효율적인 인용분석을 위한 한국 SCI 시스템의 개발
(Development of a Korea SCI System for Efficient Citation Analysis)

이 계 준 † 조 현 양 †† 최 재 황 ††† 윤 희 준 ††††
(KyeJun Lee) (Hyun-Yang Cho) (Jae Hwang Choi) (HeeJun Yoon)

요 약 정보가 생성되어 소멸되기까지의 과정에서 새로운 정보를 만들어내기 위해서 어떠한 형태로든 서비스되고 있는 정보를 수집하여 이것을 참고하고 인용하여 새로운 정보를 생성하게 된다. 본 논문에서는 인용한 문헌과 인용된 문헌 사이의 관계를 정의하고 관련 정보를 저장하여 서비스에 반영하는 KSCI (Korea Science Citation Index)을 구축하였다.

KSCI시스템은 현재 ISI사에서 서비스하고 있는 SCI(Science Citation Index)가 가지는 문제점인 순환적인 검색을 해결하기 위하여 패스인코딩(Path Encoding) 색인 기법을 사용하여 본 시스템을 구축함으로서 저장의 측면에서 약 8.98%의 효율과 검색의 측면에서 약 40% 정도의 결과를 얻었다.

키워드 : 인용색인, 패스인코딩, XML

Abstract In order to produce information the author usually reference other authors' work. A citation index leads users to papers by citations. Citations lead the user to desired information. In this paper, KSCI(Korea Science Citation Index) which defines the relationships between citing documents and cited documents has been constructed.

KSCI System is to solve problems for recursive retrieval in ISI's SCI(Science Citation Index) Path Encoding Indexing technique was used to solve the problems.

From the analysis of data, this system has efficiency about 8.98% in the aspect of data storage. In the aspect of retrieval, there was efficiency between citing documents and cited documents, especially there was over 40% of efficiency in the retrieval of cited documents. It is concluded that suggested KSCI system will provide efficient storage and retrieval system.

Key words : Cited Index, Path Encoding, XML, KSCI(Korea Science Citation Index)

1. 서 론

정보는 생성, 축적, 서비스, 인용, 소멸 등의 단계적인 과정의 생명(life cycle)을 가지게 된다. 이러한 과정 중에서 모든 정보는 인용 또는 참고라는 과정을 통해 새로운 정보를 생성함으로서 그 생명을 유지해 나간다. 과거의 많은 시스템들은 인용과 참고라는 중요한 하나의 과정을 논리적으로 표현하지 못하고 내용 중심의 시스템을 구축하여 서비스하고 있었다. 그러나 인용문헌 사이의 관계를 정의하고, 구조와 논리 정보를 저장하여 서비스에 반영하면 사용자들에게 능동적인 서비스를 제공할 수 있게 된다. 또한, 이러한 과정에서 얻어지는 통계정보를 분석하여 연구 동향을 파악하고, 논문과 저자들에 대한 평가를 위한 통계지표정보 제공이 가능해 진다. 이와 같은 것을 포함하는 것은 계량서지학(Bibliometrics)의 한 영역인 서지분석 기법의 인용색인(Citation Index)이다.

인용색인은 기존 주제색인이 가지고 있는 원하는 논문을 검색 후에 참고문헌에 수록된 논문을 검색하기 위해서는 처음부터 같은 방법으로 검색을 수행해야 한다는 문제점들을 해결하고자 인용정보를 그대로 색인정보로서 사용하는 방법이며, 현재 ISI사의 SCI(Science Citation Index)논문들 중 상위 15%를 대상으로 인용정보와 통계정보를 제공하기 위해 인용색인이 활용되고 있다.

ISI사의 SCI는 인용색인, 소스색인, 기관명색인 등으로 각각의 색인으로 구성되어 있다. 각 색인에 따른 검

† 정 회 원 : 한국과학기술정보연구원 연구원
kjlee@kisti.re.kr
†† 비 회 원 : 경기대학교 인문학부 교수
hycho@kyonggi.ac.kr
††† 비 회 원 : 경북대학교 문헌정보학과 교수
choi@knu.ac.kr
†††† 비 회 원 : 한국과학기술정보연구원 연구원
k2@kisti.re.kr
논문접수 : 2002년 4월 30일
심사완료 : 2004년 2월 2일

색은 우선, 검색 대상이 되는 색인을 정한 다음에 검색어를 입력하여 색인정보를 검색한다. 그리고 검색 결과 내에서 원하는 논문을 선택하고 마지막으로 인용된 논문정보를 선택하여 볼 수 있게 되며, 계속해서 인용된 논문을 볼 수 있게 된다. 이와 같이 각각 구성된 색인을 순환적으로 검색하면서 서비스가 이루어진다.

SCI는 검색 대상에 따라 각각 색인이 생성되기 때문에 많은 구성요소의 중복과 검색 시에 각 색인정보를 순환적으로 검색하기 때문에 검색 효율이 떨어진다. 즉, 색인이 저자명에 의존하여 연결되어 있어 구성요소들 간에 약한 연결 관계를 가지게 되어 검색 시에 많은 테이블 접근을 필요로 하며, 많은 색인 공간이 필요한 단점이 있다.

