

2차 法律情報 全文데이터베이스에 있어서 統制語 索引시스템과 自然語 索引시스템의 檢索效率 評價에 관한 研究

A Study on the Indexing System Using a Controlled Vocabulary and Natural Language in the Secondary Legal Information Full-Text Databases : an Evaluation and Comparison of Retrieval Effectiveness

盧 貞 蘭 (Jeong-Ran Roh)*

목 차

1. 서론
2. 연구방법

3. 결론

초 록

본 연구는 2차 법률정보 전문 데이터베이스 구축을 위한 기초연구(권기원, 노정란, 1998, 한국문헌정보학회지, 32(3))에서 밝혀진 법률정보의 특성을 근거로 알고리즘을 개발하고, 알고리즘에 의한 모형 통제어 데이터베이스를 구축하여 통제어 색인 시스템과 자연어 색인 시스템의 검색효율을 비교, 평가한 것이다. 연구 결과 2차 법률정보 전문 데이터베이스에서 통제어 색인 시스템은 재현율, 정확률, 자연어 시스템이 검색하지 못한 고유한 적합 문헌을 검색하는 능력에 있어서 자연어 색인시스템보다 높은 효율을 나타내었다. 또한 일반적으로 가중치를 부여하거나 접근점을 추가할 경우 데이터베이스의 정확률이나 재현율의 향상을 가져올 수 있다고 보고 있으나, 2차 법률정보 전문 데이터베이스에서는 법률정보라는 특정 지식 분야의 특성으로 인하여 가중치를 부여하거나 접근점을 추가한 경우에도 재현율과 정확률의 향상을 나타내지 않는다는 사실이 밝혀졌다. 그러므로 정보시스템 설계자는 시스템을 단순히 언어학적, 통계학적 방법으로 접근하기 보다는 정보전문가와 주제전문가가 인식하고 있는 각 주제분야의 고유 지식을 시스템에 내장시키는 것이 필요하다고 할 수 있다.

ABSTRACT

The purpose of this study is to develop the indexing algorithm of secondary legal information by the study of characteristics of legal information, to compare the indexing system using controlled vocabulary to the indexing system using natural language in the secondary legal information full-text databases, and to prove propriety and superiority of the indexing system using controlled vocabulary. The results are as follows; 1) The indexing system using controlled vocabulary in the secondary legal information full-text databases has more effectiveness than the indexing system using natural language, in the recall rate, the precision rate, the distribution of propriety, and the faculty of searching for the unique proper-records which the indexing system using natural language fails to find. 2) The indexing system which adds more words to the controlled vocabulary in the secondary legal information full-text databases does not better effectiveness in the recall rate, the precision rate, comparing to the indexing system using controlled vocabulary. 3) The indexing system using word-added controlled vocabulary with an extra weight in the secondary legal information full-text databases does not better effectiveness in the recall rate, the precision rate, comparing to the indexing system using word-added controlled vocabulary without an extra weight.

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.

* 成業公社 法規室 課長 (成均館大學校 및 同德女子大學校 講師)

접수일자 1998년 10월 16일

1. 서론

1. 1 연구의 필요성

2차 법률정보 데이터베이스는 그 필요성에서 뿐만 아니라 專門 데이터베이스 가운데에서도 수요가 높은 대표적인 데이터베이스임에도 불구하고, 법률주제분야가 갖는 학문의 난해함과 전문성 때문에 현재까지도 그 설계방법에 대하여 명확한 해법을 갖고 있지 못하고 있다. 본 연구에 앞서 수행한 2차 법률정보 전문 데이터베이스의 기초 연구(분석 내용 참조: 權起遠, 盧貞蘭, 1998, 2차 法律情報 全文데이터베이스 構築을 위한 基礎 研究, 한국문헌정보학회지 32(3))에 의하면 법률주제분야가 갖는 이러한 뚜렷한 전문성이야말로 오히려 2차 법률정보 전문 데이터베이스의 설계시 고유한 지식베이스로 적용가능할 것으로 보고 있다. 만약 위의 연구에서 추론한 법률정보의 고유한 특성과 색인어의 자동적 추출 가능성이 객관적 실험을 통하여 입증되고, 궁극적으로 검색효율의 향상을 가져온다면, 이러한 결과는 각 專門데이터베이스의 설계시 해당 주제분야의 학문적 특성을 활용하는 계기가 될 것이다. 그러므로 법률정보의 특성에 근거하여 설계한 데이터베이스의 검색효율이 기존 방식으로 설계한 데이터베이스의 검색효율보다 우월한지 여부를 평가하여, 그 타당성을 검증하는 연구가 필요하다.

1. 2 연구의 목적

본 연구는 향후 전문(full-text) 데이터베이스로 구축될 2차 법률정보를 대상으로 법률정보가 갖는 주제의 난해함과 용어상의 문제점을 해결하면서, 전문 데이터베이스가 갖는 취약점을 해결하여 정확률을 향상시킬 수 있는 방안을 모색하려는 것이다. 즉, 기존의 일반적인 자연어 색인기법이 아닌 법률정보의 고유한 특성을 주제베이스로 한 2차 법률정보 색인 알고리즘을 개발한 다음, 알고리즘에 의하여 생성된 새로운 통제어 색인방법과 자연어 색인방법의 검색 효율성을 비교하고 그 차이를 검증하는데 있다.

구체적으로 자동적으로 추출된 일종의 통제어 색인이 일반 자연어 색인시스템보다 검색효율을 향상시킬 것인지, 자동적으로 추출된 일종의 통제어 색인은 주제 집약성이 강하므로 용어를 추가하거나 특별한 장치가 없어도 높은 검색효율을 보일 것인지에 대하여 평가해 보려는 것이다.

1. 3 연구의 내용

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

첫째, 선행연구(권기원, 노정란, 1998)에서 추론된 법률정보의 내재적 특성에 따라 2차 법률정보의 색인알고리즘을 개발하며, 둘째, 이 알고리즘에 따라 추출된 데이터 요소를 지식베이스로 하는 데이터필드를 생성하고, 셋째, 새로 구축된 시스템과 원래의 시스

템의 검색 효율성을 비교하는 것이다.

1. 4 용어의 정의

본 연구에서 주로 사용하고 있는 용어는 다음과 같이 정의한다.

(1) 1차 법률정보

원문 그대로의 법률(시행령, 시행규칙 포함), 예규, 조례, 판례 등을 말한다.

(2) 2차 법률정보

1차 법률정보의 이해를 돕기 위해 마련한 법해석서, 주석서, 일반교재류, 저널 기사, 논문 등을 말한다.

(3) 자연어 색인시스템

2차 법률정보의 본문에서 사용하고 있는 자연어를 색인어로 한 시스템이다.

