

# 법률정보시스템의 색인에 관한 연구

## -특히 2차 법률정보를 중심으로-

노 정 란\*

### 목 차

1. 서론
2. 정보시스템의 색인방법
3. 법률정보시스템의 특성
4. 결론

### 1. 서론

법률정보는 개인이나 기업의 활동영역이 확대되고 전문화, 국제화되면서 법률분쟁이 잦아지게 되고, 이를 규정하는 법령의 수가 증가하고 복잡해짐에 따라 그 수요가 크게 늘고 있다.<sup>2)</sup> 현대사회에서 법과 개인(또는 법인)은 불가피하게 밀접한 관계를 맺고 있어 미국<sup>3)</sup> 등 불문법을 채택하는 서구 산업국가에서는 일찌기 법률과 판례 등 1차 법률정

\* 성업공사 종합기획부 과장  
접수일자 1997년 12월 10일

- 1) 원문 그대로의 법률(시행령, 시행규칙 포함), 예규, 조례, 판례 등을 1차법률정보라 정의한다면 이러한 1차법률정보의 이해를 돕기 위해 마련한 법해석서, 주석서, 일반교재류, 저널기사, 논문 등을 2차법률정보로 볼 수 있다.
- 2) 한국 갤럽, 한국 데이터통신(주) 공동 조사 (1987.12.12~12.24) : 전문정보로서는 생활정보를 제외하고는 요구도가 가장 높은 것으로 나타났다.
- 3) 미국 대표적 법률데이터베이스는 LEXIS와 WETLAW로 1980년 개발 완료하였다.

보에 관한 데이터베이스를 구축하여 이미 상용화 수준에 이르고 있고, 전문(Full-text) 데이터베이스의 시효가 법률데이터베이스인 것도 그와 같은 사실을 뒷받침하는 것이다. 그러나 법률주제분야의 난해한 속성으로 인하여 2차 법률정보 데이터베이스의 경우 그 중요성에도 불구하고 색인방안을 찾지 못하고 답보상태에 머물러 있다.

현실적으로 전문가 및 실무자들은 법령, 판례와 같은 1차 법률정보보다는 1차 법률정보를 알기 쉽게 풀이하거나 관점에 따라 분석한 주석, 평석, 논문과 같은 2차 법률정보에 더 많이 의존하며<sup>4)</sup> 이러한 2차 법률정보의 이용 (Digestion) 결과 생산된 또다른 2차 법률정보는 1차 법률정보의 제·개정, 폐기 (파기) 등에 영향을 미칠게 되는 연결고리를 갖고 있어, 2차 법률정보의 체계적 관리가 시급하다.

1차 법률정보 데이터베이스의 경우 법령은 법조문에 대한 '정의'의 형태를 띠고 있으며, 판례는 법령의 법조문을 그대로 사례에 적용하여 사건번호에 선고연월일, 사건명 등을 적시한 것이므로 이에 관한 데이터 요소의 추출 및 검색이 비교적 평이할 뿐더러 또 용어자체의 난해성은 시소러스<sup>5)</sup>를 통해 제어할 수 있다. 국내의 경우 1980년대 후반 이후 법률 데이터베이스<sup>6)</sup>, 판례데이터베이스<sup>7)</sup>가 각기 개발되어 상용되고 있다.

그러나 1차 법률정보와는 달리 2차 법률정보데이터베이스는 데이터

4) M. A. Voges, "Information Systems and the Law", Annual Review of Information Science and Technology 23 (1988) : 193-215.

5) 1994.12. 한국법제연구원은 개발완료(1993년)한 법령 데이터베이스 (KOLDS, Korean Law Database)의 효율적 관리와 이용을 위하여 법률정보시소러스 (책임연구원: 노정란, 연구원: 장길재)를 개발하였다. 시소러스는 색인자와 검색자가 일관되게 주제를 표현할 수 있게 하고, 검색시 용어간의 계층적 관계를 이용하여 보다 포괄적으로 또는 보다 특정한 주제로 검색할 수 있게 하는데, 본 시소러스는 모든 법률을 망라하였으며, 법률 전문가와 일반 이용자 양 측면을 고려하여 수집된 용어를 수차 클러스트하여 용어를 표준화하고 동의 관계·계층관계·관련 관계 등을 체계화하였다.

6) 한국법제연구원은 1993년부터 「대한민국 법령데이터 베이스」를 PC통신을 통하여 유료로 일반인에게 공개하고 있다. '97. 8.27 현재 대한민국 현행법령 (현재 유효한 모든 법령 약3,500건), 대한민국 조약 (현재 유효한 조약 약 250건), 대한민국 현행 영문 법률 (중요한 법령 약230건), 입법 예고 법령 ('95.10.1분부터)을 법령명별, 용어별 검색이 가능하며 시소러스를 이용한 검색은 현재 연구 중이다.

7) 법원행정처는 KAIST (한국과학기술연구원)와 공동으로 2차례 판례 데이터 베이스화를 추진하였으나 전문적인 법률용어의 난해함 때문에 실패한 후, 현직 개인 판사에 의하여 가장 먼저 성공하였다. 현재 모든 PC통신이나 상업기관의 CD-ROM 등을 통하여 해방 이후 생산된 모든 대법원 판례를 재판종류·법원·쟁점·출전별로 쉽게 검색할 수 있다.

요소의 추출 및 검색어의 구성에 있어서 보다 복잡한 문제점을 수반한다. 왜냐하면 자연과학 분야 등과는 달리 2차 법률정보의 경우 정형화된 서지정보 수준으로는 저자가 본문에서 다루고 있는 복잡한 법리를 수용 또는 표현할 수 없으므로, 본문의 정밀분석을 통하여 저자가 의도한 개념어를 추출하는 작업이 선행되어야 하는데 그러한 작업은 법률 전반에 대한 완벽한 전문지식을 요구하기 때문이다.