따라서 본 논문에서는 인용한 문헌과 인용된 문헌 사이의 구성요소들 간의 관계 정의를 통해 구성요소들의 중복을 제거하고, 상호 인용관계에서 나타나는 구조정보와 논리정보 재구성을 통해 강력한 연결 관계 정의를 위해 패스인코딩 색인 방법을 적용하여 효율적인 저장 및 검색이 용이한 KSCI(Korea Science Citation Index) 시스템을 구축하였다.

2. 인용문헌정보 모델링

SCI의 구성과 검색 방법을 분석하여 문제점을 파악하고, 이를 해결하기 위해 인용문헌정보에 대한 정의를 하였으며, 인용문헌정보 모델링을 통해 정의된 내용을 기반으로 KSCI의 색인정보와 패스인코딩 기법을 적용한 코드정보를 생성한 후에 모든 정보 표현과 확장성을 고려하여 XML 형태의 문서를 작성하였다.

2.1 SCI의 구성과 검색 방법

2.1.1 전체구성

SCI 시스템은 전거정보, 색인정보, 논문정보로 크게 구성되어 있으며, 검색을 위하여 인용색인, 소스색인, 기관명색인을 통해 검색 대상에 따라 각각 색인 정보를 생성하고 있다. 다음은 각 색인에 따른 구성요소와 기능이다.

2.1.2 검색 방법

SCI 검색 방법 중에서 저자가 쓴 논문정보에 대한 검색은 저자전거에서 저자명으로 유일한 저자를 검색하고, 이 저자명으로 소스색인에서 인용한 저자명을 검색하여

표 1 SCI 색인방법과 구성요소

구분	구성 요소	기능
인용문헌색인	[인용된문헌] 인용된저자, 발행년도, 학술지명, 권수, 시작면수	논문의 인용에 따른 정보를 얻기 위한 색인 (저자전거 포함)
	[인용한문헌] 인용한저자, 학술지명, 권수, 시작면수, 발행년도	
소스색인	인용한저자, 공저자, 본문언어, 논문제목, 학술지명, 권수, 시작면수, 발행년도, 참고문헌수, 호수	논문의 정보를 얻기 위한 색인
기관명색인	기관명, 저자명, 학술지면, 권수, 시작면수, 발행년도	기관에 포함된 저자의 논문들에 대한 정보를 얻기 위한 색인

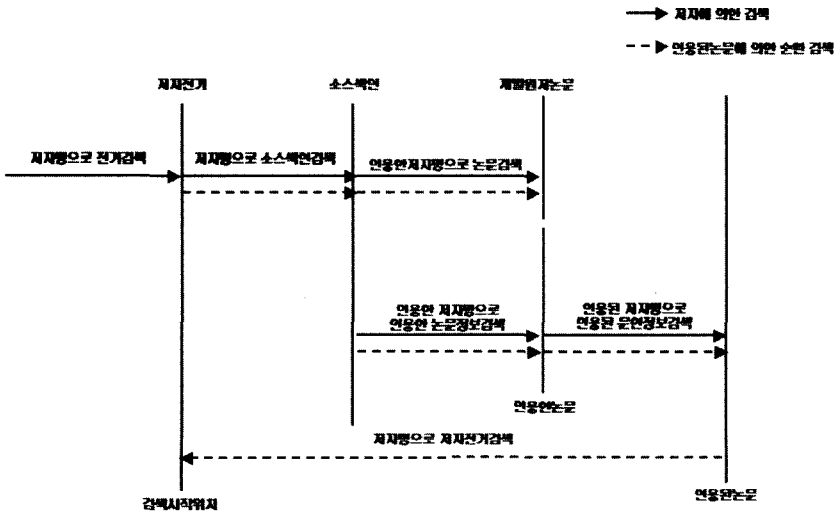


그림 1 저자명에 의한 논문 정보 검색

논문정보를 얻은 다음에 개별원저논문 정보를 검색하거나 논문에 인용된 참고논문을 검색하게 된다. 참고논문을 검색하기 위해서는 인용한 저자명에 의해 인용색인을 검색하게 되며, 인용한 저자명으로 인용한 논문을 검색한 다음 이에 해당하는 인용된 논문의 정보를 얻게 된다. 또한, 인용된 논문의 정보를 얻기 위해서는 저자전거를 검색하고 소스색인을 검색하는 순서로 순환적인 검색이 이루어진다. (논문명과 기관명에 의한 검색 또한 같은 방법으로 이루어진다.)

SCI는 각각의 색인정보를 생성하기 위한 구성요소들 간의 중복을 허용하고 있으며, 검색 시에 모든 색인정보를 순차적으로 순회하면서 검색해야 하기 때문에 검색 효율이 떨어지는 문제들을 가지고 있다. 반면에, KSCI에서는 SCI의 문제들을 효과적으로 해결할 수 있는 색인과 저장이 가능하도록 모델링을 하였다.

2.2 KSCI 구성요소

KSCI 시스템을 구성하는 구성요소를 추출하기 위해 국내 180여 학회의 과학기술 학술지, 논문지, 영문지를 대상으로 분석하여 인용한 논문과 인용된 논문의 내용을 표현할 수 있는 최소한의 정보만을 기본으로 다음과 같이 추출하였다.

