

특허 DOI에 기반한 특허원문연계체계 구축에 관한 연구

권 오 진*, 노 경 란

한국과학기술정보연구원 정보콘텐츠개발실

A Study on Implementation of Patent Fulltext Linking System based on Patent DOI

Kwon Oh Jin*, Noh Kyung Ran

KISTI, S&T Development Dept.

E-mail : dbajin@kisti.re.kr

요 약

본 연구에서는 국가별 특허문서의 고유식별체계인 특허번호 현황에 대해 살펴보고, 국가별로 다양한 특허번호체계를 하나로 통합할 수 있는 표준 디지털고유식별체계를 기술한다. 그리고 특허정보 디지털식별체계를 이용한 특허원문 연계구축방안에 대해 기술한다. 각국의 특허번호를 이용하여 특허에 대한 간단한 색인정보를 메타데이터형식으로 구축하고, 개별 특허의 원문에 대한 식별자와 위치정보를 중앙에서 관리하여 언제든지 각국의 특허원문으로 연계가 가능한 체계를 제시하고자 한다.

Abstract

This Paper reviews various formats of patent numbers that unique identifier of patent document according to country. And it describes standard DOI model that can unify each country's several patent numbers. Based on this patent DOI, it describes implementation method to link digitalized patent document to patent bibliographic record at a time.

I. 서론

정보통신기술과 인터넷의 발전으로 과거 책자형으로 제작되었던 지적 저작물들이 전자형태로 디지털화되고 있다. 이러한 추세에 따라 특허정보에 대한 디지털화가 활발하게 추진되고 있다. 그러나 특허정보가 네트워크를 통해 유통되고 있음에도 불구하고, 상이한 시스템 및 포맷으로 구성되어 있어 효율적인 접근 및 유통이 이루어지고 있지 않다.

각국의 특허청에서 발행되는 특허공보의 번호는 각국의 특허제도에 따라 표현형태가 다르며, 또한

같은 국가내에서도 공보 발행시기에 따라서 표현 형태가 다르게 표현되는 경우가 있다.

각 국가의 특허제도에 충분히 익숙하지 않은 이용자들은 특허번호를 정확히 알고 있다고 할지라도 원하는 특허를 찾기가 어려운 경우가 빈번하며, 여러 국가의 특허를 동시에 찾고자 할 경우 검색시간이 많이 소요되며, 특히 우선권 특허의 경우 관련특허가 있음에도 불구하고 해당특허가 존재하지 않는 것으로 잘못 인식되는 경우도 있다.

따라서 특허정보를 제공하는 각 기관들은 저마

다 나름대로의 번호형식 기준을 마련하려는 노력을 기울이고 있는 실정이다. 본 연구에서는 각국의 특허번호체계를 기반으로 한 DOI를 개발하고 이 DOI를 이용한 특허원문연계체계를 제시하고자 한다.

제2장에서는 국가별 특허문서의 고유식별체계인 특허번호에 대해 살펴보고, 제3장에서는 국가별로 다양한 특허번호체계를 하나로 통합할 수 있는 디지털 고유식별체계 표준안에 대해 기술한다. 그리고 제4장에서는 특허정보 디지털식별체계를 이용한 특허원문 연계구축방안에 대해 기술한다.

II. 국가별 특허번호체계 현황

1. 한국

한국의 특허번호에는 출원번호, 공개번호, 공고번호, 특허(등록)번호, 심사청구번호 등이 있다. 이들의 상호관계를 알기 위해서는 번호대조표를 이용하거나 DB에서 각종 특허번호를 검색기로 사용한다. 한국특허번호 표현형태를 살펴보면 출원번호, 공개번호, 공고번호의 경우 연도 네자리와 일련번호로 구성되어 있다. 과거 연도자리수는 2자리를 사용하여 표기하였다. 일련번호의 경우 매년 1부터 다시 시작하며, 일련번호의 자릿수가 고정되어 있지 않아 여러 형태로 나타나고 있다. 특히 PCT 출원번호의 경우 일련번호중 앞의 7이 PCT 출원을 의미하며, 98년까지 출원된 특허의 경우 "07"을, 99년 이후 출원된 특허의 경우 "70"을 표기한다. 등록번호는 연도변경에 관계없이 지속적으로 일련번호가 부여된다[1].