정보시스템의 성능을 결정하는 여러 요소<sup>8)</sup> 가운데 경제성과 신속성 등이 이미 정보기술의 발전과 더불어 상당히 향상된 수준에 이른 상태에서 앞으로 전문데이터베이스는 어느 정도의 빠른 속도로, 얼마나 많은 정보를 검색하느냐의 관점이 아닌 부적합정보를 걸러내기 위한 시간과 지적 노력과 비용을 최소화하면서 정확한 정보로 곧바로 연결될 수 있는 '정확성'의 관점에서 논의될 것이며 그 핵심에는 색인이 자리하고 있다.

따라서 본 연구는 전문 데이터베이스 기반의 2차법률정보 데이터베이스의 설계에 있어서 가장 중요하게 다루어야 할 색인에 관하여 다음과 같은 의문점을 갖고 출발한다.

법률정보의 주제적 특성상 일반적인 자연어 또는 통제어 색인기법이고 과연 유용할 것인가?

이 글은 그에 대한 해답을 도출하기 위하여 먼저 정보시스템의 색인을 위한 기초 개념으로서 '언어'의 역할을 정의하고, 두번째 최근까지 국내·외에서 연구 발표된 많은 자동색인기법에 대하여 개략적으로 살펴봄으로써 색인에 대한 이론적 연구를 수행한 다음, 세번째 법률정보의 특성을 연구하여 그 특성에 기인한 색인기법을 제안하고자 한다.

## 2. 정보시스템의 색인방법

### 2.1 색인과 정보검색과의 관계

8) 정영미, "국내문자정보 데이터베이스의 색인에 관한 연구." 情報管理學會誌 13:1(1996.6.):19-45.

정보시스템의 성능은 특정성, 망라성, 속도, 편의성, user의 만족정도 등에 의하여 측정될 수 있다. S. Dutta and P. K. Sinha, "Pragmatic Approach to Subject Indexing : A New Concept," Journal of the American Society for Information Science 35:6(1984):325.

정보검색에서 언어의 기능은 대략 다음과 같이 세가지로 요약해 볼 수있다.

첫째, 언어는 문헌이나 데이터 또는 정보를 수록하고 있는 매체의 내용을 표현한다. 둘째, 정보요구를 표현한다. 셋째, 언어는 컴퓨터로 하여금 정보검색기능을 수행토록 지시한다.<sup>9)</sup>

모든 언어는 그것이 통제된 어휘든 자연어이든간에 그 어휘와 구문론, 논리구조 및 영역으로써 각각의 특징을 갖는다.

어휘(Vocabulary)는 언어를 구성하는 요소이고, 구문론(Syntax)은 이러한 어휘들을 적합하게 배열하는 요소이며, 논리구조(Logical Structure)는 어휘간 관계를 규정한다.

영역(Domain)이란 실제로 각 개념들이 표현되어지는 형식을 말한다. 개념(Concept)은 압축된 idea로서 상징(Symbol)으로 표현되며 통제된 어휘든 자연어이든 어휘를 구성하는 요소는 상징으로 인식되어야 한다.

정보검색에 있어 중요한 점은 컴퓨터는 개념으로 인식하는 것이 아니라 상징으로 인식한다는 사실이다. 따라서 효율적인 정보검색이 되기 위해서는 축적된 정보를 나타내는 어휘에 관한 세심한 배려가 필요하게 되며<sup>10)</sup> 시스템의 어휘는 시스템의 성능에 중요한 영향을 미친다.

더구나 1980년대 이후 Technology의 발달과 함께 정보시스템이 전문기반의 데이터베이스 시대로 진행되면서 검색속도가 향상되고 접근점(Access Point)이 증가되는 등 외형적인 성능은 크게 개선되었다. 따라서 이제 시스템의 성능은 재현(Recall)의 측면에서 보다는 정확성(Precision)의 측면에서 연구되어야 할 것이다.

결국 정보시스템의 성능의 문제는 색인의 중요성<sup>11)</sup> 과 그 궤를 같이 하게 되고, 색인의 질이 정보시스템의 질을 결정하는 최종 요소가 된다

9) Stephen P. Harter, Online Information Retrieval : concepts, principles, and techniques, Orlando: Academic Press, 1986, pp.22-24.

10) Stephen P. Harter, op.,cit., pp.22-24.

Mirja Iivonen, "Consistency in Selection of Search Concepts and Search Terms" Information Processing & Management 31:2(1995):173.

S. Dutta and P. K. Sinha, "Pragmatic Approach to Subject Indexing : A New Concept." Journal of the American Society for Information Science 35:6(1984):325.

고 볼 수 있다.

대부분의 데이터베이스에서는 title, abstracts, identifiers 등의 탐색과 full-text에서의 자연어 탐색도 가능하다. 또한 데이터베이스에 따라서는 descriptors, subject headings, classification code도 활용할 수 있다. 그렇다면 어떤 선택 및 선택의 혼합이 특정탐색문제에 가장 적합한가? 각기 접근방법에 관련된 장점·단점은 무엇인가? 어떤 상태에서 natural language passage, descriptors, classification code 가 유용한가?

이러한 문제를 해결하고 효율적인 검색시스템을 설계하기 위해서 1950년대 이후 색인방법에 대한 많은 연구가 진행되어 왔는데 이 글에서는 크게 자연어 색인과 통제어 색인으로 나누어 그 원리를 고찰해 보고<sup>12)</sup>, 그 외에 이러한 방법이 갖고 있는 제한요소를 극복하고자 한 색인방법에 대하여 간략한 소개를 하고자 한다.