표 2 KSCI 구성요소

구분	구성요소
저널정보	저널명, ISSN, 저널타입, 학회명, 권, 호, 발행년도
저자정보	저자명, 소속기관, 공저자명, 공저자소속기관명, 이메일
논문정보	논문제목, 초록, 본문언어, 페이지, 키워드, URL,
인용문헌정보	저자명, 공저자명, 논문제목, 저널명, 권, 호, 페이지, 발행년도

이렇게 추출된 정보를 가지고 한번 검색으로 참고 논문까지 모두 검색이 가능한 SCI가 가지고 있는 가장 큰 장점을 수용하고, 검색 및 저장 효율을 높일 수 있도록 모델링을 하였다.

2.3 인용문헌정보 모델링

인용문헌정보를 나타내기 위해 구성요소와 문헌들 간의 인용관계를 표현할 수 있는 정보가 필요하다. 논문간의 인용관계를 나타내기 위해 우선 인용한 논문과 인용된 논문들 간의 복잡한 관계를 정형화하여 효율적인 저장과 검색을 지원하기 위하여 완전한 트리 구조에 의한 재구성은 다음과 같다.

완전 트리 구조에 의한 재구성은 루트(KSCI)로부터 시작하여 저널들을 자식으로 가지며, 저널은 논문들을 자식으로 가지며, 논문은 논문을 작성한 저자들의 정보와 논문에 인용정보를 자식으로 가지게 과정은 다음과 같다.

첫째, 저널정보는 저널들은 1부터 N까지의 유일한 값을 가진다. 둘째, 논문정보는 부모인 저널정보를 상속받아 1부터 N까지의 유일한 값을 갖는다. 셋째, 저자정보는 부모인 논문정보를 상속받아 1부터 N까지의 유일한 정보를 가지며, 형제 관계를 가지는 인용정보와 구분하기 위한 구분정보를 포함하게 된다. 또한 자식으로 저자리스트 정보를 가지는데 저자리스트는 논문을 작성한 저자와 공저자들의 정보가 된다. 넷째, 인용정보는 부모인 논문정보를 상속받아 논문에 명시된 논문들을 리스트로 가지게 되며, 인용정보리스트에 해당하는 논문은 논문정보에 등록이 되기 때문에 인용정보리스트는 논문정보의 유일한 코드 값들로 대체되어 구성된다.

이와 같이 인용문헌정보를 부모와 자식과 같은 상속관계를 고려하여 완전한 트리 구조 형태로 재구성하였으며, 구성요소들 간의 중복을 제거하고 인용관계에서 나타났던 복잡한 관계를 간단하게 정형화하였다.

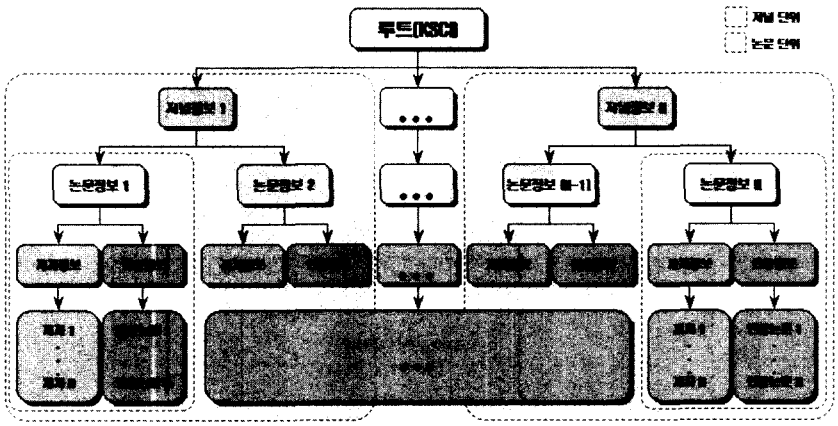


그림 2 트리 구조에 의한 인용정보의 재구성

3. KSCI 시스템 설계 및 구축

3.1 코드정보와 색인정보 설계

인용정보의 저장에 대한 효율성을 높이기 위해 트리 구조에 의해 재구성된 인용정보를 기반으로 색인정보를 생성하고, 색인정보를 트리 위치정보에 매핑하기 위해 패스인코딩 방식에 의한 유일한 값의 코드정보를 생성하였다.

3.1.1 코드정보

코드정보는 모델링을 통해 만들어진 이진트리에서 색인정보, 소스정보, 리스트정보를 정확하게 검색하기 위해 유일한 값을 나타낼 수 있도록 패스인코딩 기법을 적용하였다. 코드생성은 부모로부터 상속받은 코드와 자식의 코드와의 결합으로 구성되어 지며 이렇게 만들어진 코드 또한 유일한 값을 가지게 된다. 코드 생성규칙과 트리에 대한 위치 정보 매핑은 다음과 같다.