(4) 통제어 색인시스템

2차 법률정보의 본문에 나타난 그대로의 자연어가 아니라 본문에서 인용하고 있는 법률조문 또는 판례를 추출한 후 일종의 용어집 기능을 수행하는 법령 데이터베이스에서 색인어를 추출하는 시스템이다. 인용사항이 법률조문인 경우에는 법령 데이터베이스에서 그대로 해당 법률조문의 괄호 안의 용어로 변환하고, 판례인 경우에는 먼저 판례 데이터베이스에서 참조 법률조문을 읽어 들인 후 이 조문을 법령 데이터베이스에서 같은 방법으로 법령용어로 변환하여 색인어를 생성한다.

(5) 용어 수 추가 통제어 색인시스템

통제어 색인시스템과 같은 방법으로 색인어를 추출하되 법률조문의 괄호 안

용어와 더불어 괄호 뒤의 설명문에 나타나는 용어들을 추가 색인어로 하는 시스템이다.

(6) 가중치 부여 통제어 색인시스템

용어를 추가한 통제어 색인시스템에서 법률조문의 괄호 안의 법률용어와 괄호 뒤 설명문에 나타난 법률용어간에 가중치를 달리 하면서 색인어로 생성한 시스템이다.

괄호 안 법률용어의 가중치 값을 1로 하고 괄호 뒤 설명문에 나타난 용어의 가중치 값을 각각 0.25, 0.50, 0.75, 0.90으로 한다.

(7) 검색효율

재현율, 정확률, 적합성 분포 판정 및 상호 중복되지 않는 적합문헌수와 비율을 말한다.

(8) 성능곡선

표준 재현율을 이용하여 모든 질문의 평균적인 검색성능을 표현한 곡선을 말한다.

2. 연구방법

본 연구는 법률정보의 특성을 토대로 특히 2차 법률정보 전문데이터베이스에서 본 연구자가 제안하고 있는 통제어 색인방법과 일반적 자연어 색인방법을 비교하고 통제어 색인방법이 높은 검색효율을 보일 것이라는 가설을 검증함으로써 연구의 목적을 이루고자 하였다.

2. 1 가설

본 연구에서 설정한 영가설(null hypothesis)은 첫째, 2차 법률정보의 자연어 색인 시스템은 통제어 색인시스템, 용어수를 추가한 통제어 색인시스템과 비교하여 검색효율이 높거나 적어도 차이가 없을 것이며, 둘째, 2차 법률정보의 용어 수 추가 통제어 색인시스템은 가중치 통제어 색인시스템과 비교하여 검색효율의 차이가 없다는 것이다.

영가설의 검증을 통하여 입증하고자 한 대립가설 또는 연구가설(research hypothesis)은 다음과 같다.

법률정보의 특성상 2차 법률정보는 1차 법률정보를 인용하는 형식을 취하고 있으며, 인용된 1차 법률정보는 다른 주제분야에서의 인용자료의 기능보다 주제를 집약하는 기능이 강하다. 따라서 2차 법률정보에 있어서 주제 집약 기능이 강한 인용 1차 법률정보를 이용하여 색인어를 자동 추출하는 통제어 색인시스템은 오히려 본문에 나타난 일반 자연어를 색인어로 한 시스템과 비교하여 높은 검색효율을 보이게 될 것이다.

둘째, 인용 1차 법률정보를 이용한 통제어 색인방법은 통제어 자체의 집약기능이 강하므로 용어 수를 추가하여도 검색효율의 향상을 나타내지 않을 것이다.

셋째, 인용 1차 법률정보를 이용한 통제어 색인방법은 통제어 자체의 집약기능이 강하므로 가중치를 부여하는 경우에도 검색효율의 향상을 나타내지 않을 것이다.

2. 2 가설 검증 방법

본 연구의 가설은 동일한 환경 아래에서 통제어 색인시스템, 자연어 색인시스템, 용어 수 추가 통제어 색인시스템의 검색효율을 비교하고, 아울러 용어 수 추가 통제어 색인시스템과 추가한 용어간에 가중치를 부여한 통제어 색인시스템의 검색효율을 비교함으로써 궁극적으로 통제어 색인시스템의 가능성을 검증하려는 것이다.

가설의 검증은 실제 모형 실험환경을 구성하여 각 시스템에서 탐색을 수행한 후 결과를 수집하고, 그 결과들을 통계적 방식을 이용하여 분석하는 단계로 진행하였다.

가설을 검증하기 위한 실험환경은 크게 일곱 단계로 나누어 구성하였다. 각 단계를 보면,

첫째, 알고리즘 개발 단계

둘째, 실험용 데이터베이스 구축단계

셋째, 통제어, 자연어, 용어 수 추가 통제어, 가중치 부여 통제어 색인방식을 위한 색인 프로그래밍 단계

넷째, 탐색을 위한 탐색질문 작성 단계

다섯째, 탐색질문에 대한 각 색인시스템별 실제 탐색 단계

여섯째, 탐색결과 분석을 위한 적합성 판정기준 설정 단계

일곱째, 체크리스트에 의한 시스템 효율 측정단계로 진행되었다.

2. 2. 1 알고리즘 개발

즉, 법률정보의 구성체계를 분석한 결과 2차 법률정보의 속성에서 하나의 일정한 논리

를 발견할 수 있는데, 그 논리는 기사의 본문에 나타난 법률조문이 그 본문의 내용을 표현하는 개념어(content representation keyword)가 된다는 것이며, 이 개념어를 기사의 색인어로 채택할 수 있다는 것이다.

본 연구는 이러한 사실을 근거로 인용 법령과 판례로부터 색인어를 추출하는 알고리즘을 개발하고자 하였다.

2. 2. 2 알고리즘에 의한 실험용

데이터베이스 구축

법률정보의 특성에 따라 개발된 2차 법률정보의 통제어 색인알고리즘의 효율성을 검증하기 위하여 다음과 같이 실제 실험용 모형데이터베이스를 구축하였다.

모형의 개발은 다음과 같이 세 단계로 나누어 수행하였다.

<제1단계>

알고리즘 구현을 위한 기초 데이터베이스 구축

(1) 2차 법률정보 데이터베이스 구축

전문 데이터베이스 실험환경을 조성하기 위하여 국내 최대 법률도서관인 대법원도서관이 소장하고 있는 국내 2차법률정보 중 저작권 사용에 동의한 저자들의 문헌을 대상으로 하였다. 이들 문헌 중 1991년에서1996년 사이에 생산된 476개 문헌의 전문을 실험 2차 법률정보 데이터베이스로 구축한다.

(2) 법령 데이터베이스 구축

2차 법률정보의 전문에 포함된 법률조문에서 자동적으로 색인어를 추출하기 위하여 1997. 2. 28 현재 대한민국의

유효한 2,219개 현행법령【법률 607개 (25320 조문), 시행령 458개 (12711 조문), 시행규칙 등 1,154개 (27672 조문)】을 법령데이터베이스로 구축한다.