## 2. 2 자연어 색인과 통제어 색인

자연어 색인은 다음과 같은 문제점들이 있다.<sup>13)</sup>

1) 동의어 및 유사동의어의 존재가 on-line검색시스템의 어려움을 증가시킬 것이며, 그 이유는 컴퓨터가 정보검색을 할 때 symbol로 대조되기 때문이다. 따라서 탐색문이 개념에 관한 모든 관련 동의어를 포함하는데 실패한다면, 즉 탐색문에 사용된 단어나 구가 데이터베이스안의 개념을 기술하는데 어떤 단어나 구와 대조할 수 없다면, 검색결과는 오류정보가 많아지게 될 것이다.

11) D. Davidson, "The Effects of Individual Difference of Cognitive Style on Judgement of Document Relevance," *Journal of the American Society for Information Science* 25:1(1978):67-80.

12) 자연어 색인과 통제어 색인에 관한 대립적 연구에 대하여 초기의 연구에서부터 최근의 연구에 이르기 까지 개괄적 요약을 한 기사는 1) Elaine Svenonius 의 "Unanswered Questions in the Design of Controlled Vocabularies," *Journal of the American Society for Information Science* 37:5(1986):332-334. 와 2) J. Rowley, "The Controlled versus Natural Indexing Languages Debate revised: a Perspective on Information Retrieval Practice and Research," *Journal of the American Society for Information Science* 20:2(1994):108-119.

3) 정영미, "국내문자정보 데이터베이스의 색인에 관한 연구," *情報管理學會誌* 13:1(1996.6.):22-25.가 있다.

13) Stephen P. Harter, *Online Information Retrieval : concepts, principles, and techniques*, Orlando: Academic Press, 1986, pp.22-24.

2) 자연어에서는 주어진 개념을 표현하는데 많은 방법이 있을 수 있으며, 따라서 전문 데이터베이스를 탐색할 때, 주어진 주제를 다루는 가능한 한 많은 항목(items)을 찾기 위하여 탐색자는 관련 개념을 표현하는데 사용하고 있는 모든 가능한 단어와 구를 참여 시키도록 노력하여야 한다.

3) 동질적이고 (homogeneous), 고도로 전문화되고 (highly specialized), 기술적인(technical) 문헌에서는 주어진 단어나 구가 거의 동일한 개념으로 사용되거나 사회과학과 같은 복합적 주제분야에서는 자연어의 모호성은 심각한 문제가 되고 있으며, 동형이의어에 의해 야기되는 의미모호성의 문제점은 근본적으로 극복되어야 할 문제이다. 일반적으로 학문의 성격과 특성<sup>14)</sup>에 따라 "hard" 학문과 "soft" 학문으로 구분하고 있는데 "hard" 학문은 상대적으로 잘 정의된 구조 등을 갖고 있는 학문의 특성을 가질 때 사용되며, 반대 학문의 개념은 "soft" 학문으로 구분하고 있다. 학문분야간의 이러한 용어의 차이점은 시스템의 어휘 설계시 고려할 가치가 있다.<sup>15)</sup>

4) 자연어 색인은 동형이의어관계에서 야기되는 연관관계 또는 맥락상의 모호성(contextual ambiguity)의 문제점을 내포하며, 개념들 사이에 존재하는 계층적 관계를 다루는데 있어서도 무력하다. 통제어휘는 동의어, 유사동의어, 동형이의어 등을 통제함으로써 유사주제의 분산 및 유사하지 않은 주제의 배열을 방지하여 색인 및 탐색에 있어서 주제에 관하여 일관된 표현이 가능토록 하고, 포괄적인 탐색활동을 촉진시키기 위해 어의적으로 관련된 용어들을 함께 묶은 것이다.

14) 인문과학분야 연구자들은 검색어로서 별도의 주제어를 선택하기보다 연대순이라든가 지명 등을 선호하며, 자연과학분야 연구자들은 그 분야의 주제어 또는 일반 용어를 주제어로서 선택한다. Bates, M. J., Wilde, D. N., & Siegfried, S., "An Analysis of Search Terminology used by Humanities Scholars: The Getty Online Searching Project Report Number 1," Library Quarterly 63 (1993):1-39.

15) 몰토(Molto, Mavis)에 의하면 전문(Full-text) 데이터베이스가 증가되면서 그에 관한 연구도 증가하고 있으나 대부분이 언어적인 문제 등 자연어 색인의 보편적인 결함을 보완하는 방향에서 이루어지고 있다면서 학문주제별 Full-text를 심층적으로 분석하여 그 결과를 시스템 설계에 반영할 가치가 있음을 주장하였다. 왜냐하면 학문주제별로 특정한 용어가 특정되어 사용되고 있으며, 그 분야의 저자들이 공통적으로 사용하는 언어습관이 있으므로 이에 관한 연구 결과가 시스템에 내장될 때 시스템의 성능을 향상시킬 수 있을 것으로 보았다. Mavis Molto, "Improving Full Text Search Performance through Textual Analysis," Information Processing & Management 29:5(1993):615-616.