예시) 논문정보 코드 = 저널코드(부모코드) + 논문코드 (자신의 레벨에서 구별 코드)로 구성됨

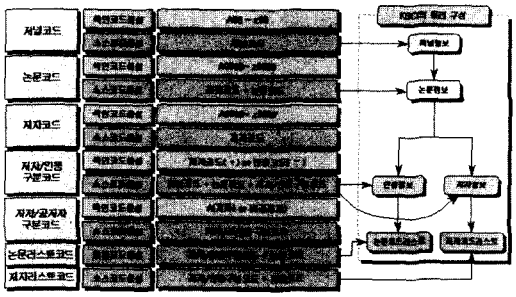


그림 3 코드정보 생성 규칙과 트리에 대한 위치정보 매핑

3.1.2 색인정보

모델링 통해 정립된 정보와 위치표현을 위한 코드정보를 정확히 저장하기 위하여 색인 정보를 생성하였다. 색인정보는 인용문헌정보들 간의 논리적인 관계 정의 및 위치 정보를 저장하기 위한 것이며, 소스정보는 실제 논문정보를 포함하게 된다. 또한, 저널, 논문, 저자에 대한 전거 정보를 생성하여 정보를 유일하게 나타낼 수 있는 정보와 논문에 인용된 문헌정보를 나타내기 위한 리스트정보가 필요한데 이러한 색인정보를 구분하여 구성하면 다음과 같다.

3.2 시스템 전체 구성 설계

인용문헌의 내용정보, 구조정보, 논리정보를 추출한 다음 데이타베이스에 저장하고 검색할 수 있도록 하기 위해 전체적인 시스템을 설계하였다.

3.2.1 시스템 설계

인용문헌정보를 데이타베이스화하기 위해 유일한 색인정보를 참조하여 내용정보, 구조정보, 논리정보를 생

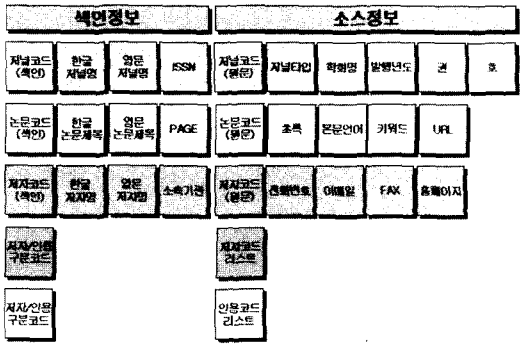


그림 4 KSCI 색인정보 설계

성하여 저널정보, 논문정보, 저자정보, 저자/공저자리스트, 인용문헌리스트의 구축을 위한 전체구성은 다음과 같다.

4. KSCI 시스템 구축

KSCI 시스템을 구축하기 위해 인용문헌정보를 분석하고, 색인정보와 코드정보 설계, 그리고 시스템 설계 등의 과정을 통해 얻은 결과를 가지고 인용문헌정보를 데이타베이스에 저장하는 시스템을 구축하였다.

4.1 시스템 구축환경

KSCI 시스템을 구축하기 위해서 서버는 IBM P860 이며 운영체제는 AIX 2.7.x 버전을 기반으로 데이타를 저장하기 위해 Oracle 8.1.7 버젼을 사용하였다. 시스템 개발 언어는 JAVA 1.3을 사용하였고, 저장시스템과 데이타베이스의 연결을 위하여 JDBC 2.0을 이용하였으며, XML 파서 및 DOM 트리를 위해서는 XML4j 3.1.1 (xerces) JAVA Parser를 이용하여 본 시스템을 구축하였다.

4.2 저장 시스템 구현

KSCI 시스템은 인용정보 입력자로부터 입력받은 정보의 유효성을 검사하여 유효한 정보만을 선별하여 색인정보 관리기로 보낸다. 색인정보 관리기는 저장될 정보에 대한 색인정보의 유무를 알기 위해 저널색인, 논문색인, 저자색인 테이블을 검색하여 얻은 색인 코드정보를 추가하여 저장을 위한 완전한 정보를 생성하게 되다. 이때, 색인정보에 검색한 정보가 없을 경우 신규로 색인을 생성하는 과정을 거치게 된다. XML문서를 작성하기 위하여 원문을 원문정보 분석기를 통해 분석하여 XML Tag 변환기로 보내진다. XML Tag 변환기에서는 표준 DTD를 참조하여 XML문서를 작성한다. 이렇게 만들어진 XML문서는 Oracle 데이타베이스에 저장이 된다. 그리고 논문에 인용된 논문이 이미 구축되어 있을 경우라면, 새로 구축할 필요 없이 정보를 참조할 수 있는 코드값만을 가지도록 구축하였다. 이러한 과정에 필요한 것