(3) 판례 데이터베이스 구축

전문에 포함된 판례를 인식하여 판례와 관련있는 참조 법률조문을 추출하기 위하여 1945년~1996년까지 생성된 대법원의 판례 데이터베이스를 실험환경의 데이터베이스로 구축한다.

<제2단계>

알고리즘에 의한 데이터필드 구축

(1) 제1단계에서 구축한 2차 법률정보

데이터베이스의 476개 개별문헌의 전문을 대상으로 본문에 나타난 법률조문, 판례를 추출한다.

(2) 추출된 법률조문과 판례는 제1단계

에서 구축한 법령·판례 데이터베이스와 연계시켜 두가지 방법으로 서로 다른 필드를 생성하게 하는데, 그 하나는 위 법률조문과 판례를 해당 법률조문의 괄호 안에 있는 용어로 변환하여 이 용어들로서 키워드군을 형성하는 것이다.

(3) 추출된 법률조문과 판례는 (2)의

방법과 동일하게 법령·판례 데이터베이스와 연계시킨 다음 해당 법률조문의 괄호 안에 있는 용어 뿐 아니라 괄호 뒤 법률조문의 설명문에 나타난 용어로 변환하여 별도의 키워드군을 형성한다.

<제3단계>

알고리즘 실험을 위한 시스템 구축

- (1) 알고리즘에 의하여 형성된 두개의 서로 다른 필드는 개별 문헌의 전문 필드와 같이 본 연구의 가설을 검증하기 위한 3가지 기본 시스템을 구성한다.
- (2) 주제어 사전이 붙은 2차 법률정보 데이터베이스의 각 본문들은 자연어 색인방법을 실험하는 색인시스템이 된다.
제2단계에서 법률조문의 괄호 안 용어로 구성된 키워드군은 탐색시 통제어 색인방법을 실험하는 색인시스템으로 한다.
- (3) 제2단계에서 법률조문의 괄호 안 용어 및 괄호 뒤 법률조문 설명문에 나타난 용어로 구성된 키워드군은 탐색시 용어 수를 추가한 통제어와 가중치 부여 통제어 색인방법을 실험하는 색인시스템으로 한다.

구축된 실험용 2차 법률정보 데이터베이스의 환경을 요약하면 <표 1>과 같다.

2. 2. 3 자연어, 통제어, 용어 수 추가 통제어, 가중치 부여 통제어 색인방식을 위한 색인 프로그래밍

각 색인시스템의 검색효율을 비교하기 위한 색인 알고리즘은 그림1과 같이 구성하였다.

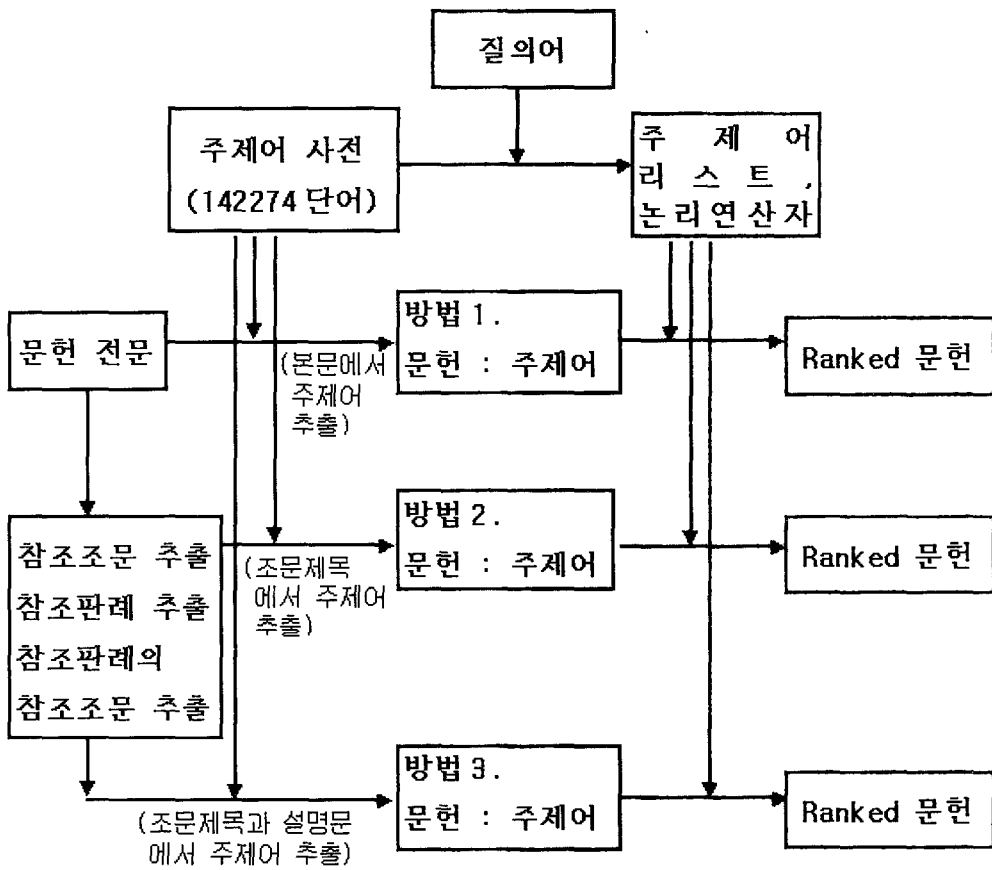
<그림 1>의 색인어를 추출하는 3가지 방법(방법1, 2, 3)은 다음과 같다.

• 방법 1

자연어 색인방식이다. 색인된 문헌의 가중치를 1로, 색인되지 않은 문헌은 가중치를 0으로 하여 확장 불리언 모델(Extended Boolean Model)을 이용하여 검색을 수행하고, 본 모델의 랭킹 알고리즘(Ranking Algorithm)을 사용하여 상위

<표 1> 실험용 2차 법률정보 데이터베이스

구 분	자연어 시스템	통제어 시스템	용어 수 추가 통제어시스템	가중치 부여 통제어시스템
데이터베이스명	본 연구를 위한 실험용 2차 법률정보 데이터베이스			
데이터베이스 크기	Sentence : 652,008 Word : 2,410,211 Byte : 20,217,804			
문 헌 길이 (Range)	Sentence : 160-14,644 Word : 580-38,053 Byte : 4,879-332,309			
검 색 엔 진	확장 불리언 모델 (Extended Boolean Model)			
Ranking 여부	Ranking			
색인어 화일	본문 (전문)	본문에 나타난 법령 또는 판례		
색인어 평균 단어수	참조조문수:min.1-max.1,475 mean.50.75			



〈그림 1〉 색인 알고리즘 구성도

순위의 문헌을 선택한다.