통제어휘는 정보검색과정의 커뮤니케이션을 촉진하기 위해 존재하나 모든 문제를 해결하지 못하며 또한 그 자신의 새로운 문제를 만들어 낼 수 있다. 통제어 시스템의 실패는 어휘의 특정성 결여, 용어간의 모호하고 허위적인 관계, 어휘의 부적절한 구조에 기인하는데 이러한 문제점을 해결하기 위해서는 다음과 같은사항이 반영되어야 한다.<sup>16)</sup>

- 1) 실질적이며 잠재적인 이용자의 정보요구 및 문헌의 용어로부터 추출한 '적합성(warrant)을 갖을 수 있어야 한다.
- 2) 대다수 탐색자의 탐색실험시 어느 정도 수준의 정확율을 보장할 수 있도록 충분히 전문성(specific)을 가질 수 있어야 한다.
- 3) 그릇된 조합 및 부정확한 용어관계에서 비롯되는 대부분의 문제점을 피하기 위하여 전조합(pre-coordinate)을 하여야 한다.
- 4) 동의어, 유사동의어 등의 통제를 통해서 색인과 탐색에 일관성을 증진시킬 수 있어야 한다.
- 5) 동형이의어의 분리나 용어가 불명확한 것은 그 의미나 범위를 지정함으로써 모호성을 줄일 수 있어야 한다.
- 6) 계층적, 상호참조적인 구조를 통하여 특정 주제 표현에 가장 적합한 용어를 선정할 수 있도록 하여야 한다.

자연어 색인시스템은 모든 정보를 검색에 대비하여 방대하게 조직할 수 있으며, 검색에서는 이용자가 어떤 특정의 검색용어에 익숙하지 않더라도 쉽게 적응하여 검색이 가능한 장점을 갖고 있고, 또한 전문 데이터베이스의 증가추세로 자연어 색인을 이용한 데이터베이스가 늘어날 전망이다. 통제어 시스템은 고유의 장점에도 불구하고 자연어시스템이 갖는 경제적 효과, 특정성 측면 등에서 효과가 뒤지는 것으로 연구되었다.<sup>17)</sup>

그렇다면 통제어 시스템은 모든 검색시스템에서 제외시켜야 할 것인가? 자연어 시스템은 완전한가? 독특한 지식 베이스를 갖는 법률정보

16) 이에 관하여 자세히 기술한 저서는 F. Wilfred Lancaster, Vocabulary Control for Information Retrieval, Arlington:Information Resources Press,1986.

17) Full-text데이터베이스의 증가는 자연어 색인의 증가를 가져왔으며 그에 따라 자연어색인의 결점을 보완하는 많은 연구가 있어 왔다. 대표적으로 Jung Soon Ro, An Evaluation of the Applicability of Lanking Algorithms Improving the Effectiveness of Full-text Retrieval, Indiana University:1985. 의 연구가 있으며 Carol Tenorpir & Jung Soon Ro, Full-text Databases, Westport: Greenwood Publishing Group, Inc., 1990, 제4장에서 자연어검색과 통제어검색의 검색성능을 다룬 선행연구에 대하여 자세히 기술하고 있다.

시스템에서도 그러할 것인가?

### 2. 3 여러가지 색인방법

이보넨(Iivonen, M.)<sup>18)</sup>은 탐색개념은 탐색어로 표현되므로 정보검색은 탐색어와 매우 밀접한 관련이 있다고 보고, 탐색개념이 탐색어로 표현되어지는 관계에 관한 연구를 수행하였다. 네개의 서로 다른 시스템 - 피험자마다 각기 서로 다른 탐색어를 쓰는 환경, 동일한 탐색개념을 표현하기 위하여 여러가지 탐색어를 사용토록 한 환경, 탐색질의 계층관계를 통하여 통제된 환경, 여러가지 탐색어가 동일한 탐색개념을 표현하도록 탐색개념을 포괄적으로 준 환경 - 에서 32명의 피험자를 대상으로 12개의 탐색질문으로 시험을 실시한 결과 특정성이 높은 용어를 사용하는 전문직종은 탐색어로서 디스크립터를 사용하면서 탐색자 간에 용어의 일치성이 높은 반면 공공도서관 등에서는 자연어를 탐색어로 선택하면서 탐색자간 용어의 일치율이 낮다는 사실과, 가장 최선의 탐색표현을 찾으려 하기 보다는 탐색개념에 대한 다양한 표현장치(시소러스, 질의어 확장)를 마련하는 것이 탐색효율을 높일 수 있다는 사실을 발견하였다. 그러나 이 경우에도 재현의 문제는 해결방안이 될 수 있으나 정확성의 문제는 여전히 남는다.

번스타인과 윌리엄슨(Lionel M. Bernstein and Robert E. Williamson)<sup>19)</sup>의 연구결과에 의하면 의학분야 전문 데이터베이스 검색시 언어적, 확률적 방법 외에 의학분야 문헌이 갖고 있는 고유의 지식을 내장시킴으로써 검색효율을 증진시킬 수 있다는 것이다.

의학분야 시스템에서는 문헌의 처음에 나타나는 수 개의 문장 내에 정보요구를 해결할 단서(answer)가 나타나는 경우가 85 ~ 90%임을 밝힘으로써 본문에 대한 전반적 검색이나 정보질의에 대한 재분석 없이도 이 결과를 지식베이스로 하여 자동 색인이 가능하다고 보았다. 베루(Catherine Berrut)<sup>20)</sup>는 학문주제별 지식이 색인에 활용될 수 있을 것으로 보았다. 의학분야의 경우 특정한 용어가 특정한 법칙 - 임상관측

18) Mirja Iivonen, "Consistency in Selection of Search Concepts and Search Terms," *Information Processing & Management* 31:2(1995):173-190.

19) Lionel M. Bernstein and Robert E. Williamson, "Testing of a Natural Language Retrieval System for a Full Text Knowledge Base," *Journal of the American Society for Information Science* 35:4(1984):235-247.



