

2차 法律情報 全文데이터베이스 構築을 위한 基礎 研究

A Primary Study on Building the Secondary Legal Information Full-Text Databases

權 起 遠 (Kie-Won Kweon)*
盧 貞 蘭 (Jeong-Ran Roh)**

목 차

- | | |
|---------|-------------|
| 1. 서 론 | 3. 법률정보의 특성 |
| 2. 선행연구 | 4. 결 론 |

초 록

최근 주제별 데이터베이스의 구축이 활발해지고 있다. 현재까지 이루어진 연구결과를 보면 자연어 색인 시스템의 검색효율이 통제어 색인시스템의 검색효율보다 일반적으로 높은 것으로 나타나고 있어 각 주제 부문에서는 데이터베이스의 구축시 이러한 결과를 수용하여 자연어 색인시스템을 채택하는 경향이 높다. 본 연구는 專門데이터베이스로는 그 수요가 매우 높은 2차 법률정보 全文데이터베이스를 대상으로 법률정보의 내재적 특성에 근거한 통제어 색인시스템의 이론적 가능성을 제안하려는 것이다. 본 연구를 통해 색인어의 자동적 추출 가능성이 추론된다면, 그 결과는 2차 법률정보시스템의 설계자들에게 주제배경이 없이도 특정의 원리에 의하여 자동색인을 가능케 하고, 다른 주제분야의 정보시스템 설계자들에게는 해당 주제분야의 고유한 지식베이스를 활용하는데 있어서의 시사점을 제공하게 될 것이다.

ABSTRACT

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.

As this study proves that the cited primary legal information within the secondary legal information 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 index terms can be possible through revising the legal text cited in the secondary legal information 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.

* 成均館大學校 文獻情報學科 教授

** 成業公社 法規室 課長(성균관대학교, 동덕여자대학교 강사)

접수일자 1998년 9월 5일

1. 서론

1.1 연구의 필요성

최근 개인이나 기업의 활동영역이 확대되고 전문화, 국제화되면서 법률분쟁이 증가할 뿐 아니라 이를 규정하는 법령의 수가 증가하고 복잡해짐에 따라 법률정보의 수요가 크게 늘고 있다. 이러한 추세에 부응하여 미국을 비롯한 서구 산업국가에서는 일찌기 법령과 판례 등의 1차 법률정보에 관한 데이터베이스를 구축하여 이미 상용화 수준에 이르렀다. 국내의 경우도 1987년 법체처와 총무처에 의해 “대한민국 법령 데이터베이스”가 구축되어 사용되고 있으며, 판례 데이터베이스 역시 1990년 처음으로 개발된 이래 상용되고 있다.

1차 법률정보¹⁾ 데이터베이스는 이용자의 접근이 비교적 용이하다. 그 이유는 법령이란 법률조문에 대한 정의의 형태를 띄고 있으며, 판례 역시 법령의 법률조문을 그대로 사례에 적용하여 사건번호에 선고연월일, 사건명 등을 적시한 것이므로 이에 관한 데이터 요소의 추출 및 검색이 비교적 평이한 방법으로 가능하며 또 용어자체의 난해성은 시소러스를 통해 제어할 수 있기 때문이다.

그러나 1차 법률정보와는 달리 2차 법률정보²⁾는 데이터 요소의 추출 및 검색어의 구성에 있어서 보다 복잡한 문제점을 수반하는데 그 문제점은 크게 두가지로 분석해 볼 수 있다.

첫째, 2차 법률정보의 표제에 사용되는 용어는 법학상의 여러 분야와 여러 법률조문과 관련되어 있는 경우가 대부분이어서 표제어를 하나의 특정 법률분야로 한정하는 것이 어렵고, 그 결과 표제어만으로 본문을 포괄하여 표현하기 어렵다는 사실이다. 또한 본문에 사용되고 있는 용어들은 그 자체가 전문성을 갖고 있는 용어이지만 법률정보 내에서는 특정한 전문용어로 볼 수 없는 일반 자연어이기 때문에 본문의 내용을 대표하는 색인어의 추출없이 본문에 나타난 모든 자연어에 시소러스를 적용하여 용어를 확대 또는 축소하는 경우에는 정보시스템의 성능 향상을 기대하기 어렵다.(Green, Rebecca and Bean, Carol A., 1995, Schuegraf, E. J. and Bommel M. F., 1993, Leung, Chi-Hong and Kan, Wing-Kay, 1997)

둘째, 정보시스템의 전문(full-text) 데이터베이스화 경향에 따라 2차 법률정보도 향후 전문데이터베이스로 구축되어 갈 것이므로 2차 법률정보 데이터베이스의 설계시 전문 데이터베이스가 갖는 취약점을 해결하여야 한다는 사실이다. 전문 데이터베이스의 경우 접근점의 과다는 일반적으로 재현율은 높일 수 있으나 정확률을 낮춘다는 사실이 여러 선행연구에서 밝혀지고 있어 향후 전문 데이터베이스 시대의 과제는 정확률을 향상시키는 것이 될 것이다. 이와 같은 사실은 전문 데이터베이스 중심의 정보시스템 설계시 핵심은 곧 정보의 내용을 정확하고 단순하게 재현할 수 있는 색인에 있다는 것을 의미하

1) 1차 법률정보란 원문 그대로의 법률(시행령, 시행규칙 포함), 예규, 판례 등을 말한다.