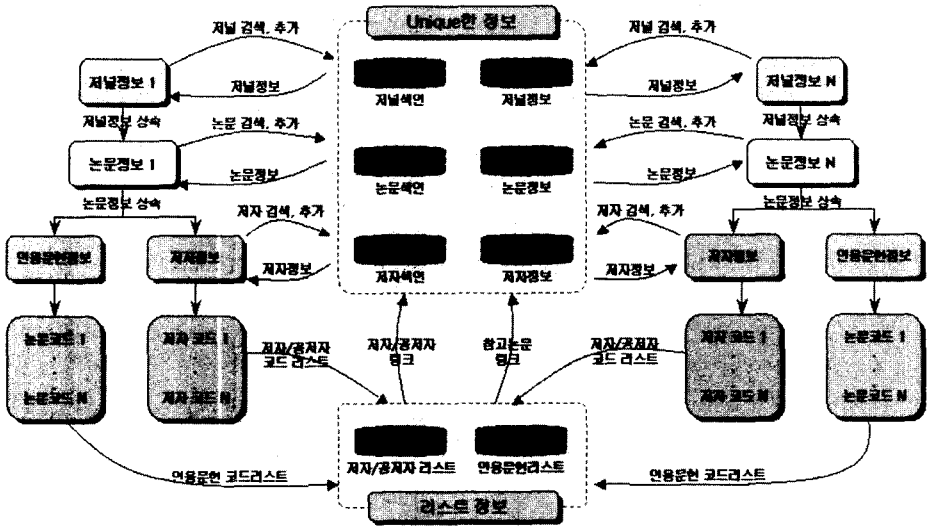


그림 5 데이타 참조에 의한 전체 구성

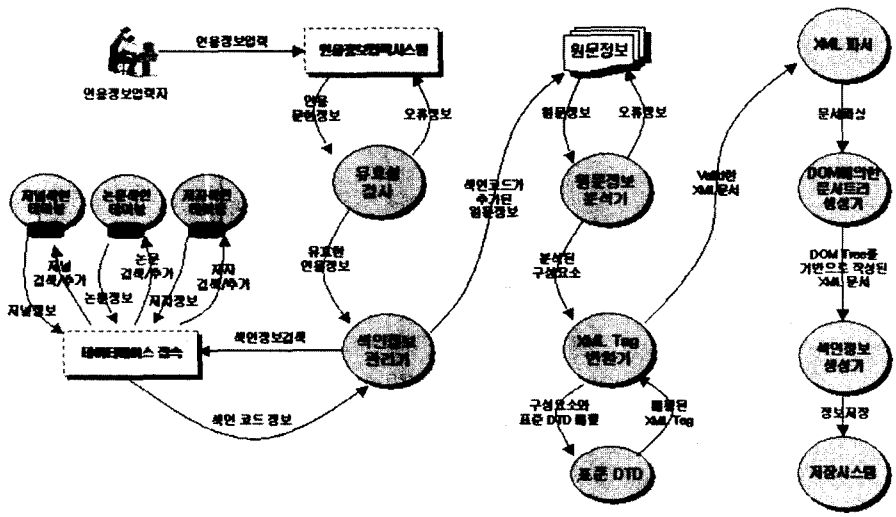


그림 6 KSCI 시스템 구현

이 기구축된 정보에 대한 XML 문서 과정을 통해 원하는 정보를 추출하게 된다. 이러한 과정을 나타낸 것은 다음과 같다.

4.3 검색 시스템 구현

KSCI에서 검색을 살펴보면, 저자가 쓴 논문에 대한 정보를 검색하기 위해서는 저자색인에서 저자명에 의해 유일한 저자를 검색한다. 그리고 저자원문에서 저자원문코드를 검색한 다음 저자원문코드를 분석하여 분석된 결과를 가지고 저널정보, 논문정보(원문정보포함) 정보를 검색하게 된다. 인용정보검색 또한 저자원문코드를 분석한 결과를 가지고 검색이 이루어진다.

예) 저자원문코드가 A01A0001+AA0001일 경우 코드 분석 결과는 다음과 같다.

표 3 저자소스 코드의 분석과 검색

코드분석	코드에 의한 검색
저널코드 : A01 논문코드 : A0001 저자리스트 : + 인용정보리스트 구분코드 : - 저자 구분코드 : A 공저자 구분코드 : U 저자코드 : A0001	A01 : 저널색인, 저널원문검색 A0001 : 논문색인, 논문원문검색 A01A0001+U : 공저자 검색 A01A0001- : 인용된 논문 검색

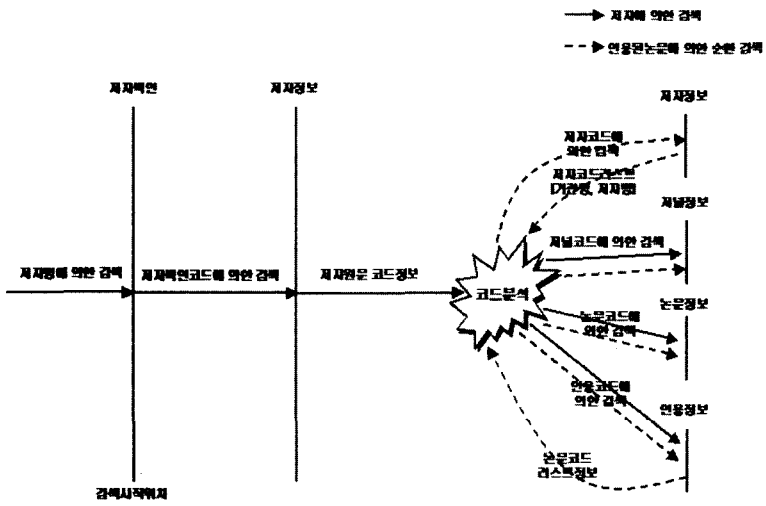


그림 7 저자명에 의한 논문 정보 검색(KSCI 방법)

그리고 코드 분석 결과를 가지고 모든 정보를 한번에 검색을 하며, 인용된 논문의 검색은 인용정보리스트 코드에 의해 검색되어지는 논문코드로서 검색이 가능하게 된다. 검색 과정은 다음과 같다. (논문명과 기관명에 의한 검색 또한 같은 방법으로 이루어진다.)

5. 시스템 성능 평가

인용문헌정보를 대상으로 데이타베이스 구축 후의 데이타 크기에 의한 실험결과와 검색을 통한 실험결과로 분류하여 각각을 SCI 방법과 비교하는 성능평가를 하였다.