→ 진단→증상 등 - 에 의하여 기술되는 사실을 밝히고 이러한 요소를 문헌의 내용을 표현하는 지식베이스로 한다면 보편적인 자연어 색인에서 발생할 수 있는 언어의 문제에서 자유로울 수 있을 것으로 보았다.

샬리니(R. Shalini)<sup>21)</sup>는 문헌내에 인용된 자료는 저자가 “주제로서 나타내고자 한 개념”과 관련있다고 판단한 참고자료이므로 이용자는 인용문헌프로파일에 대하여 자신의 탐색개념과 심리적으로 적합하다고 느낀다는 사실에 근거하여, 인용문헌프로파일이 실제로 적합률을 향상시키는데 대한 실험을 수행하였다. 실험은 두 단계로 나뉘어 수행되었는데 첫번째 단계는 서지시스템내에서 서지정보, 디스크립터 등으로 주제개념을 검색토록 한 시스템과 문헌내에 인용된 문헌의 서지정보, 초록 등을 이용하여 주제개념을 검색토록 한 시스템 두 가지를 제작한 다음 두번째 단계에서는 이용자에게 전문(full-text)를 주고 첫번째 단계의 활동이 탐색하고자 한 주제개념에 적합한지 여부를 측정하였다. 그 결과 인용문헌 프로파일은 색인어의 기능을 수행하면서 적합판단자료로 적용될 수 있음을 밝혔다.

파오와 워텐(Miranda Lee Pao & Dennis B. Worhten)<sup>22)</sup>의 공동연구는 동일한 데이터베이스 내에서 주제어라고 할 수 있는 디스크립터를 이용한 검색시스템과 문헌내 인용자료를 이용한 검색시스템의 검색효율을 비교하고 데이터베이스의 검색효율을 증진하는 방안에 관하여 연구한 것이다. 연구결과 문헌내 인용자료를 이용한 검색과 디스크립터를 사용한 검색은 서로 다른 결과의 문헌을 검색해내는 것을 발견하고 양자의 통합 검색이 재현율을 증진시킬 수 있다고 주장하였다.

보이스와 맥레인(Bert R. Boyce and P. MaLain)<sup>23)</sup>은 통제어를 이용하여 정확률은 유지하면서도 재현률을 높이는 방안에 대하여 연구하였다. 이와는 달리 자연어색인을 이용하여 재현율은 그대로 유지하되

20) Catherine Berrut, "Indexing Medical Reports: The RIME Approach," *Information Processing & Management* 26:1(1990):93-109.

21) R. Shalini, "Citation Profiles to Improve Relevance in a Two-stage Retrieval System: a proposal," *Information Processing & Management* 29:4(1993):463-470.

22) Miranda Lee Pao & Dennis B. Worhten, "Retrieval Effectiveness by Semantic and Citation Searching" *Journal of the American Society for Information Science* 40:4(1989):226-235.

23) Bert R. Boyce and P. MaLain, "Entry Point Depth and Online Search Using a Controlled Vocabulary" *Journal of the American Society for Information Science* 40:4(1989):273-276.

정확률을 높이는 방안으로서 문헌내 높은 빈도를 차지하는 용어를 제거하는 등의 방안도 고려될 수 있으나 이 경우에는 낮은 빈도를 보이면서 색인어로서 가치가 없는 용어는 제거할 수 없는 단점이 있다. 따라서 이들의 연구는 통제어가 갖고 있는 장점은 살리면서 디스크립터 수를 늘려 자연어색인이 갖을수 있는 접근점의 증가 효과를 가져오도록 함으로써 결국 재현율도 높일 수 있다고 본 것이다. 그 방법은 본래 문헌에 할당된 표준 디스크립터 외에 문헌이 인용하고 있는 인용자료의 디스크립터를 색인어로 추가하는 것이다.

### 3. 법률정보시스템의 특성

각 문헌에서 어느 법률문제를 어느 법규로써 다루고 있는지 정확히 표현(Precise content Representation)될 수 있다면 이용자는 일일이 전문을 읽어야 하는 노력을 대폭 절감할 수 있을 것이며 이것은 곧 저자가 정보 생산시 의도한 개념(Producer's Viewpoint)이 왜곡되거나 누락없이 표현되는 이상적인 상태를 말한다.<sup>24)</sup>

법률정보시스템에 있어서도 색인의 중요성은 이미 입증되었으며 미국 의회법률도서관(The Law Library of Congress)조차도 색인이나 초록을 갖춘 수많은 데이터베이스와 장서를 보유하고 또한 특별예산 등을 책정하여 효율적인 정보검색을 위한 색인, 분류, 자동화 등에 노력하고 있으나 만족할 만한 결과는 얻지 못하고 있는 상태이다.<sup>25)</sup>

국내에서도 대규모 법률도서관을 중심으로 2차법률정보의 전문(full-text) 데이터베이스를 전제로 한 여러가지 색인방안이 제시되고 있다. 전문데이터베이스의 경우 오히려 접근점의 과다로 인한 비효율을 초래하므로 저자로 하여금 초록을 작성케 하여 제목과 초록을 대상으로 색인을 하는 방법, 소제목을 입력하여 소제목을 대상으로 색인을 하는 방법 등을 고안하였으나 그 어떤 방법도 효율성을 보장할 수 없다. 왜냐하면 모든 문헌의 저자들로 하여금 일관성 있게 초록을 작성토록 강제할 수 없을 뿐 만 아니라 일정시점을 기준으로 그 이후 발생된 문헌에 대하여만 그러한 방안을 강제한다고 하더라도 이미 생산, 축적된 소급

24) Robert Fugmann, "The Complementarity of Natural and Indexing Languages," *International Classification* 9(1982): 140-144.