2) 2차 법률정보란 1차 법률정보의 이해를 돕기 위해 마련한 법해석서, 주석서, 일반 교재류, 저널기사, 논문 등을 말한다.

는 것이다.(정영미, 1996) 그러므로 2차 법률 정보 전문데이터베이스의 설계시에도 가장 핵심을 이루게 될 부분은 이러한 문제를 해결할 색인에 관한 사항으로서(Nazareth, A. M., 1990) 다수의 임의어를 이용함으로써 낮은 정확률을 감수하고 재현률을 높이는 색인방법보다는 주제 개념을 정확하게 표현하여 정확률을 높일 수 있는 장치를 필요로 한다.

이와 관련하여 법률정보가 다루고 있는 여러 난해한 개념들이 언어학적·확률분포적·문헌구조적인 방법 등 일반적인 자연어 색인기법으로 추출가능할 것인지, 이러한 방법으로 추출된 색인어들이 적합한 탐색결과를 나타낼 것인지에 대하여는 회의적 견해가 많이 발표되었다.(Green, Rebecca, 1995, Spink, Amanda and Saracevic, Tefko, 1997)

스메아톤(Smeaton, Alan F., 1990)에 의하면 벡터(vector) 또는 확률분포를 이용한 수학적 모델은 자연어 색인방법이 갖고 있는 많은 문제를 해결하였으나 아직도 상당부분은 미해결의 상태로서 이와 같은 방법만으로는 한계에 이르렀다고 보고 있다. 실제로 자연어 방식을 이용한 법률정보시스템의 이용자들은 검색결과에 대하여 강한 불만을 나타내고 있으며 오히려 주제 전문가가 할당할 통제 색인어 또는 그와 유사한 방법의 색인 방법을 제안하고 있다.(문용호, 1997)

법률주제부문의 세계 최대 법률도서관인 미국 의회법률 도서관에서도 많은 데이터베이스와 장서를 보유하고 있으며 또한 특별예산 등을 책정하여 이상적이고 효율적인 시스

템을 만들기 위한 색인, 초록, 분류, 자동화 등에 노력하고 있으나 만족할 만한 결과는 얻지 못하고 있다.(Price, Kathleen, 1991)

국내의 경우 역시 지금까지 이러한 문제점에 관하여 많은 논의가 있었으며 최근에는 대규모 법률도서관을 중심으로 2차 법률정보의 전문 데이터베이스를 전제로 한 여러가지 색인방안이 제시되고 있다. 저자로 하여금 초록을 작성케 하여 제목과 초록을 대상으로 색인을 하는 방법, 소제목을 입력하여 소제목을 대상으로 색인을 하는 방법 등이 현재까지는 최선의 색인방안으로 채택되고 있으나(문용호, 1997) 그 어떤 방법도 효율성을 보장할 수 없다. 그 이유는 모든 문헌의 저자들로 하여금 초록을 작성토록 강제할 수 없을 뿐만 아니라 일정시점을 기준으로 그 이후 발생한 문헌에 대하여만 그러한 방안을 강제한다고 하더라도 이미 생산, 축적된 소급자료와 같은 형식으로 연결할 방법은 불가능에 가깝다. 또한 소제목을 입력하는 방안에 있어서도 역시 소급자료에 있어서는 동일한 방식을 적용할 수 없으므로 이러한 색인 방법은 일정 시점을 기준으로 그 이전에 생산된 법률정보를 사장시키던가 또는 막대한 비용과 시간을 투입하여 동일한 형식으로 재현시키지 않는 한 수용하기 어려울 것으로 보인다.

이와 같은 현상은 2차 법률정보의 중요성이 크게 강조되어 가고 있음에도 불구하고 법률정보가 갖고 있는 여러가지 난해한 문제점 때문에 결국 현행 2차 법률정보 데이터베이스가 표제, 저자, 출판사 등의 기본적 서지사항을 탐색어로 하는 서지 데이터베이스이

거나 일반 자연어 색인기법을 그대로 이용하는 정체성을 빚게 되었음을 반증하는 것이다.

그러므로 2차 법률정보 전문 데이터베이스에서 법률정보가 갖는 주제의 난해함과 용어상의 문제점을 해결하고, 저자가 나타내고자 한 개념어를 정확히 추출하는 방안의 모색이 절실하다. 궁극적으로 검색의 효율성을 보장하기 위해서는 기존의 일반적인 자연어 색인기법이 아닌 법률정보의 고유한 특성을 주제베이스로 새로운 측면에서의 통제어 색인에 관한 연구가 필요하다.

1. 2 연구의 목적

본 연구의 목적은 법률정보의 특성을 연구하여 그 고유한 특성을 주제베이스로 한 2차 법률정보 데이터베이스의 색인 방안을 제안하는데 있다.

각 법률문헌이 어느 법률문제를 어느 법규로 다루고 있는지 정확히 색인어로 표현될 수 있다면 이용자는 일일이 전문을 읽어야 하는 노력을 대폭 절감할 수 있을 것이며, 이럴 경우 저자가 정보 생산시 의도한 개념이 정보시스템에 의하여 왜곡되거나 또는 누락 없이 표현된 이상적인 상태가 되었다고 할 수 있다.(Fugmann, Robert, 1982) 따라서 본 연구는 향후 2차 법률정보 데이터베이스가 전문으로 구축되어 갈 것을 전제로 개선된 색인방안을 모색하려는 것이다.