5.1 실험 대상 소스데이타

인용문헌정보를 대상으로 데이타 크기와 검색에 의한 성능평가를 위해 사용되어지는 소스데이타는 다음과 같으며, 하나의 레코드를 하나의 건수로 한 결과이다.

표 4 실험 대상 데이타

구 분	내 용(단위 : 건)
저널 수	1,667
논문 수	142,680
저자 수	54,190
기관명 수	19.462
인용된 논문	6,250
인용한 논문	56,276

5.2 데이타베이스 구축 결과 비교

인용문헌에 대한 실험 대상 데이타를 기준으로 SCI방법과 KSCI 방법을 통해 데이타베이스를 구축한 결과에 대해 비교하였으며, 본 논문에서 제안한 KSCI 시스템을 통해 인용문헌정보를 저장하였을 때에 어느 정도 효율이 있는가에 대해 실험을 하였다.

표 5 데이타베이스 구축 결과

구 분	용 량(단위 : KB)	
	SCI 방법	KSCI 방법
기관명색인	44,544	720
소스색인	179,920	812,705
인용정보색인	1,988	2,167
	17,895	2,251
원문	715,264	44,003
합계	잘못된 계산식	잘못된 계산식

SCI 방법을 통했을 때는 959,611KB의 저장 공간이 필요한 반면에 KSCI 방법을 통했을 경우에는 861,864의 공간만을 필요로 한다. 즉, 97,747KB의 중복을 줄였으며 약 8.98%의 절감 효과를 가져왔다.

5.3 검색에 의한 실험 결과

검색에 의한 실험은 검색 대상 데이타베이스 크기를 고정해 놓고 질의어 수만 증가시키는 실험과 질의어 수를 고정시켜 놓고 검색 대상 데이타베이스의 크기를 증가시키는 실험으로 크게 두 가지로 나누어서 수행하였다. 모든 검색 실험은 SCI방법과 KSCI방법으로 나누어서 비교하였다.

5.3.1 데이타베이스 크기 고정과 질의어 수의 증가에 관한 실험

첫 번째 실험은 검색에 대한 성능 평가를 위해 질의어 수를 저자명, 논문명, 기관명 각각에 대해 1개, 10개, 100개, 1000개로 증가시키면서 전체를 대상으로 검색을 수행하였으며, 검색은 인용한 논문과 인용된 논문의 검색으로 나눠서 수행하였다. 실험 결과는 각 질의어 수에 의해 검색되는데 걸리는 시간(초)을 기준으로 하여 SCI 방법과 KSCI 방법으로 검색한 결과는 그림 8~그림 13과 같다.

