest Press, 1989.
- [7]Johnsonbaugh, Richard, *Discrete Mathematics*, Rev. ed. New York, Macmillan Publishing Company, 1986.