• 방법 2

본 연구의 알고리즘에 의한 방식이다. 문헌에 포함된 법률조문, 판례(판례데이터베이스에서 참조조문을 읽어 들인 후 법령데이터베이스에서 법률조문으로 변환)의 괄호 안 용어를 주제어로 선정하여 문헌의 색인을 수행한다. 어떤 임의의 주제어가 괄호 안 용어에 포함되어 있는 경우, 그 조문을 참조하는 문헌의 가중치는 1이고, 그렇지 않은 경우는 0이다. 문헌의

검색은 이들 문헌의 가중치를 이용한 확장 불리언 모델을 이용하여 검색을 수행하고, 본 모델의 랭킹 알고리즘을 사용하여 상위 순위의 문헌을 선택한다.

• 방법 3

방법2와 동일한 알고리즘에 의하여 동일한 방식으로 수행되 어떤 임의의 주제가 괄호 안 용어 및 괄호 뒤 설명문에 포함되어 있는 경우, 그 조문을 참조하는 문헌의 가중치는 1이고, 그렇지 않은 경우는 0이며, 가중치 실험에서는 괄

호 안 용어에 포함되어 있는 경우는 1로, 조문 설명문에 포함된 경우는 4회에 걸쳐 0.25, 0.50, 0.75, 0.90의 가중치 값을 갖고, 그렇지 않은 경우는 조문 제목에서와 마찬가지로 0이다. 문헌의 검색은 이들 문헌의 가중치를 이용한 확장 불리언 모델을 이용하여 검색을 수행하고, 본 모델의 랭킹 알고리즘을 사용하여 상위 순위의 문헌을 선택한다.

2. 2. 4 탐색을 위한 탐색질문 작성 방법

각 시스템간의 효율성을 비교하기 위하여 시스템마다 동일한 탐색질문을 부여하여 탐색을 수행하였다.

일반적으로 탐색질문을 작성하는 방법은 피험자인 탐색자 스스로 탐색질문을 선택하게 하는 방법, 탐색에 참여하지 않는 이용자들로부터 탐색질문에 대하여 신청을 받는 방법, 탐색자에게 다수의 탐색질문을 제공하는 방법 등의 3가지 방법이 있다.(Walker, G., 1988, Sullivan, M. V., Borgman, C. L. and Wippem, D, 1990) 본 연구에서는 실제 탐색에 참여하지 않는 이용자들로부터 탐색질문에 대하여 신청을 받는 방법의 일종으로서 특정 분야의 법률에 국한되지 않고 전 분야의 법률에 걸쳐 실제로 질의가 이루어져서 정보요구과일에 이미 등록된 탐색질문 중에서 임의의 질문 15개를 선정하였다.

이러한 파일을 선정한 이유는 두가지 측면을 고려했기 때문이다. 첫째, 특정 법률 이용자 집단의 요구에 의해 작성된 정보요구사항은 대개 특정 분야의 법률로 제한되므로 전 법률분야에 걸쳐 고르게 분포된 정보요구

를 채취하기 위해서는 본 화일이 가장 적합하기 때문이다. 둘째, 본 연구에서는 자연어 색인시스템이 법률분야에서도 적용될 수 있는가 또는 통제어 방식이 더 우월한가를 면밀하게 측정하여야 하므로, 예를 들어 일반 정보시스템에서와 같이 단순히 “저당권”에 대하여 검색명령을 수행한다면 시스템간의 효율 비교시 변별성을 파악할 수 없기 때문에 본 파일을 선정하였다. 본 파일의 경우 실제 구체적으로 제기된 정보요구를 법률전문가가 정교하게 정리한 고도로 정밀한 수준의 탐색질문 파일이므로 정교한 질문이 요구되는 본 연구의 방법과 일치하였다.

탐색질문의 선정은 1991년 이후 발생한 것으로 실험 2차 법률정보 데이터베이스 내에서 한 법률조문이 10개이상의 문헌에서 인용되고 있으면서, 1991년 이후 개정되지 않은 법률조문을 주요 참조조문으로 하고 있는 것으로 한정하였다.

탐색질문의 선정을 한정된 근거는 다음과 같다. 첫째, 법률은 빈번하게 제정 또는 개·폐되는 분야이기 때문에 정보요구의 시점과 적용 법률의 시점이 일치하여야 정확한 정보의 제공이 가능하게 된다. 따라서 본 연구의 실험 2차 법률정보 데이터베이스의 저술 시점이 1991년 이후이므로 정보요구도 1991년 이후 발생한 것 중 1991년 이후 개정이 이루어지지 않은 법률조문과 관련있는 정보요구로 제한한 것이다.

둘째, 본 연구는 유사계수에 의하여 관련성이 높은 순위대로 검색결과가 표현되므로 순위에 따라 각각의 시스템을 비교하기 위해서는 적어도 일정 개수 이상의 관련문헌을

검색해 내기 위한 방안이 필요하다. 그러므로 본 연구의 알고리즘에 의하여 각 문헌에서 인용하고 있는 법률조문 또는 판례를 추출한 다음 통합 실험 데이터베이스 내에서의 법률조문이 10개이상의 문헌에서 인용되고 있으면서, 1991년 이후 개정되지 않은 법률조문을 대상으로 그 조문을 주요 참조조문으로 하고 있는 임의의 질문을 선정하였다.

2. 2. 5 탐색질문에 대한 각 색인시스템별 탐색 수행

실제 정보요구파일에서 선정된 탐색질문으로 탐색을 수행하기 위하여 다음과 같은 세부기준이 마련되었다.

(1) 탐색식

탐색식은 각 시스템에 동일하게 적용하였으며, 확장 불리언 모델을 사용하였다.

(2) 탐색어 선정 및 탐색 수행

15개의 탐색질문에 대한 탐색어(부록 1참조)는 법률정보 탐색전문가인 본 연구자가 1차로 추출한 후 법률전문가와 협의를 거쳐 최종 탐색어를 선정하였으며, 탐색은 본 연구자가 직접 수행하였다.

(3) 탐색결과 표현방법

탐색결과는 각 시스템마다 확장 불리언 모델의 랭킹알고리즘을 사용하여 유사도가 큰 순서대로 순위를 부여하였다.

2. 2. 6 탐색결과 분석을 위한 적합성 판정 기준 설정

15개의 탐색질문에 대하여 서로 다른 색인시스템을 통한 탐색결과를 비교하고 시스템

간의 검색효율을 분석하기 위해서는 동일한 적합성 판정기준이 필요하다. 탐색결과와 적합성을 판정하기 위하여 시스템내 적합문헌 총수 산출, 적합성 판정자, 적합성 측정도구, 적합성 판정을 위한 제시자료 등에 대한 기준을 다음과 같이 설정하였다.