구체적으로 법률정보가 지니고 있는 특성을 기초로 다음과 같은 가능성의 근거를 제시하려는 것이다.

가. 법률정보가 지니고 있는 특성을 기초

로 한 주제의 자동 추출 즉, 자동 색인이 가능할 것인가?

나. 자동적으로 추출된 일종의 통제어 색인은 주제 집약성이 강하므로 일반 자연어 색인시스템의 대체시스템으로 가능할 것인가?

본 연구를 통해 색인어의 자동적 추출 가능성이 추론된다면, 그 결과는 2차 법률정보 시스템의 설계자들에게 주제배경이 없이도 특정의 원리에 의하여 자동색인을 가능케 하고, 다른 주제분야의 정보시스템 설계자들에게는 해당 주제 분야의 고유한 지식베이스를 활용하는데 있어서의 시사점을 제공하게 될 것이다.

1. 3 연구의 내용

본 연구의 목적을 이루기 위해 본 연구는 구체적으로 다음과 같은 3가지 측면의 연구를 수행하였다.

첫째, 다른 주제부문에서 해당 주제분야의 특성을 근거로 데이터베이스를 구축한 이론적 근거를 조사하고,

둘째, 현재까지 이루어진 법률정보데이터베이스의 장·단점을 조사하고,

셋째 법률정보의 특성을 추론하는 것이다.

본 연구는 법률정보의 특성상 본문에 포함된 법률조문이나 참조 판례를 그대로 통제어로 사용할 수 있다는 가정을 문헌의 구조 분석을 통하여 추론하는 것에 의해 이루어진다.

2. 선행연구

본 장에서는 법률정보시스템에 관한 선행 연구외에도 1950년대 룬(Luhn, H. P.)의 연구 이후 최근까지 생산된 여러가지 자동색인 기법에 관한 연구 중 지식베이스 중심의 색인 및 인용색인에 관한 선행연구들을 개관하였다.

2. 1 지식베이스 기반의 색인에 관한 연구

이보넨(Iivonen, M., 1995)은 탐색개념이 탐색어로 표현되므로 정보검색은 탐색어와 매우 밀접한 관련이 있다고 보고, 탐색개념이 탐색어로 표현되는 관계에 관한 연구를 수행하였다. 이 연구는 32명의 피험자를 대상으로 12개의 탐색질문에 대하여 서로 다른 4개의 시스템 탐색결과를 측정하는 방법으로 진행되었다. 연구 결과 특정성이 높은 용어를 사용하는 전문직종은 탐색어로서 디스크립터를 사용하여 탐색자 간에 용어의 일치성이 높은 반면 공공도서관에서는 자연어를 탐색어로 선택하여 탐색자간 용어의 일치율이 낮다는 사실과, 최적의 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 해소될 수 있으나 정확성의 문제는 여전히 남게 된다.

몰토(Molto, Mavis, 1993)는, 전문 데이터베이스가 증가함에 따라 그에 관한 연구도 증가하고 있으나 대부분이 언어적인 문제 즉

자연어 색인의 보편적인 결함을 보완하는 방향에서 연구가 이루어지고 있다는 사실을 지적하고 학문주제별로 전문을 심층적으로 분석하여 그 결과를 시스템 설계에 반영할 가치가 있음을 주장하였다. 이는 학문주제별로 특정한 용어가 독특한 의미로 사용되고 있으며, 그 분야의 저자들이 공통적으로 사용하는 언어습관이 있으므로 이에 관한 연구 결과가 지식베이스로 시스템에 내장될 때 시스템의 성능을 향상시킬 수 있다는 것을 의미하는 것이다.

베루(Berrut, Catherine, 1990)는 학문주제별 지식이 색인에 활용될 수 있을 것으로 보았다. 의학분야의 경우 특정한 용어가 특정한 법칙(임상관측 → 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.

2. 2 인용색인에 관한 연구

샬리니(Shalini, R., 1993)는 문헌내에 인용된 자료는 저자가 주제로 나타내고자 하는 개념과 관련이 있다고 판단한 참고자료이므로 이용자는 인용문헌 프로파일에 대하여 자신의 탐색개념과 심리적으로 적합하다고 인식하고 있다는 사실에 근거하여, 인용문헌 프로파일이 실제로 적합률을 향상시키는지에 대한 실험을 수행하였다. 그 결과 인용문헌 프로파일은 색인어의 기능을 수행하면서 적합판단자료로 적용될 수 있음을 밝혀 내

고, 인용하고 있는 인용자료의 디스크립터를 색인으로 추가하는 방안을 제시하였다.

파오와 워텐(Pao, Miranda Lee and Worhten, Dennis B., 1989)은 동일한 데이터베이스 내에서 주제어라고 할 수 있는 디스크립터를 이용한 검색시스템과 문헌내 인용자료를 이용한 검색시스템의 검색 효율을 비교하고 데이터베이스의 검색효율을 증진하는 방안에 관하여 연구를 수행하였다. 연구결과 문헌내 인용자료를 이용한 검색과 디스크립터를 사용한 검색이 서로 다른 결과의 문헌을 검색해낸다는 사실을 발견하고 양자의 통합 검색이 재현율을 증진시킬 수 있다고 주장하였다. 정확률을 증진시키기 위해서는 두 시스템의 검색결과를 통합하여 중복된 문헌만을 받아들이는 방법을 제안하고 있다.