25) Kathleen Price, "Building a Law Library to Serve the World Legal Community," *Law Library Journal* 83(1991):225-233.

자료와 같은 형식으로 연결할 방법은 불가능에 가깝다. 또한 소제목을 입력하는 방안이 있어서도 역시 소급자료에 있어서는 동일한 방식을 적용할 수 없으므로 이러한 색인방법은 일정 시점을 기준으로 이전의 지식을 사장시키던가 또는 막대한 비용과 시간을 투입하여 동일한 형식으로 재현시키지 않는 한 수용하기 어려울 것으로 보인다.

2차 법률정보 데이터베이스에 있어서 표제상의 용어나 일반 자연언어만으로 접근하는 것이 매우 불안정하여 원하는 정보에의 정확한 접근이 어려움에도 불구하고 표제, 저자, 출판사 등의 기본적 서지사항을 이용한 검색수준에 머물고 있는 이유는 무엇인가? 그 이유는 법률정보의 중요성이 다른 부문보다도 더 크게 인정되고 있으면서도 오히려 뒤늦게 개발된 점이나, 일반 주제부문의 정보시스템의 경우 일반 엔지니어에 의해서 비교적 용이하게 개발될 수 있었던 것과 비교하여 법률정보시스템은 일반 엔지니어들이 오랜동안 개발을 시도하였음에도 불구하고 결국 판사 등의 주제전문가에 의하여 개발될 수 밖에 없었던 사실과 무관하지 않다.

즉, '법률'이라고 하는 주제의 난해함이 개발 저해 요인이 될 수 있으며, 이러한 '난해함'은 엄밀한 의미에서 곧 법률정보의 특성으로 표현될 수 있다. 그러나 역설적으로 이러한 법률정보의 고유한 특성은 법률정보시스템 설계시 지식 베이스로 적용될 수 있을 것이다.

### 3. 1 법률정보의 특성

법률이란 모든 자연인(또는 법인)의 법률행위를 규율하기 위한 것으로 그 외적 표현형식은 일종의 코드이자 시소러스 또는 용어집 기능을 수행한다. 임의의 법률을 예를 들어 그 구성체계를 보면,

#### (법률구성 예) 주택임대차보호법 제3조

제3조(대항력등) ①임대차는 그 등기가 없는 경우에도 임차인이 주택의 인도와 주민등록을 마친 때에는 그 익일부터 제3자에 대하여 효력이 생긴다. 이 경우 전입신고를 한 때에 주민등록이 된 것으로 본다.

②임차주택의 양수인(기타 임대차 권리를 승계한 자를 포함한다)은 임차인의 지위를 승계한 것으로 본다.

③민법 제575조 제1항·제3항 및 제578조의 규정은 이 법에 의하여 임대차의 목적이 된 주택이 매매 또는 경매의 목적물이 된 경우에 이를 준용한다.

④민법 제536조의 규정은 제3항의 경우에 이를 준용한다.

여기에서 제3조로 표기된 형식은 코드의 기능을 수행하며, 괄호안의 용어인 “(대항력등)”은 코드의 값 또는 정의를 나타내며, ①,②,③,④항에서 이어지는 문장은 본 법률 해당 條의 정의를 구체적으로 설명하고 있다.

판례란 발생사건에 대하여 법률적 적용을 가한 것으로 역시 정형화된 형식으로 나타난다. 법률과 마찬가지로 예를 들어 그 표현 형식을 살펴보면,

(판례구성 예)

1997. 2. 28. 선고 96다46033,46040 판결(임대료·임대보증금)

【판결요지】주택임대차보호법 제3조 제1항에서 주택의 인도와 더불어 대항력의 요건으로 규정하고 있는 주민등록은 거래의 안정을 위하여 임차권의 존재를 제3자가 명백히 인식할 수 있는 공시방법으로 마련된 것이므로, 주민등록이 어떤 임대차를 공시하는 효력이 있는지의 여부는 일반 사회통념상 ... (중략)

【참조조문】주택임대차보호법 제3조 제1항

【참조판례】대법원 1994.11.22.선고 94다13176판결(공1995, 64), ...

(중략)

【원고(반소피고), 피상고인】(생략)

【피고(반소원고), 상고인】(생략)

【원심판결】서울지법 1996.9.25.선고 96나31016, 31023

【주 문】원심판결을 파기하고 사건을 서울지방법원 본원 합의부에 환송한다.

【이 유】상고이유를 판단한다.

(중간생략)

이와 같이 모든 판례는 선고년월일, 사건번호, 참조조문 등 정형화 된

형식으로 표현되는데, 법령을 모든 법률정보를 파생시키는 가장 기초적 법률정보라 정의한다면 판례란 이러한 법령을 개별 사건에서 구체적으로 해석하고 적용시킨 것이다. 이들을 제외한 2차 법률정보란 법률이나 판례에 관하여 해석, 평석 등을 가한 연구결과물로서 1차법률정보와 사용자회 내지 근거범위를 같이 하게 된다.

(법률문헌 구성 예)

제목 : 대항력 있는 임차인의 배당요구  
저자 : 서울고등법원 판사 겸 법원행정처 사법대책연구심의관 장성원  
출전 : 법조 1997. 7.