(1) 적합문헌 총수의 산출

본 연구는 제1단계에서 자연어, 통제어 및 용어 수 추가 통제어 색인시스템의 3가지 시스템을 비교하고, 제2단계에서 용어 수 추가 통제어와 가중치를 달리하는 4가지 통제어 색인시스템을 비교하는 방식으로 진행되고 있어 시스템내 존재하는 적합문헌 총수도 두 단계로 나누어 산출하였다.

첫번째 단계는 자연어, 통제어, 용어 수를 추가한 각 색인시스템이 수행한 탐색결과와 적합문헌을 합쳐 잠정적인 상대 적합문헌집단을 만든다. 그 다음 3개의 시스템에서 각각 랭킹 1위에서 10위까지로 절단하고 중복된 적합문헌을 제외하여 최종 적합문헌집단을 만든다. 이 집단의 문헌건수가 각 시스템의 재현율, 정확도를 산출시 기준 적합문헌총수가 된다.

두번째 단계는 용어 수를 추가한 통제어 시스템과 원래의 통제어를 1의 가중치 값으로 하고, 추가한 용어들을 각각 0.25, 0.50, 0.75, 0.90의 값으로 하는 가중치 부여 통제어시스템들이 개별적으로 수행한 탐색결과와 적합문헌을 합쳐 잠정적인 상대 적합문헌집단을 만든다. 그 다음 비교하려는 5개의 시스템에서 각각 랭킹 1위에서 10까지로 절단하고 중복된 적합문

헌을 제외하여 최종 적합문헌집단을 만든다. 이 집단의 문헌건수가 가중치를 달린 통제어 시스템의 재현율, 정확률 산출시 기준 적합문헌총수가 된다.

(2) 적합성 판정자

본 연구는 이미 실제로 정보요구화일에 등록된 질문을 탐색질문으로 선정하였으므로 일반 이용자를 실험대상으로 하지 않았다. 따라서 탐색결과의 적합성 판정(Swanson, Don R., 1986)은 이용자가 아닌 고도의 법률지식을 습득하고 사법 실무를 담당하고 있는 현직 재판연구원 3명으로 하여금 객관적 적합판정을 수행토록 하였다.

(3) 적합성 측정도구

탐색결과의 적합성 측정방법(Myke, G., 1995)에 대한 다양한 주장이 제기되고 있으나 본 연구에서는 매우 적합, 적합, 부적합, 매우 부적합의 4등급 척도를 사용하였으며 판단자를 홀수인 3명으로 하여 문헌당 “2인이상”이 “적합 이상(적합, 매우 적합)”의 판정을 내렸을 경우 “적합”한 것으로 판정하였다. “적합”의 기준을 “매우 적합”으로 하지 않고 “적합 이상”으로 한 이유는 본 연구의 목적이 재현율 뿐만 아니라 정확률도 측정하여 시스템간의 성능을 비교하기 위하여 탐색질문을 정교하게 작성하였으므로 “매우 적합”기준을 적합 기준으로 사용할 경우 적합문헌 건수가 매우 적어 시스템간의 변별성을 측정하는 것이 어렵다고 판단되었기 때문이다.

(4) 적합성 판단을 위한 제시자료

적합성 판단을 위한 자료로는 결과문헌의 전문을 제시하였다. 그 이유는 첫째, 적합성 판단을 위해 제시할 수 있는 자료로 표제, 목차, 소제목, 초록 등과 같은 여러 필드가 있으나 476개 문헌의 각 구성양식이 상이하여 동일한 환경하에서의 적합성 판단이 불가능하였기 때문이다.

두번째 이유는 전문을 통한 적합성 판단이, 다른 필드를 통한 적합성 판단보다 정확한 판단을 유도할 수 있기 때문이다. 적합성 판단자들은 편의상 표제를 판단자료로 선호하는 경향이 많으나(Shaw, D., 1995) 사라세빅에 의하면 표제에 의한 적합성 판단은 초록이나 전문이 부가적으로 제공될 때 판단자들이 판단결과를 수정하는 것으로 나타나고 있다. (Saracevic, Tefko, 1969) 따라서 적합성 판단이 각 시스템 효율의 측정에 결정적 요소로 작용하게 되는 본 연구에서는 연구의 취지를 충분히 이해하고 있는 전문가들을 대상으로 전문을 판단자료로 제시하였다.

(5) 적합성 판단을 위한 체크리스트 작성 방법

본 실험에서는 적합성 판단자의 거부감을 최소화하면서 정확한 판단을 유도하고 아울러 시스템별 효율 판별이 가능하도록 각 시스템이 검색해 낸 모든 결과 문헌에 대하여 판정하기보다 랭킹한 최 초 결과는 무시하고 각 시스템의 결과 문헌을 동일하게 상위1위에서 10위까지로 절단하였다.

본 연구는 두 단계로 나뉘어 먼저 통제어, 자연어, 용어 수를 추가한 통제어

색인시스템간 검색효율 비교가 이루어지고 다음 단계에서 가중치 실험이 진행되었으므로 적합성 판단을 위한 체크리스트도 두 단계로 나누어 작성하였다.

첫번째 단계로 자연어, 통제어 및 용어 수 추가 통제어 색인시스템이 각각 검색해 낸 상위 10위까지의 결과문헌을 병합하여 적합성 판정을 위한 제1차 체크리스트를 작성하였다. 두번째 단계는 용어 수를 추가한 통제어 시스템과 4개의 가중치 통제어 시스템이 탐색을 수행한 결과 얻어진 결과문헌을 병합하여 제2차 체크리스트를 작성하였다. 적합성 판단은 동일한 판정자에 의하여 이루어져야 하므로

제1,2 체크리스트의 결과를 병합하여 최종 체크리스트로 확정하였다.

랭킹순위가 적합성 판단에 영향을 미치지 않게 하기 위하여(Eisenberg, M. and Barry, C., 1988) 체크리스트에 제시되는 문헌의 순서는 랭킹 순위별로 제시하지 않고 <표 2>와 같이 각 시스템이 검색해 낸 10위까지의 문헌을 통합한 뒤 중복문헌을 제외한 다음 문헌별 고유번호 순서대로 제시하였다.