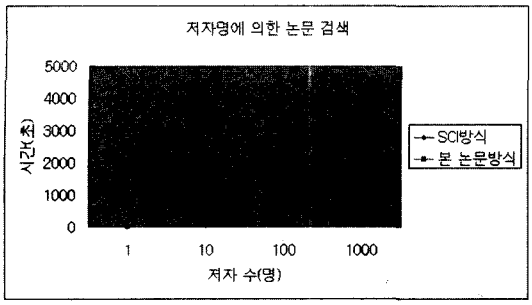


그림 8 저자명에 의한 논문 검색

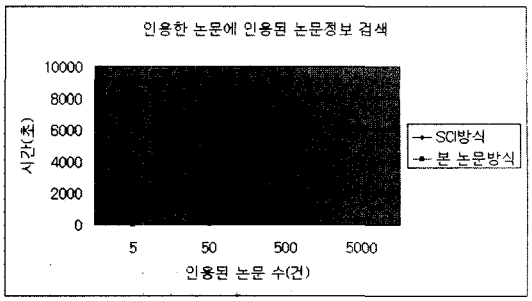


그림 11 인용된 논문 검색

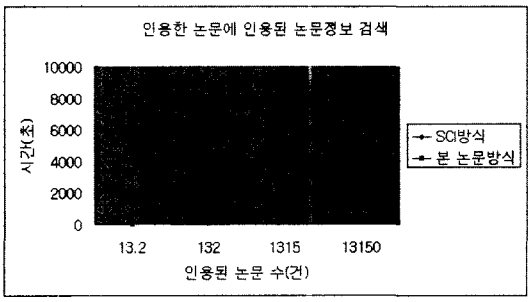


그림 9 인용된 논문 검색

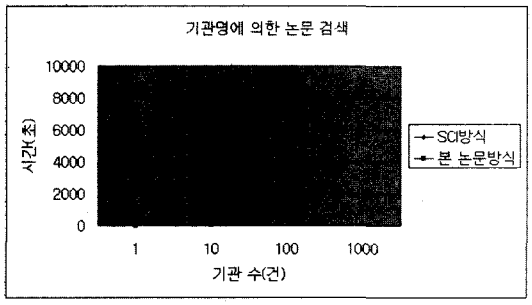


그림 12 기관명에 의한 논문 검색

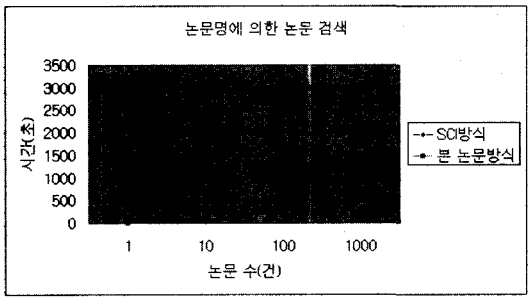


그림 10 논문명에 의한 논문 검색

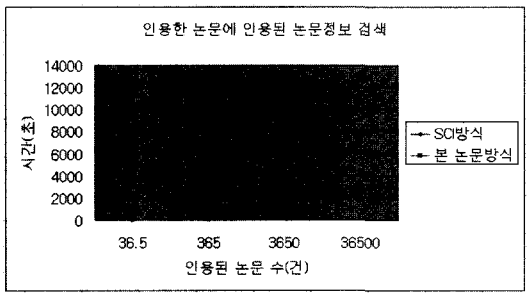


그림 13 인용된 논문 검색

위의 실험내용 결과를 보면, 저자정보, 논문정보, 기관정보 등에 의한 정보의 검색에서는 인용한 논문 검색과 인용된 논문 검색을 나누어서 실험을 하였으며, 각각의 검색은 그 대상 정보의 범위를 설정하여 실험을 하였다. 검색결과를 토대로 분석을 종합해 얻은 결론은 다음과 같다.

표 6 검색 결과에 대한 비교

	인용한 논문		인용된 논문	
	검색(초)	검색(%)	검색(초)	검색(%)
	0.222	14.6	0.316	43.78
	0.066	2.16	0.777	45.51
	0.402	34.04	0.154	42.64
	논문 1건을 검색하는데 걸리는 시간			

5.3.2 데이타베이스 크기 증가와 질의어 수의 고정에 따른 실험

두 번째 실험은 검색 대상이 되는 질의어수를 고정시켜 놓고, 데이타베이스의 크기의 증가에 따라 단계별로 나누어서 검색 결과를 비교해 보기 위한 실험을 하였다. 검색 실험은 실험 대상이 되는 데이타베이스 크기를 결정한 후에 질의를 위해 100건의 샘플데이타를 추출하여 실험을 하여 얻을 결과들이다. 검색 대상 데이타의 크기에 따른 검색 성능을 평가하기 위하여 전체 대상이 되는 데이타를 3단계에 의해 대상 데이타를 선정한 것은 다음과 같다.

각 단계를 100건의 샘플데이타를 가지고 SCI 방법과 KSCI 방법을 통해 저자정보에 의한 검색, 논문정보에

표 7 실험 대상 데이타 구분

구분	1단계	2단계	3단계
저널수	556	1,100	1,667
논문수	47,568	91,325	142,680
저자수	18,007	36,926	54,190
기관명수	6,492	13,032	19,462
인용한 문헌 수	1,780	4,148	6,250
인용된 문헌 수	17,855	38,571	56,276

의한 검색, 기관정보에 의한 검색을 수행한 결과는 다음과 같다.

5.4 검색 결과 분석

본 논문에서 제안한 KSCI 방식에 의해 구축한 시스템을 SCI 방식을 사용하여 구축한 시스템과 비교했을 때 다음과 같은 결과를 얻었다. 첫째, 어떤 저자 1명이 쓴 논문(인용한 논문) 모두를 검색하는 데에는 저자명에 의한 검색은 14.6%가 감소하였으며, 저자가 쓴 논문을 기준으로 인용된 논문을 검색하는 데에는 약 43.78%가 감소하였다. 둘째, 논문명에 의해 특정 논문(인용한 논문)을 검색하는 데에는 2.16%가 감소하였으며, 특정 논문을 기준으로 인용된 논문을 검색하는 데에는 45.51%가 감소하였다. 셋째, 어떤 기관에 소속된 저자들이 쓴 논문(인용한 논문) 모두를 검색하는 데에는 34.04%가 감소하였으며, 기관에 소속된 저자들이 쓴 논문을 기준으로 인용된 논문을 검색하는 데에는 42.06%가 감소하였다. 넷째, 검색 대상의 크기와는 관계없이 인용문헌정보를 검색하는데 있어서 일정한 수치로의 증가를 가져왔으며, SCI방식 보다 KSCI 방식에서 검색 성능이 높았다.

6. 결 론

정보의 생성에서 소멸까지의 과정 중에서 인용이라는 정보가 가지는 하나의 중요한 과정을 기반으로 얻을 수 있는 논리적인 정보의 데이타베이스를 통해 사용자들에게 능동적인 서비스를 제공하기 위하여 인용문헌이 가지는 내용정보, 구조정보, 논리정보를 색인과정을 통해 데이타베이스에 저장하고 효율적인 검색의 제공이 가능한 KSCI 시스템을 구축하였다.

KSCI 시스템은 SCI 방법을 통했을 때 보다 데이타베이스를 구축하였을 경우에 약 8.98%의 저장 공간에 대한 효율이 있었으며, 인용한 논문 검색에 있어서는 각각 저자명, 논문명, 기관명 각각의 검색에서 절감 효과가 있었으며, 검색된 인용한 논문을 기준으로 인용된 논문을 검색할 경우에는 약 40% 이상의 많은 검색 절감 효율이 있었다. 또한 질의어를 고정시켜 놓고 데이타베이스의 크기를 증가하면서 수행한 실험에서도 검색 효율이 있었다.