一. 서

주택임대차보호법 제3조의 2 제1항은 임대차계약서에 확정일자를 갖춘 임차인의 임차보증금에 관하여 우선변제권을 인정하고 있고, 법 제8조 제1항은 소액임차인의 임대차보증금에 관하여 최우선변제권을 인정하고 있다. 임대차보증금반환청구권은 담보물권과 달리 등기부에 의하여 공시되지 않으며, 채무명의를 얻지 않고서도 강제집행절차에서 우선변제권을 주장할 수 있다는 점에서 일종의 우선특권에 해당된다고 설명된다. (중략)

三. 임대차 종료전의 대항력 있는 임차인의 우선변제권 행사(중략)

3. 판례

가. 종전의 판례

· · · · · 다만, 대법원 1995. 5.26. 선고 95다132258 판결 및 1996.

6.25. 선고 96다12474 판결은주택임대차보호법 제4조 제1항은 임차인을 보호하기 위한 규정으로서, 기간을 2년미만으로 정한 때에는 . . .

(중략)

4. 사건

(중략)

민사소송법 제603조의2제1항에 의하면 법원이 경매개시결정후 집행관에 게 하는 현황조사명령의 내용에는 부동산의 점유관계, 임차내용이 포함되어 있고, 민사소송법 제617조의2에 의하면 . . . . . 그 중대한 하자는 경락불허가 사유가 된다(민사소송법 제633조 제6호). 나아가 대항력 있는 임대차의 존부 및 내용은 임대차공고에 기재되어야 하고 그 기재가 누

락되거나 잘못되면 원칙적으로 경락불허가사유가 된다. (민사소송법 제633조 제5호)  
(중략)

이 글의 저자가 다루고자 한 최상위 개념(topicality)은 대항력을 갖춘 입차인인 경우 목적 주택이 경매될 때 입차기간 종료전이라도 우선변제를 요구할 수 있는지, 아니면 우선변제권을 포기하고 경매법원에 대하여 배당요구를 해야 하는지, 또는 그 효과는 무엇인지에 대한 것이다. 저자는 자신이 말하고자 하는 다수의 개념을 주택임대차보호법 제3조·제4조·제8조, 민사소송법 제603조의 2, 제617조의 2·제633조 등과 같은 법률조문과 이미 생성, 확립된 판례 등을 인용하여 근거를 제시하고 설명하고 있다.

이와 같은 법률문헌의 예를 보면 제목필드만으로는 저자의 의도가 포용되지 않는다는 사실을 곧 알 수 있다. 따라서 2차 법률정보의 색인어를 추출하고자 한다면 내용 전반에 관한 검토가 필요하고 그러한 작업은 법률전문가가 아니면 결정할 수 없을 것이다. 그러나 여기에 주목할 만한 점은, 위의 예에서 알 수 있는 바와 같이 2차 법률정보의 내용에는 2차 법률정보가 다루고자 하는 주제와 관련된 여러 법령 또는 판례가 나타나게 된다는 사실이다. 2차 법률정보는 1차 법률정보인 법령, 판례를 떠나서는 존재할 수 없으며, 결국 2차 법률정보는 법령, 판례와 사용어휘 또는 근거범위를 같이 하는 속성을 갖게 되는 것이다.

법령, 판례, 법률문헌의 구성체계를 통해서 추론한 법률정보의 특성은 다음과 같이 요약될 수 있다. 즉, 1차 법률정보는 2차 법률정보를 통제하는 기능을 갖고, 2차 법률정보는 1차 법률정보를 인용하는 형식을 갖는 주제적 특성을 지니고 있으며 2차 법률정보에 인용된 1차 법률정보는 저자의 저술의도를 명확히 표현하는 개념어가 된다는 사실이다. 따라서 이와 같은 고유한 특성을 법률정보 시스템의 지식베이스로 할 수 있을 것이라는 추론이 가능하다.

결국 법률정보란 1차 법률정보가 2차 법률정보를 통제하는 기능을 갖는 주제적 특성을 지니고 있으며 이와 같은 고유한 특성을 법률정보 시스템의 지식베이스로 할 수 있을 것이라는 추론이 가능하다.

### 3. 2 법률정보의 특성을 활용한 색인방안

자연어나 통제어의 속성이 정보시스템에서 의미하는 것은 분명히 그 장단점을 구별할 수 없다. 즉 어떠한 접근방법을 택하였다고 하더라도 그것을 잘못 사용했다고 비난할 수 없다.<sup>26)</sup>

현재 이들 연구의 결과 통제어와 자연어의 혼합형태가 가장 효과적이라는 결론에 이르고 있는데, 그렇다면 모든 시스템에 획일적으로 이 결론이 적용될 수 있을 것인가?

자동색인의 방법은 학문주제에 따라 적합성결과가 달라지며,<sup>27)</sup> 이용자의 유형에 따라 검색전략이 달라질 수도 있다.<sup>28)</sup>

법률정보에 있어서는 어떠한가?

법률정보의 용어는 상용어가 아니므로 주제배경 없이는 거의 접근이 불가능하고, 기사 또는 문헌의 표제(title)가 표현하는 개념 외에도 여러 관점에서 논하는 여러 개념들이 그 기사 또는 문헌의 내용의 많은 부분을 차지하는 경우가 많고, 예를 들어 민법상의 '시효'는 형법에서도, 기타 법령에서도 사용되는 등 용어의 범주화가 용이하지 않는 점 등 색인에 있어서 많은 어려움이 예상될 수 있다.

그러나 2차 법률정보 데이터베이스에 있어서는 다음과 같은 이유에서 자연어 색인시스템보다는 통제어 색인시스템을, 통제어 시스템 중에서도 인용색인시스템의 구현을 제안할 수 있다.