보이스와 맥레인(Boyce, Bert R. and McLain, P., 1989)은 통제어를 이용하여 정확률은 일정 수준으로 유지하면서도 재현율을 높이는 방안에 대하여 연구하였다. 그와 반대로 자연어색인을 이용하여 재현율은 그대로 유지되되 정확률을 높이는 방안으로서 문헌내 높은 빈도를 차지하는 용어를 제거하는 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 유지시키면서 디스크립터 수를 늘려 자연어 색인이 가질수 있는 접근점의 증가 효과를 가져와 결국 재현율도 높일 수 있다고 본 것이다. 그 방법은 본래 문헌에 할당된 표준 디스크립터 외에 문헌이 인용하고 있는 인용자료의 디스크립터를 색인으로 추가하는 것이다.

쇼오(Shaw, W. M.Jr., 1990)는 1976년부터 1979년까지 생산된 1,239개의 의학부문 자료를 대상으로 99개의 탐색질문을 4가지 주제어 색인방법과 2가지 인용색인방법으로 검색 효율을 비교하였다. 실험 결과 인용 색인방법이 주제어 색인방법보다 우월하되 주제어 색인방법과 인용 색인방법을 통합할 경우 효율을 증진시킬 수 있는 것으로 나타났다.

2. 3 법률정보시스템에 관한 연구

2. 3. 1 국내 연구

국내에서의 법률정보를 다룬 최초의 연구는 홍명자에 의하여 이루어졌다. 홍명자(1975)는 국내의 법률관계 참고자료를 조사하여 실용서지를 작성하였는데 수록범위는 해방이후부터 1975년 현재까지 국내에서 출판된 자료로 하였다.

이어 홍명자(1989)는 법제도의 일반적 특성과 법률전문직 제도상의 특성을 기술하고, 제도상의 특성이 법률정보의 생산과 이용에 미치는 영향에 대하여 개괄적으로 설명하였다. 이 연구를 바탕으로 홍명자(Myung -Ja Hong, 1992)는 미국 시티 대학교 에서 한국의 법률정보시스템과 법률 전문가의 이용행태 및 요구사항을 조사한 논문을 발표하였다.

류인모(1994)는 법률정보 데이터베이스는 일반 데이터베이스와는 달리 법학의 전문성이 요구되는 대표적인 전문적 데이터베이스이므로 법학적 사고가 내재된 키워드의 부여와 구조 설계가 필요하다고 주장하였다.

노정란(1990)은 법률정보 이용자의 법률전문 공 여부에 따른 요구유형과 이용관습에 대한

연구를 수행하여 법학전공자나 법률실무자는 법률학 분야의 주제 전문적 배경을 갖고 있기 때문에 사서에 대한 의존도가 낮고, 학술잡지와 같은 저널 등 최신 정보에 대한 이용률이 높다는 사실들을 발견하였다. 그 결과 이용자 지향적인 법률정보 서비스를 위해서는 법률정보시스템의 체계적인 발전과 법률전문사서 연수프로그램의 개발 등이 필요하다고 주장하였다.

1970년대 중반부터 법률정보시스템에 대한 연구가 학계에서 시작되어 왔으며, 최근에는 법률정보시스템을 직접 운영하는 현장에서 법원도서관을 중심으로 국회도서관, 서울대학교 법과대학 도서관 등이 단일 또는 공동으로 연구를 수행하고 있다. 우리나라 최대 법률도서관인 대법원 도서관은 초고속 정보망사업으로 「전자법원도서관 구축」승인을 획득하여 1998년까지 국내외 모든 법률정보를 전산망과 인터넷을 연결, 국내외 어디서나 정보서비스를 제공받을 수 있도록 하고 법률문헌 전자출판, 법률문헌 목록과 색인의 자동화, 법률문헌에 대한 전문 데이터베이스 구축 등을 목표로 하고 있다. 특히 색인시스템 구축의 필요성과 중요성을 인식하고 있음에도 불구하고 기초 서지정보를 이용한 색인 외에는 적절한 방안을 찾지 못하고 있다.

현재 국내에서 제작하였거나 제작 중에 있는 법률정보 데이터 베이스는 법령, 판례 데이터베이스 외에 서울대학교 도서관의 법률문헌색인, 국회·서울대학교 법과대학·한국법제연구원의 GLIN(Global Legal Information Network), 대법원 도서관의 장서 목록, 국회도서관 정기간행물색인 중 법률분

야 등으로 이들 법률정보 데이터베이스들은 주제별로 서지사항을 수록하거나 기초 서지정보를 대상으로 색인하는 수준에 머물러 있다.

상기의 국내 선행연구를 정리하여 볼 때 학계에서는 법률정보의 체계적 관리를 위한 기초연구를, 현장의 법률도서관에서는 이를 바탕으로 실무적 접근을 시도하고 있으나 학계나 현장에서 해결하여야 할 가장 핵심적 과제인 법률정보의 개념추출, 즉 색인방안에 관한 연구가 미흡한 것으로 나타나고 있다.

2. 3. 2 외국의 연구

법률·판례 중심의 1차 법률정보를 컴퓨터를 이용하여 접근하려는 연구 (Computer Assisted Legal Research : CALR, Computerised Legal Information System: CLIS)는 이미 1950년대와 1960년대에서도 계획 또는 실시되었다고도 볼 수 있으나 본격적인 시도는 미국의 대표적 상업 CALR시스템인 LEXIS와 WESTLAW가 개발된 1970년대 부터이며, 이후 두 시스템에 관한 비교·연구가 활발하게 이루어졌다.