따라서 KSCI 시스템은 패스인코딩 코드생성을 기반으로 색인정보를 생성하고 저장하였으며, 인용한 문헌과 인용된 문헌의 구성요소들 간의 중복을 제거하였다. 그리고 내용정보, 논리정보, 구조정보 등의 정보를 손실없이 저장하고 검색에 이용함으로서 인용된문헌을 검색할 때에 코드 정보를 활용할 수 있게 되었으며, 검색 대상에 따른 검색 위치 선정이 가능하여 검색 효율을 강화하였다. 표준 DTD의 작성을 통한 인용문헌정보에 대한 수정, 추가가 용이한 확장성을 제공하며, 상호인용관계에서 나타나는 복잡한 논리 관계를 정의하여 강력한 연결 관계를 제공하여 인용문헌정보에 대한 효율적인 검색을 가능하게 하였다.

참 고 문 헌

[1] Glanzel, Wolfgang and Urs Schoepflin. 1999. "A bibliometric study of reference literature in the sciences and social sciences," Information Processing & Management, 35(1):31-44.

[2] Giles, C. Lee, Kurt D. Bollacker, and Steve Lawrence. 1998. "CiteSeer: An automatic citation indexing system," Proceedings of the third ACM Conference on Digital libraries, 89-98.

[3] Zhang, Haiqi and Yuhua Zhang. 1997. "Scientometric study on research performance in China," Information Processing & Management, 33(1):81-89.

[4] Observation des Sciences et des Technique (OST), White Paper on Science and Technology, 1994.

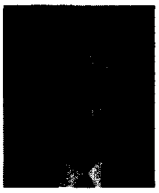
[5] Science and Technology Agency Japanease Government, White Paper on Science and Technology, 1996.

표 8 실험 대상 데이타의 크기에 따른 검색 결과

구분	1단계				2단계				3단계			
	SCI방법		KSCI방법		SCI방법		KSCI방법		SCI방법		KSCI방법	
	Citing	Cited	Citing	Cited	Citing	Cited	Citing	Cited	Citing	Cited	Citing	Cited
저자	2.84	6.79	2.51	3.91	3.62	8.61	3.29	5.13	4.27	10.2	3.85	6.0
논문	2.25	5.91	2.08	3.25	2.84	7.55	2.73	4.26	3.34	8.9	3.19	4.98
기관	6.19	9.46	4.33	5.21	7.88	12.13	5.63	6.82	9.3	14.2	6.59	7.98

[4] 김연희, 김성완, 신판섭, 이재호, 임해철, "문서중심 XML 문서를 위한 데이타 모델", 홍익대학교 컴퓨터 공학과, 2001 가을 정보과학회 학술발표 논문집, 2001.
[5] 김창수, 정희경, "XML 응용 개발환경", 배제대학교 전자계산학과, 정보과학회지 제19권 1호, 2001.
[6] 박상원, 정재목, 정태선, 김형주, "XML과 데이타베이스", 서울대학교, 정보과학회지 제19권 1호, 2001.
[7] 정영선, "국내 학술지 인용색인의 구축 방안에 관한 연구", 이화여자 대학교 대학원, 석사학위논문, 2000.
[8] 연제원, 김상균, 이규철, 나중찬, 김명준, "XML 문서의 효율적 검색 및 변경을 위한 저장 관리기의 설계 및 구현", 충남대학교 컴퓨터공학과, 1999.
[10] 안영선, "학위논문의 XML DTD 설계에 관한 연구", 성균관대학교 대학원 석사학위논문, 1999.
[11] 이종설, 강형일, 손충범, 강승헌, 정연수, 민영수, 박종관, 유재수, "XML 저장관리시스템 설계 및 구현", 충북대학교 컴퓨터 정보통신 연구소, 컴퓨터 정보통신 연구 제7권 2호, 1999.
[12] 유재수, "효과적인 문헌 정보 관리를 위한 데이타언어와 질의처리기", 충북대학교 컴퓨터 정보통신 연구소, 컴퓨터 정보통신 연구 제6권 1호, 1998.
[14] 한국학술진흥재단, "국내 학술지 인용색인 데이타베이스 개발 연구", 연구보고서, 연세대학교, 1998.

이 계 준
2000년 2월 충남대학교 컴퓨터공학과 졸업(석사). 2004년 현재 한국과학기술정보연구원 재직. 관심분야는 정보검색, 색인, XML, Web Services 등



조 현 양
1981년 2월 중앙대학교 도서관학과 졸업(학사). 1985년 5월 Rutgers University (문헌정보학 석사, MLS). 1996년 2월 중앙대학교 대학원 문학박사(정보학전공) 2004년 현재 경기대학교 인문학부 문헌정보학전공 교수. 관심분야는 색인, 시소러스, 정보검색 등

최 재 황
1990년 2월 성균관대학교 문헌정보학과 졸업(학사). 1993년 University of North Texas, School of Library and Information Sciences. MS, in Information Science. 1997년 Florida State University. School of Information Studies Ph. D, in Library and Information Science. 2004년 현재 경북대학교 문헌정보학과 조교수

윤 희 준
1995년 2월 한밭대학교 전산학과 졸업(공학사). 1997년 2월 충남대학교 대학원 컴퓨터공학과 졸업(석사). 2004년 현재 한국과학기술정보연구원 근무. 관심분야는 대용량데이타베이스, 멀티미디어 데이타베이스, 표준화