2. 2. 7 체크리스트에 의한 시스템 효율 측정 방법

각 색인 시스템의 성능은 검색효율로 측

<표 2> 적합성 판정 체크리스트(예시)

[질문 5] 부가가치세의 납세고지서에 과세기간과 과세표준 및 세액의 산출근거 등 이외에 세액산출의 실질적 근거와 경로, 경위 등을 기재하여야 하는지 여부					
문헌번호	매우 적합	적합	부적합	매우 부적합	Title
029					납세고지서 기재사항 하자의 치유
048					조세신고납부 경우의 부당이득반환청구권
189					상속재산의 평가방법
191					당초의 과세처분과 경정처분의 법률관계
193					납세고지서의 기재사항과 송달
197					부가가치세법상 매입세액공제
199					가산세의 성질과 몇가지 관련 문제
203					조세채권의 소멸시효
204					과세처분 무효확인소송
205					조세채무의 성립과 확정
256					조세 및 임금채권의 우선권과 배당
278					소득세법상의 몇가지 문제
310					항고 소송의 심리
312					항고소송의 제소기간
430					항고소송의 대상인 행정처분
459					감액수정신고에 따른 환급세액의 환급청구

정한다. 보편적으로 검색효율의 측정척도는 재현율과 정확률로 삼고 있으나 본 연구에서는 재현율, 정확률 외에 적합성 분포의 차이, 시스템간 중복되지 않는 적합문헌수의 차이 등 네 가지 측면에서 측정한다.

적합성 분포(Green, Rebecca, 1995)를 정보 시스템 검색효율 평가척도의 하나로 채택한 이유는 개개의 시스템이 전 탐색 과정을 통하여 검색해 낸 각각의 결과문헌에 대한 개별 판정자의 적합여부의 분포도 시스템의 검색효율을 평가할 수 있는 또 다른 기준이 될 수 있기 때문이다.

재현율, 정확률은 각 탐색질문에 대한 결과문헌에 대하여 일정 수준을 보장하고 있는 경우에 판정자가 적합하다고 판단한 것이므로 어떤 경우에 있어서는 실제 상황과 유리될 수도 있어 판정자 마다의 개별 판정 결과도 부가적으로 관찰할 필요가 있다. 또한 통계적인 측면에서 보더라도 각 탐색질문의 수는 대개 수 개 정도로 제한될 수밖에 없어 수 개의 데이터를 근거로 한 검증 결과보다는 데이터의 크기를 크게 확대할 경우 객관적인 신뢰도 측면에서 의미가 강조될 수 있다. 적합성의 분포는 각 탐색질문에 대한 모든 결과문헌에 대하여 각 판정자마다 내린 판정결과를 근거로 한 것이므로 본 연구에서만 보더라도 15개 질문이 아닌 3,375개의 데이터로 증가되어 데이터의 크기가 상대적으로 크게 확대, 통계적인 객관성을 점검하는 하나의 기준이 될 수 있어 본 연구에서 시스템 검색효율 척도의 하나로 본 것이다.

또한 시스템간 중복되지 않는 적합문헌

수의 차이를 검색효율 척도로 삼은 이유로 두가지를 들 수 있다. 첫째, 만약 어떤 시스템이 다른 시스템에서는 검색해내지 못하는 고유한 적합문헌을 검색해 내는 능력을 갖고 있다면 그러한 능력은 경우에 따라서는 오히려 재현율, 정확률보다도 우수한 성능의 지표로 삼을 수 있다는 것이다. 둘째, 지금까지 이루어진 연구결과에 의하면 자연어 시스템과 통제어 시스템은 서로 다른 적합문헌을 검색해 낸다고 밝히고 있어 본 연구에서도 만약 각 시스템이 서로 다른 적합문헌을 검색해 낸다는 사실이 검증된다면 시스템간 검색효율의 차이를 비교하는 측면에서 뿐만 아니라 통합 시스템의 제안 등 효율적인 정보 시스템 운영을 도모할 수 있을 것이라는 가정에 기인한다.

2. 3 자료의 수집

본 연구의 수행에 필요한 자료는 1997. 8. 11 ~ 8. 18 3명의 판정자에게 각각 체크리스트 1부와 각 결과문헌의 전문을 제시하고, 결과문헌에 대한 판정결과를 4등급척도(매우 적합-적합-부적합-매우 부적합) 중 한 곳에만 표기하도록 한 후 1997. 9. 8 ~ 9. 13 그 결과를 수집하였다.

2. 4 자료의 분석

수집된 체크리스트는 SAS System for Windows를 사용하여 분석하였다.

재현율과 정확률, 상호 중복되지 않는 적합문헌수와 비율의 차이는 T-검증 (T-test)

을, 적합성 분포는 χ^2 -검증 (Chi-Square)을 실시하였다. 유의수준(p)은 <0.05로 하였다.

분석은 두 단계로 나누어 수행하였다. 먼저 자연어 색인시스템, 통제어 색인시스템, 용어 수를 추가한 통제어 색인시스템간의 검색효율을 두 시스템씩 각각 비교하는 방법으로 분석하였다. 그 다음 가중치가 검색효율에 영향을 미치는지 여부를 검증하기 위하여 용어 수를 추가한 통제어 시스템에서 원래의 통제어와 추가한 용어의 가중치 값의 비율을 각각 1 대0.25, 1 대0.50, 1대0.75, 1대0.90 등 간격을 달리하면서 각 시스템의 검색효율을 비교하였다. (본 연구는 확장 불리언 모델의 랭킹 알고리즘을 사용하여 검색 결과를 유사 계수가 가장 큰 순위부터 출력하였으므로 각 순위에서의 재현율, 정확률을 이용한 평균

검색성능곡선으로도 각 시스템을 비교할 수 있다. 각 시스템의 성능곡선 비교결과는 가설의 검증 여부와는 직접적 관계가 없으므로 지면 관계상 생략하기로 한다.)

통계적 분석을 위하여 산출한 기본 데이터는 <표 3>에서 <표 6>까지 나타나 있다. <표 3>과 <표 4>는 자연어, 통제어, 용어 수를 추가한 시스템의 데이터이며, 표5와 표6은 가중치 시스템의 데이터이다.

<표 3>의 데이터는 15개 탐색질문에 대하여 자연어, 통제어, 용어 수를 추가한 통제어 시스템이 각각 상위 10위까지 검색한 결과에 의해 산출된 적합문헌총수, 검색문헌수, 검색해낸 적합문헌수이다. 각 시스템의 검색문헌총수(144,146,144)에 대하여 3명의 판정자 판정 결과(x3)가 있게 되므로 세 시스템의

<표 3> 자연어, 통제어, 용어 수 추가 통제어 색인시스템의 검색결과문헌 및 적합문헌수(상위 10위)

A: 검색결과문헌수 B: 적합문헌수

질문번호	적합문헌총수	자연어시스템		통제어시스템		용어 수 추가 통제어시스템	
		A	B	A	B	A	B
1	6	10	3	10	3	10	4
2	8	10	4	10	5	7	6
3	7	10	3	10	5	10	6
4	7	10	4	8	5	10	3
5	5	10	4	9	5	10	4
6	5	10	2	9	4	10	4
7	8	10	2	10	6	10	4
8	6	10	2	10	4	10	2
9	9	10	4	10	3	10	6
10	7	10	4	10	5	9	5
11	9	10	2	10	5	10	4
12	8	7	2	10	4	10	5
13	9	10	5	10	7	10	7
14	4	7	1	10	3	8	3
15	6	10	3	10	6	10	3
합계	104	144	45	146	70	144	66