법률정보의 검색방법에 대하여 베커와 헤이즈(Becker, J. and Hayes, R. M., 1967)는 법률정보에 능통한 중개자에 의한 검색이 최종 이용자에 의한 검색보다 효과적이라고 주장한 반면 로이드(Lloyd, M., 1986), 베룰과 레위스(Berul, L. H. and Lewis, R. E., 1986)는 최종 이용자에 의한 검색이 더 효과적이라고 주장하였다.

빙(Bing, J., 1987)은 CLIS가 원문검색, 초록을 통한 관련 문헌 검색, 서지사항을 통

한 검색이 가능하다고 설명하면서, 초록만을 통한 검색은 재현율의 저하를 초래하게 되므로 재현율의 증가를 위해서는 원문과 초록이 결합된 형태의 검색이 요구된다고 주장하였다.

쿤즈(Kunz, C. L., 1996)는 CLIS의 장·단점을 분석하였는데 장점은 이용자가 자연어를 통해 신속하게 검색할 수 있다는 점이며, 단점은 개념을 인식하지 못함으로 인한 문제, 오·탈자를 인식 못하는 컴퓨터 처리상의 문제, 유사 용어를 인식하지 못하는 데서 기인하는 정확률 저하의 문제, 상호 연관관계를 규정해 주지 못하는 문제 등이라고 지적하였다.

슬레이드와 그레이(Slade, M. and Gray, R.)는 검색되는 정보의 질을 높이기 위해서는 이용자의 이용행태적 특성을 관찰하여 인간의 지능수준을 시스템에 반영시켜야 한다고 주장하였으며, 스보보다(Svoboda, W. R., 1981)는 효율적인 법률정보시스템을 구축하기 위해서는 법률정보의 특성과 법률분야 전문가의 이용행태적 특성을 고려하여야 한다고 주장하였다.

스프로울(Sprowl, J. A., 1976)은 법률문헌의 검색 가능성은 저자가 선택한 어휘와 이용자가 선택한 용어에 의해서 결정되며, 보즈(Voges, M. A., 1988)는 현행 법률 검색시스템의 문제점으로 여러가지로 해석 가능한 법률적 개념이 단지 구체적으로 제시된 용어에 의해서만 검색되고 있다는 점을 지적하였다.

고에당(Goedan, J. C., 1986)은 수많은 법률정보 속에서 관련정보를 검색하기 위해서는 부가적으로 판결의 요지와 서지사항의 각 부

분들을 구비하는 것이 필요하다고 주장하였고, 실버스타인(Silverstein, S. H., 1990)은 검색질문을 작성하는 방법을 제시하였다.

전문 데이터베이스의 자연어 시스템과 통제언어 데이터베이스의 유용성 비교에 관한 연구는 보즈(1988), 빙(1987), 테노피르(Tenopir, C., 1984), 페레(Perez, E., 1982), 다브니(Dabney, D. P., 1986) 등 많은 학자들에 의하여 수행되어 왔는데, 현재까지 이루어진 연구 결과를 보면 통제어 데이터베이스의 유용성(동의어·유사어·용어간의 계층관계·상호관련어의 통제)에도 불구하고 전문 데이터베이스 시스템의 우월성을 인정하고 있다. 그러나 이용자로 하여금 검색을 위해 수많은 시도를 하게 하는 지적 노력과 비용에 대한 부담 등 전문 데이터베이스 시스템이 지니고 있는 결점을 해결하기 위해서는 색인어와 같은 제한적인 용어통제가 필요하다고 보았다. 겔바트와 스미스(Gelbart, D. and Smith, J. C., 1992)는 향후 CLIS의 발전방향에 대하여 연구하였다. CLIS는 현재까지 연구된 자연어 검색시스템을 토대로 한 새로운 방안으로, 신속한 정보검색을 위한 컨설턴트의 도입, 문헌내용에 기초한 지식기반시스템, 불리언 논리가 아닌 새로운 검색기법, 질문식 수정을 통한 검색방법, 적합성 환류방법을 통한 검색 효율향상 등 현재 일반 정보검색시스템에서 활발하게 연구되고 있는 여러가지 새로운 검색기법을 도입할 것을 제안하였다.

빙(1987)은 일반인들이 법률정보를 이용할 때에 겪게 되는 어려움은 하나의 법률사건이 여러 법률조문과 관련되어 발생하게 되는 법

물정보의 특성상 그 적용법률을 파악하는 것이 매우 힘들다는 것이라고 지적하고 일반인들의 문제해결을 위한 법률정보 시스템을 제안하였다. 이 시스템은 법률주제가 갖는 특성에 근거한 일종의 지식베이스 - 즉, 의사결정과정을 추론화 한 정형 규칙을 생성한 다음 발생사건별로 적용되는 법률조문을 자동적으로 도출시키는 시스템이다.