먼저 통제어 시스템을 제안할 수 있는 이유로는,

1) 2차 법률정보 표현에 사용되는 언어는 1차 법률정보 표현에 사용된 용어, 조문형식을 그대로 사용하므로 이러한 형식이 그대로 일종의 통제어로서 사용될 수 있다. (이러한 사실을 용어의 특정성(specificity)으로 표현한다면 오히려 자연어 시스템이 갖고 있는 속성이기

26) 자연어 시스템이든 통제어 시스템이든 어느 한쪽이 다른 한쪽보다 더 우월하다고는 말할 수 없다. 즉 어느 특정한 목적의 탐색에서는 자연어 탐색이, 또 다른 목적의 탐색에서는 통제어 탐색이 적절할 것이며, 궁극적으로 특정분야의 user가 가장 잘 안다.

Raya Fidel, "Towards Expert Systems for the Selection of Search Keys," *Journal of the American Society for Information Science* 37:1(1986):43-44.

27) Mary E. Rowbottom & Peter Willett, "The Effect of Subject Matter on Automatic Indexing of Full-text," *Journal of the American Society for Information Science* (1982):139-141.

28) 정보 검색시 검색경험이 적은 초보자는 숙련자에 비하여 덜 통제된 색인어를 검색어로 선택하면서 그 수도 많지 않은 경향을 보인다. Hsieh-Yee, "Effect of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers," *Journal of the American Society for Information Science* 44:3(1993):161-174.

도 하나, 문헌에 나타난 임의의 키워드가 아니라 1차 법률정보라고 하는 일종의 사전에서 정의된 용어라는 점에서 일반적 자연어 시스템으로 보지 않는다.)

2) 2차 법률정보 표현에 사용된 언어는 이미 1차 법률정보에서 정의된 정제된 용어로서 자연어 처리에 필요한 구문적 문제가, 해결될 수 있다.

3) 용어의 확장이 필요한 경우 보조장치로서 그대로 시소러스를 용할 수 있다.

인용색인시스템을 제안할 수 있는 이유는

1) 2차 법률정보는 그 속성상 법률정보의 해설 또는 주석이므로 그 표현은 반드시 1차 법률정보를 인용하는 형식을 갖게 된다.

2) 인용된 1차 법률정보는 2차 법률정보가 표현하고자 하는 개념(concept)과 관련이 있다.

#### 4. 결 론

국내외 법률정보시스템 연구자들은 2차 법률정보 데이터베이스의 색인방법을 채택하는데 있어서 법률주제가 갖고 있는 고도의 전문성과 기존 자연어 방식, 통제어 방식 각각의 결점으로 인하여 바람직한 대안을 찾지 못하고 있는 실정이다.

본 연구는 그 대안을 모색하기 위한 기초 연구로서 2단계로 나누어 진행되었다. 첫째, 지금까지 이룩된 자동 색인에 관한 선행연구 중 색인어의 기본 원리를 정리하고, 학문주제별 또는 기타 특성에 따라 색인방법이 달라질 수 있는 가능성을 제시한 저작에 대하여 분석하여 그 결과를 법률정보 특성에 적용가능한지 검토하였다. 둘째, 법률정보의 특성을 연구한 결과 2차 법률정보 데이터베이스의 색인방법으로는 인용색인방식과 통제어 색인방식을 채택할 수 있다는 근거를 제시하였다. 즉, 법률정보만의 고유한 학문적 특성을 기반으로 한 색인시스템을 제안할 수 있으며, 그 결과 다음과 같은 4가지의 기대효과를 가져올 수 있을 것으로 보인다.

첫째, 색인어 선정시 법률주제 전문가의 지적 노력을 필요로 하지 않으므로 그에 수반되는 비용, 시간의 투입을 절약할 수 있으며, 주관의 개입을 배제하고 일관성을 유지할 수 있다.



둘째, 소급자료와의 통합 데이터베이스화가 가능하다.

셋째, 법률의 빈번한 제·개정시에도 별도 장치의 개발없이 데이터베이스내 모든 색인어의 일괄 개정이 용이하다.

넷째, 보조 장치로서 시소라스를 활용했을 경우 시소라스의 가치를 극대화시킬 수 있다. 왜냐하면 시소라스란 본질적 색인어가 아닌 검색의 보조장치로서, 문헌의 내용을 정확히 표현하는 개념어의 추출 없이 문헌에 출현하는 모든 단어에 대하여 시소라스를 적용하는 것은 법률 시소라스의 개발·유지보수에 투입되는 비용 대비 효과를 거둘 수 없으며, 적합한 수 개의 색인어가 추출된 후 그 용어에 대하여 시소라스를 적용할 때에야 시소라스의 적용 효과를 거둘 수 있기 때문이다.

요컨대 정보시스템 설계시 정보전문가와 주제전문가가 인식하고 있는 각 주제 분야의 지식을 내장시키는 것이 필요하며, 그것이 더욱 본질적인 지능형 정보시스템이 될 수 있다는 가능성을 모색하려는 것이다.

## ABSTRACT

### A Study on the Index Model for Secondary Legal Information Databases

Roh, Jeong-Ran

This study proves that the quoted legal text functions as the index which represents the contents of the text because of the characteristics of legal information, the automatic indexing in the secondary legal full-text databases can be possible without the assistance of the experts. In case of the establishment, amendment or repealing of law, change of words of index can be possible through revising the legal text quoted in the secondary legal full-text databases. Even when we don't input the full-text about retrospective documents, automatic indexing is also possible, and the establishment and the practice of expert knowledge and integrated databases are possible in case of the retrospective documents.

This study indicates that it is necessary to have characteristic information the information experts recognize - that is to say, experimental and inherent knowledge only human being can have - built-in into the system rather than to approach the information system by the linguistic, statistic or structuralistic way, and it can be more essential and intelligent information system.