〈표 4〉 자연어, 통제어, 용어 수 추가 통제어 색인시스템의 평균 재현율, 평균 정확률

구 분	자연어시스템	통제어시스템	용어수 추가 통제어시스템
전체검색문헌수	144	146	144
적합문헌수 (매우적합+적합)	45	70	66
부적합문헌수 (부적합+매우부적합)	59	76	78
평균 재현율	0.43	0.69	0.64
평균 정확률	0.30	0.48	0.49

〈표 5〉 용어 수 추가 통제어와 가중치 부여 통제어 색인 시스템의 검색결과 문헌 및 적합문헌수 (상위10위)

A: 검색결과문헌수 B: 적합문헌수

질문 번호	적합문헌 총수	1: 0.25		1:0.50		1:0.75		1:0.90		1:1	
		A	B	A	B	A	B	A	B	A	B
1	6	10	5	9	4	9	4	9	4	10	4
2	9	10	6	8	6	8	6	8	6	7	6
3	7	7	5	10	6	10	6	10	6	10	6
4	6	8	5	8	4	8	4	8	4	10	3
5	5	9	5	8	5	8	5	8	5	10	4
6	6	10	5	8	5	8	5	8	5	10	4
7	8	10	5	8	3	8	3	8	3	10	4
8	6	10	4	10	4	10	4	10	4	10	2
9	6	10	4	10	4	10	4	10	4	10	6
10	5	10	3	10	4	10	4	10	4	10	5
11	8	10	5	8	6	8	6	8	6	10	4
12	11	9	6	10	5	8	6	10	5	10	5
13	8	10	7	10	7	10	7	10	7	10	7
14	3	10	3	8	3	8	3	8	3	8	3
15	7	10	7	10	7	10	7	10	7	10	3
합계	101	143	75	135	73	133	74	135	73	145	66

적합성 분포 비교시 총 데이터는 1,302개의 데이터가 된다.

〈표 3〉의 데이터 값을 기준으로 산출한 세 시스템의 평균 재현율, 평균 정확률 값은 〈표 4〉와 같다.

〈표 5〉의 데이터는 용어 수를 추가한 통제어 시스템에서 법률조문의 정의를 1로 하고, 법률조문의 설명문에 나타난 용어를 각각 0.25, 0.50, 0.75, 0.90 으로 변화시킨 시스템들

이 각각 상위 10위까지 검색한 결과 산출된 적합문헌총수, 검색문헌수, 검색해낸 적합문헌수이다. 각 시스템의 검색문헌총수에 대하여 3명의 판정자 판정 결과(x3)가 있게 되므로 세 시스템의 적합성 분포 비교시 총 데이터는 2,073개의 데이터가 된다.

〈표 5〉의 데이터 값을 기준으로 산출한 세 시스템의 평균 재현율, 평균 정확률 값은 〈표 6〉과 같다.

〈표 6〉 용어 수 추가 통제어와 기준치 부여 통제어 색인시스템의 평균재현율, 평균 정확률

구 분	1:0.25	1:0.50	1:0.75	1:0.90	1:1
전체검색문헌수	143	135	133	135	145
적합문헌수(매우적합+적합)	75	73	74	73	66
부적합문헌수(부적합+매우부적합)	68	62	59	62	79
평균 재현율	0.76	0.75	0.75	0.75	0.68
평균 정확률	0.53	0.53	0.55	0.54	0.46

2. 5 연구의 범위 및 제한

본 연구는 여러 유형의 2차 법률정보 중 단행본과 학위논문은 제외하고 학술기사를 대상으로 하였다.

학술기사를 색인어의 추출을 위한 자료로 한정하는 이유는 단행본과 학위논문의 경우 '상법학', '채권양도에 관한 연구', '물권법', '부동산등기법' 등으로 이미 표제상에서, 또는 예외없이 갖춰진 목차색인을 통하여 범주화가 용이한데 비하여 학술기사는 다루고자 하는 개념의 추출이 표제수준에서는 불충분하고 전문의 분석을 통해서만 가능하다고 판단되었기 때문이다.

따라서 본 연구는 저작권을 제공한 법률분야 대표저자가 1991년부터 1996년까지 국내에서 발표한 판례평석, 연구논문 등을 대상으로 실험을 위한 2차 법률정보 전문 데이터베이스를 구축하였다. 실험 전문 데이터베이스를 구축함에 있어서 저작권의 문제가 야기될 수 있어 정보 입력량의 크기가 제한되었으나 이러한 문제를 극복하고 객관성을 확보하기 위한 방안으로 질문의 수를 크게 늘렸다.

대상자료의 생산년도를 1991년부터 1996년까지로 제한한 이유는 법률정보는 개·폐가

빈번한 주제분야이므로 현행 법령의 구조와 동일한 색인어를 추출하기 위해서이다.

적합성 판단자로서 현직 대법원 재판연구관들을 선정한 이유는, 본 연구에서 객관적 신뢰성을 확보하기 위하여 특정 법률로 한정하지 않고 전체 법률분야를 대상으로 탐색질문을 작성하였으므로, 이들 전문가 집단이 전체 법률분야의 전문성을 비교적 충실하게 수용할 수 있는 집단으로 판단되었기 때문이다. 기업법무가, 검사, 변호사 등의 경우 특정 주제의 법률전문가로 활동하는 것이 최근의 일반적인 현상이므로 본 연구의 판단자로서 적합하다고 보기 어렵다. 이러한 연구 설계 결과 본 연구는 실제 이용자를 피험자로 했을 경우 변수로 작용될 수 있는 그들의 주제배경 여부(Ingrid, Hsieh-Yee, 1993) 등을 통제하면서 개념어의 자동적 추출과 그 효율측정에 집중할 수 있다.

3. 결 론

본 연구에서는 법률정보의 특성 연구를 통하여 2차 법률정보의 색인 알고리즘을 개발하고 이 알고리즘에 의한 통제어 색인시스템과 자연어 색인시스템, 용어 수 추가 통제

어 색인시스템, 가중치 통제어 시스템을 비교하여 2차 법률정보 전문 데이터베이스에 있어서 통제어 색인시스템의 타당성과 우월성을 입증하려고 하였다. 본 연구를 통하여 얻어진 결론은 다음과 같다.