상기의 외국의 선행연구들에서 볼 수 있는 바와 같이 법률정보시스템에 대하여 전반적으로 연구가 빈약한 국내와는 달리 체계적으로 연구가 진행되고 있음을 알 수 있다. 외국의 선행연구들은 개괄적으로 세가지 방향에서 수행되어 왔다. 첫번째로는 기존 통제언어 데이터베이스와 전문 데이터베이스의 효율에 관하여 활발하게 연구를 진행한 결과 전문검색의 우월성을 입증하는 것이고, 둘째 법률정보 자체의 특성과 그 이용행태에 근거한 시스템 설계를 제안하고, 세째 최신 검색기법을 활용한 검색시스템의 구축에 대한 것이다. 그러나 본 연구가 다루고자 하는 2차 법률정보에 관해서는 구체적으로 논의되지 않았으며, 자동색인에 관한 연구의 경우도 일반 전문 데이터베이스에서 사용하고 있는 빈도, 인접관계 등 일반적 자동색인 방법을 이용한 것이다. 또한 본 연구에서 다루고자 하는 것과 같이 법률정보의 특성을 토대로 법률정보의 내부를 규율하는 기초원리를 추론하고, 그 원리에 근거하여 자동적으로 색인어를 추출하는 색인방법에 대한 연구는 아직까지 이루어지지 않고 있다.

3. 법률정보의 특성

본 장에서는 법률정보의 특성을 기초로 한 색인어의 추출방안을 제안하기 위하여 법률정보의 일반적 특성을 추론화하는 작업을 수행하였다.

일반적으로 법률정보란 법령, 판례, 법률문헌 등을 말한다. 법령과 판례는 모든 법률정보의 기초를 이루는 원시자료라는 점에서 1차 법률정보로 볼 수 있으며, 원시정보에 대하여 특정 관점에서 견해를 피력하거나 원시정보 자체에 대하여 연구 또는 설명을 하는 법률문헌은 원시정보에 대한 2차적 결과물이라는 의미에서 2차 법률정보로 볼 수 있다.

3. 1 법령의 구성체계

법령이란 모든 자연인(또는 법인)의 법률행위를 규율하기 위한 것으로 외적 표현형식은 숫자로 표시된 법령조문과 법령조문을 정의하는 표준화된 용어로 표시된다. 법령은 모든 법률정보를 파생시키면서 법률용어의 기본 개념을 정의하고 있으므로 일종의 코딩이면서 시소러스 또는 용어집 기능을 수행하는 것으로 볼 수 있다. 예를 들어 임의의 법률인 "주택임대차보호법"은 8개의 조문으로 구성되어 있고 제1조에서부터 제8조에 이르기까지 순차적 일련번호를 갖는다. 그 중 제3조를 보면,

제 3 조(대항력등)

- ① 임대차는 그 등기가 없는 경우에도 임차인이 주택의 인도와 주민등록을 마친 때에는 그 익일 부터 제3자에 대항

하여 효력이 생긴다. 이 경우 전입신고를 한 때에 주민등록이 된 것으로 본다.

- ② 임차주택의 양수인(기타 임대차 권리를 승계한 자를 포함한다)은 임차인의 지위를 승계한 것으로 본다.
- ③ 민법 제575조 제1항·제3항 및 제578조의 규정은 이 법에 의하여 임대차의 목적이 된 주택이 매매 또는 경매의 목적물이 된 경우에 이를 준용한다.
- ④ 민법 제536조의 규정은 제3항의 경우에 이를 준용한다.

여기에서 “제3조”로 표기된 형식은 해당 조문을 다른 조문과 구별하여 인식하기 위한 “코드”의 기능을 수행하며, 괄호안의 용어인 (“대항력등”)은 정의(코드의 값)를 나타내며, ①,②,③,④ 항에서 이어지는 문장은 본 법률 해당 조문의 정의를 구체적으로 설명하는 설명문의 기능을 수행한다.

법률정보 사용자들간에는 일반적으로 “주택임대차보호법 제3조에 의하면 . . .”이라는 어구가 상용되고 있는데 이 어구를 사용하는 발신자와 수신자 간에는 이 어구의 의미는 즉시 “대항력”이라는 의미로 대체되어 사용되고 있다. 그러므로 법령 데이터베이스 내에서 “주택임대차보호법 제3조”는 숫자로 표현되어진 코드이고 “대항력”은 이 코드가 갖는 고유의 값이 되는 것이다. 만약 법률문헌 내에서 “주택임대차보호법 제3조”라는 표현이 사용되고 있다면 이 표현은 문헌 내에서는 주택임대차보호법 제3조라는 자연어가 되지만 모든 법률을 정의하고 있는 일종의 용어집인 법령 데이터베이스에서는 임대

차라는 법률관계에 있어서 ‘주택의 인도’와 ‘주민등록’이라고 하는 공시행위를 마치면 “대항력”을 인정한다고 정의함으로써 “주택임대차보호법 제3조”는 곧 “대항력”이라고 하는 법령용어로 변환되는 것이다.

따라서 개별 법률조문의 괄호 안의 법령용어는 법령 데이터베이스내에서 하위 개념을 정의하고 통제하는 기능을 수행하는 상위 통제어가 된다는 사실을 알 수 있다.

3. 2 판례의 구성체계

법령을 모든 법률정보를 파생시키는 가장 기초적 법률정보로 정의한다면 판례란 이러한 법령을 개별 사건에서 구체적으로 해석한 것이다.

자연인(또는 법인)간에 분쟁이 발생하면 이러한 분쟁을 자체적으로 원만히 해결할 수도 있지만 많은 경우 법의 해석을 빌어 해결하고 있다. 이 경우 법원에 소를 제기하여야 하고 이 단계부터는 분쟁의 당사자가 자연인(또는 법인)이 아니라 원고와 피고로서 존재하게 된다. 법원은 원고와 피고간의 분쟁사건을 심리하게 되는데, 심리란 분쟁사건을 해결할 법리를 연구하는 것으로 구체적으로는 분쟁사건에 적용시킬 법령을 찾아 내고 적용하는 과정이다. 심리가 종료되면 법원은 소 제기 사건에 대한 최종 결정을 판시하게 되는데, 판시의 내용이 결정문 형식으로 나타나게 된 것이 판례이다. 선행 판례는 이후 발생하는 판례에 대하여 법적인 구속력과 기판력을 가짐으로써 법령과 마찬가지로 모든 자연인(또는 법인)의 법률 행위시 준거가 되

는 법률정보이다.

모든 판례는 고유의 사건번호, 선고연월일, 사건명으로 표기되고 판시의 근거가 된 법률 조문(참조조문), 선행 판례(참조판례) 등이 명기되어 정형화된 형식으로 표현된다. 예를 들어 임의의 판례를 보면,

1997. 2. 28. 선고 96다 46033, 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(공995,64), ... (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)

여기에서 선고연월일(1997. 2. 28.), 사건번호(96 다 46033, 46040)는 해당 판례를 다른 판례와 구별하여 인식하게 하는 “코드”의 기능을 수행하고, 판결요지는 코드의 값 또는 정의를 나타내며, 참조조문, 참조판례는 본 코드와 밀접한 관련이 있는 “인용자료”의 성격을 갖는다. 기타 원고, 피고, 원심판결, 주문, 이유 등은 본문을 구체화하여 설명하는 성격의 데이터 요소이다.

만약 위의 판례에 대하여 전문의 내용을 요약하는 색인어를 추출하려고 한다면 판결요지, 판결이유 필드 등에 출현하는 모든 자연어를 대상으로 추출하는 방법도 있을 수 있으나 참조조문 필드에 나타난 법률조문을 근거로 “주택임대차보호법 제3조” ⇒ “대항력”이라는 색인어 추출도 가능할 것이다. 이와 같은 사실을 가정할 수 있는 근거는 판례요지 필드의 내용에서 보는 바와 같이 본 판례의 초점은 주택임대차의 대항력을 다루는 것이기 때문에 출현빈도, 언어처리 등을 통한 자연어 색인방법과 유사한 또는 일정 수준 이상의 색인어 추출이 가능하기 때문이다.

일반적으로 주택임대차보호법 제3조에 관하여 논의되고 있는 상황이라면 이와 같은 사실은 곧 법령 데이터베이스에서의 용어정의에 의하여 “대항력”을 말하고자 하는 것이며, 이와 관련하여 위의 예와 같은 주택임대차보호법 제3조에 관한 판례가 이미 생성되어 있는 경우에는 판례 또한 법령과 마찬가지로 무시되어서는 안되는 준거정보로서 작용한다.

3. 3 법률문헌의 구성체계

법령과 판례를 제외한 2차 법률정보는 법률이나 판례에 관하여 해석, 평석 등을 가한

연구결과물이므로 법령이나 판례와 사용어휘 또는 근거범위를 같이 하게 된다. 예를 들어 임의의 법률문헌을 보면,

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

一. 서

주택임대차보호법 제3조의2 제1항은 임대차계약서에 확정일자를 갖춘 임차인의 임차보증금에 관하여 우선변제권을 인정하고 있고, 법 제8조 제1항은 소액임차인의 임대차보증금에 관하여 최우선변제권을 인정하고 있다. 임대차보증금 반환청구권은 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 판결은” 주택임대차보호법 제4조 제1항은 임차인을 보호하기 위한 규정으로서, 기간을 2년미만으로 정한 때에는..... ”

(중략)

4. 사건

(중략)

민사소송법 제603조의2제1항에 의하면 법원이 경매개시결정후 집행관에게 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 누락되거나 잘못되면 원칙적으로 경락불허가사유가 된다. (민사소송법 제633조 제5호)

(중략)

이 글의 저자가 다루고자 한 최상위 개념(topicality)은 대항력을 갖춘 임차인인 경우 목적 주택이 경매될 때 임차기간 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 제603조의2·제617조의2·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.

4. 결 론

본 연구는 2차 법률정보 전문데이터베이스의 구축시 핵심 과제인 색인어의 추출문제를 법률정보의 내재적인 특성에 근거하여 해결하고자 이론적인 분석을 시도한 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. 그러므로 이와 같은 2차 법률정보의 특성은 2차 법률정보 데이터베이스 색인시스템의 지식베이스로 할 수 있을 것이라는 추론이 가능하며, 그 결과 2차 법률정보 데이터베이스는 모든 전문데이터베이스가 안고 있는 개념어의 추출 부담 문제로부터 상당히 자유로울 수 있을 것이다.

본 연구는 2차 법률정보에서 인용되고 있는 법률과 판례가 2차 법률정보의 내용을 표현 또는 집약한다는 사실을 밝혀냄으로써 이러한 사실을 근거로 2차 법률정보내의 인용 법령과 판례로부터 색인어의 자동적 추출 가능성을 제안하고자 하였다.