- (1) 2차 법률정보 전문 데이터베이스에서 통제어 색인시스템은 재현율, 정확률, 적합성의 분포, 자연어 시스템이 검색하지 못한 고유한 적합문헌을 검색하는 능력에 있어서 자연어 색인시스템보다 높은 효율을 나타낸다.
- (2) 2차 법률정보 전문 데이터베이스에서 용어 수 추가 통제어 색인시스템은 통제어 시스템과 비교하여 재현율, 정확률, 적합성의 분포, 통제어 시스템이 검색하지 못한 적합문헌을 검색하는 능력에 있어서 효율의 향상을 가져오지 않는다.
- (3) 2차 법률정보 전문 데이터베이스에서 가중치를 부여한 통제어 색인시스템은 가중치를 부여하지 않은 통제어 색인시스템과 비교하여 재현율, 정확률, 가중치를 부여하지 않은 시스템이 검색하지 못한 적합문헌을 검색하는 능력에 있어서 효율의 향상을 가져오지 않으며, 적합성의 분포에서는 높은 효율을 보인다.

이와 같이 2차 법률정보 전문 데이터베이스에서는 법률정보가 지니고 있는 고유한 특성을 연구하여, 그 특성을 시스템에 내장시킨 통제어 색인시스템이 일반적인 자연어 색인시스템보다 재현율, 정확률, 적합성 분포의 향상을 가져올 수 있으며 자연어 시스템에서

는 검색되지 않는 적합문헌을 검색해 내는 능력도 우월하다는 사실이 밝혀졌다.

이러한 사실은 정보시스템을 단순히 언어학적, 통계학적, 문헌구조적인 방법으로 접근하기 보다는 정보 전문가와 주제 전문가가 인식하고 있는 각 주제 분야의 고유 지식, 즉 인간만이 가질 수 있는 보다 경험적이고, 고전적인 지식을 시스템에 내장시키는 것이 필요하며, 그것이 더욱 본질적인 지능형 정보시스템이 될 수 있다는 점을 시사하고 있는 것이다. 따라서 각 주제 분야의 색인시스템을 구축할 때에는 각 주제 분야의 고유한 학문적 특성을 기초로 하는 것이 필요하다고 할 수 있다.

또한 2차 법률정보의 특성상 본문에 인용된 법률조문이 본문의 내용을 대표하는 색인어의 기능을 수행한다는 사실이 검증됨으로써 전문인력의 확보 없이도 일정 수준 이상의 2차 법률정보 전문 데이터베이스에서의 자동색인이 가능하며, 법률의 제정, 개정 및 폐지시에도 2차 법률정보 전문 데이터베이스의 본문에 인용된 법률조문을 일괄 수정함으로써 색인어의 변경이 가능할 것으로 보인다. 아울러 소급자료에 대한 전문을 입력하지 않은 경우에도 같은 원리로 자동적 색인어 추출이 가능하여 소급자료의 경우 색인어만으로 전문자료와 통합 데이터베이스의 구축과 활용이 가능할 것이다.

본 연구의 결과는 다음과 같은 측면에 기여할 수 있을 것으로 기대된다. 첫째, 본 연구에 의해 2차 법률정보 전문 데이터베이스의 색인어가 자동 추출됨으로써 시소러스의 활용에 도움이 될 것이다. 둘째, 외국 법률조

문이 국역화 될 수 있다면 본 연구의 결과에 의하여 자동 추출된 색인어를 근거로 국내외

통합 2차 법률정보 전문 데이터베이스의 운영이 용이하게 될 것이다.

참 고 문 헌

- 권기원, 노정란. 1998. 2차 법률정보 전문데이터베이스 구축을 위한 기초 연구. *한국문헌정보학회지* 32(3):281-296.
- 유재욱. 1996. "적합판단 영향 요인에 관한 이론적 고찰." *情報管理學會誌* 13(2):143-163.
- 정영미. 1987. 정보검색론. 서울: 정음사.
- Davidson, D. 1978. "The Effects of Individual Difference of Cognitive Style on Judgement of Document Relevance." *Journal of the American Society for Information Science* 25(1):67-80.
- Dutta, S. and Sinha, P. K. 1984. "Pragmatic Approach to Subject Indexing : A New Concept." *Journal of the American Society for Information Science* 35(6): 325-331.
- Eisenberg, M. and Barry, C. 1988. "Order effects: A Study of The Possible Influence of Presentation Order on User Judgements of Document Relevance." *Journal of the American Society for Information Science* 39(5):293-300.
- Eisenberg, M. and Schamber, L. 1988. "Relevance: The Search for a Definition." *ASIA Mid-Year Proceedings* 25:164-168.
- Green, Rebecca. 1995. "Topical Relevance Relationships. I .Why Topic Matching Falls." *Journal of the American Society for Information Science* 46(9):646-653.
- Ingrid, Hsieh-Yee. 1993. "Effects 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):161-174.
- Jung-Soon Ro. 1985. An Evaluation of the Capability of Ranking Algorithm to Improving the Effectiveness of Full Text Retrieval. Doctoral Dissertation, University of Indiana.
- Lancaster, F. W. 1979. *Information Retrieval Systems: Characteristics, Testing and Evaluation*, 2nd ed. New York: John Wiley & Sons, Inc.
- Myke, G. 1995. "Understanding Performance in Information Systems:

- Blending Relevance and Competence." *Journal of the American Society for Information Science* 46(6):446-460.
- Saracevic, Tefko. 1969. "Comparative Effects of Titles, Abstracts, and FullTextsonRelevance Judgements." *Proceedings of ASIS* 6:293-299 .
- Shaw, D. 1995. "Bibliographic Database Searching by Graduate-Students in Language and Literature-Search Strategies, System Interfaces, and Relevance Judgements." *Library and Information Science Research* 17(4):327-345.
- Shaw, W. M.Jr. 1990. "Subject Indexing and Citation Indexing-PART II: An Evaluation and Comparison." *Information Processing & Management* 26(6):705-718.
- Su, Louise T. 1992. "Evaluation Measures for Interactive Information Retrieval." *Information Processing & Management* 28(4):503-516.
- Sullivan, M. V., Borgman, C. L. and Wippern, D. 1990. "End-Users, Mediated Searchers, and Front-End Assistance Programs on Dialog: A Comparison of Learning, Performance, and Satisfaction." *Journal of the American Society for Information Science* 41(4):27-42.
- Swanson, Don R. 1986. "Subjective versus Objective Relevance in Bibliographic Retrieval Systems." *Library Quarterly* 56:389-398.
- Tenopir, C. 1984. "Full-text Databases." *Annual Review of Information Science and Technology* 19: 215-246.
- Tenopir, Carol. 1984. *Retrieval Performance in a Full Text Journal ArticleDatabase*. Doctoral Dissertation, University of Illinois.
- Walker, G. 1988. "The Search Performance of End-User." In: *National Online Meeting Preceedings-1988*, Medford:Learned Information Inc.