참 고 문 헌

- 노정란. 금융기관에서의 법률정보 요구와 이용 관습에 관한 연구. 석사학위논문, 성균관대학교, 1990.
- 류인모. 법령정보데이터베이스와 법적 문제. 제 5 회 법제 세미나 주제발표 (법령 정보서비스와 국민의 법생활). 한국법제연구원, 1994.
- 문용호. "법률정보 검색의 동향." 국회도서관보 34:4(1997.7·8):3-26.
- 정영미. "국내문자정보 데이터베이스의 색인에 관한 연구." 情報管理學會誌 13:1(1996.6.):19-45.
- 홍명자. 한국법률학관계 참고자료에 관한 고찰. 석사학위논문, 이화여자대학교, 1975.
- 홍명자. "한국의 법제도와 법률정보와의 관계에 관한 고찰." 도서관 44:4(1989):3-17.
- Myung-Ja Hong. Access to Legal Information in Korea. Doctoral Dissertation, City University, 1992.
- Becker, J. and R.M.Hayes. Information Storage and Retrieval : Tools, Elements, Theories. New York : John Wiley & Sons, Inc., 1967.
- Berul L. H. and R.E.Lewis. The Legal Profession's New Look : End Users of Information Technology. In:ASIS '86 Proceedings of the 49th Annual Meeting of ASIS 23(1986):16-19.
- Berrut, Catherine. "Indexing Medical Reports : The RIME Approach." Information Processing & Management 26:1(1990):93-109.
- Bing, J. "Developing Knowledge Based Legal Systems for the Public Administration." LAW/Technology 20:1(1987):1-59.
- Bing, J. "Performance of Legal Text Retrieval Systems : the Curse of Boole." Law Library Journal 79:2 (1987):187-202.
- Boyce, Bert R. and P. McLain. "Entry Point Depth and Online Search Using a Controlled Vocabulary." Journal of the American Society for Information Science 40:4(1989) :273-276.
- Dabney, D. P. "The Curse of Thamus : an Analysis of Full Text Legal Document Retrieval." Law Library Journal 78:1(1986):5-40.
- Edwards, John D. "LEXIS & Training Centers : Law School Opportunities." Law Library Journal 80:3 (1988) :459-472.
- Fugmann, Robert. "The Complementarity of Natural and Indexing Languages." International Classification 9(1982) :140-144.
- Gelbart, D. and J.C.Smith. "Towards Combining Automated Text

- Retrieval and Case - based Expert Legal Advice.” *Law Technology Journal* 1:2(1992):19-24.
- Goedon, J. C. “Legal Comparativists and Computerized Information System : General Problems and Present German Status of Computerized Legal Information.” *International Journal of Legal Information* 14:1/2 (1986):1-49.
- Green, Rebecca. “ Topical Relevance Relationships. I .Why Topic Matching Falls.” *Journal of the American Society for Information Science* 46:9(1995):646-653.
- Green, Rebecca and Carol A.Bean. “Topical Relevance Relationships. II .An Exploratory Study and Preliminary Typology.” *Journal of the American Society for Information Science* 46:9(1995):654-662.
- Iivonen, Mirja. “Consistency in Selection of Search Concepts and Search Terms.” *Information Processing & Management* 31:2(1995):173-190.
- Kunz, C.L. *The Process of Legal Research*. Little Brown and Co., 1996.
- Larson, Signe E. and Martha E.Williams. “ Computer Assisted Legal Research.” *Annual Review of Information Science and Technology* (1980):251-285.
- Lloyd, M. *Legal Database in Europe : User Attitude and Supplier Strategies*. Amsterdam : North-Holland, 1986.
- Molto, Mavis. “Improving Full Text Search Performance through Textual Analysis.” *Information Processing & Management* 29:5 (1993):615-632.
- Nazareth, A. M. “Legal Databases, Legal Epistemology, and the Legal Order.” *Law Library Journal* 86 (1994):700.
- Pao, Miranda Lee and Dennis B.Worhten. “ Retrieval Effectiveness by Semantic and Citation Searching.” *Journal of the American Society for Information Science* 40:4(1989):226-235.
- Perez, E. “Text Enhancement : Controlled Vocabulary vs. Free Text.” *Special Libraries* 73:33(1982):183-192.
- Price, Kathleen. “Building a Law Library to Serve the World Legal Community.” *Law Library Journal* 83 (1991):225-233.
- Shalini, R. “Citation Profiles to Improve Relevance in a Two-stage Retrieval System: a Proposal.” *Information Processing & Management* 29:4 (1993):463-470.
- Shaw, W. M.Jr. “Subject Indexing and

- Citation Indexing-PART II : An Evaluation and Comparison." *Information Processing & Management* 26:6(1990):705-718.
- Silverstein, S. H. "An Index Model for Query Formulation." *Legal Reference Services Quarterly* 10:3 (1990):115-124.
- Slade, M. and R.Gray. *The Impact and Potential of Online Legal Research System on Academic Legal Education*. British Libray R&D Report 5812.
- Smeaton, Alan F. "Natural Language Processing and Information Retrieval." *Information Processing & Management* 26:1(1990):19-20.
- Spink, Amanda and Tefko Saracevic. "Interaction in Information Retrieval : Selection and Effectiveness of Serch Terms." *Journal of the American Society for Information Science* 48:8(1997):741-761.
- Sprowl, J. A. "Computer-Assisted Legal Research : an Analysis of Full-Text Document Retrieval Systems, Particularly the Lexis system." *American Bar Foundation Research Journal* 1976:175-226.
- Svoboda, W. R. *Study on Common Standards for Query Languages in Computerised Legal Retrieval Systems*. Strasbourg : Council of Europe, 1981.
- Tenopir, C. "Full-text Databases." *Annual Review of Information Science and Technology* 19(1984) : 215-246.
- Voges, M. A. "Information Systems and the Law." *Annual Review of Information Science and Technology* 23 (1988):193-215