2. 일본

한국특허와 마찬가지로 일본특허는 특허상태에 따라 출원번호, 공개번호, 공고번호, 등록번호로 구분한다. 일본특허번호체계는 앞의 3자리가 연도를 의미하고 뒷부분은 특허고유번호로 구성된다. 출원번호와 공개번호의 경우 2000년 이후 서기년도 4자리를 사용하고 뒷부분은 일련번호로 된 고유식별번호를 사용하는데, 일련번호의 자릿수는 정해져 있지 않다. 등록번호 즉 특허공보와 등록실용신안, 실용신안등록의 공보번호는 일본연호가 없으며, 연도가 바뀌더라도 일련번호를 다시 시작

하지 않고, 전년도에 사용했던 일련번호에 이어 번호를 부여한다[2].

표1. 일본특허번호유형

구분	번호유형		사례
공표 (공개) 번호	1999년 이전번호	일본연도- 일련번호	S54-123456 昭62-234 昭62-123456
	2000년 이후번호	일본연도- 일련번호	H12-123456 平1-7345
		서기년도 4자리- 일련번호	2000-123456
공고번호		일본연도- 일련번호	H01-123456 平1-7345
재공표, 공개기보		0+서기년 도2자리-일 련번호	098-12345
등록	특허(실용) 명세서,청 구번호	일련번호	1234567

3. 미국

미국특허번호는 출원번호와 공고번호로 구분된다. 미국의 특허번호체계는 각 유형별 코드와 일련번호로 구성된다. 일련번호의 자릿수는 정해져 있지 않으며 연도가 바뀌더라도 다시 시작하지 않고, 전년도에 사용한 일련번호에 이어서 순차적으로 부여된다. 미국의 특허번호에는 연도를 나타내는 정보가 존재하지 않는다[3].

표2. 미국특허번호유형

구분	번호유형	사례
특허	일련번호	5,146,634
의장특허	D+일련번호	D339,456
식물특허	PP+일련번호	PP8,901
재발행특허	RE+일련번호	RE35,312
방어출원	T+OG volume 기반번호부여	T109,201
법정발명등록	H+일련번호	H1,523

III. 특허정보 디지털고유식별체계 표준안

인터넷환경에서 디지털 콘텐츠의 위치를 발견하

고 접근하는 방법으로 URL이 활용되고 있으나 URL은 인터넷의 단순한 물리적 주소로써 이에 대응되는 디지털 정보원의 위치변경이나 시스템주소 변경이 발생하면 해당 정보자원에 접근하지 못하는 문제점을 안고 있다[4].

전통적인 정보자원에 대한 식별기호체제로 ISBN, ISSN, SICI, 특허번호 등이 존재하듯이 디지털세계에서도 디지털 콘텐츠를 식별해줄 수 있는 관리체계와 식별기호가 필요하다. DOI는 콘텐츠에 대한 유일한, 영구적인 고유식별기호로, URL로 변환되어 인터넷상에 존재하는 해당정보로 접근할 수 있도록 하는 체계이다. DOI는 네트워크상에서 연계서비스를 제공하기 위한 중요한 연계도로 자리매김하고 있다. DOI의 활용가치를 높이고 유용한 서비스를 제공하기 위해서는 DOI 자체만으로 대상 콘텐츠에 대한 최소한의 의미를 제공할 수 있는 식별자 구성체계가 요구된다.

DOI는 '/'로 구분되는 접두부와 접미부로 구성된다. 접두부는 디렉토리코드와 등록자코드로 구성되며 '.'에 의해 분리된다. 접미부는 등록기관이 정한 임의의 문자나 숫자열로 구성되며 기존의 콘텐츠식별자들을 사용할 수 있다. DOI는 콘텐츠를 구분할 수 있는 이름으로 유일한 식별자 역할을 하기 위해 중복성을 피해야 한다. 그래서 DOI등록당시 유일성을 항상 검토하게 된다.

최근 인터넷환경에서 여러 소스로부터 입수한 각국의 특허정보에 대한 효과적인 통합관리가 요구되고 있다. 특허의 서지정보와 원문정보를 연계하기 위해 원문에 대한 서지정보를 중앙에서 통합하고 미리 계산되어 변환서버에 저장된 식별자를 이용하여 위치정보를 파악하게 된다. 이 위치정보는 각국의 특허번호를 기반으로 생성된 DOI이며 유일하게 특허콘텐츠를 식별할 수 있도록 하는 도구이다. DOI를 등록신청하면 접두부를 할당하고 접미부를 부여한 다음 DOI를 다시 검증하고 오류가 없으면 DOI 등록승인과정을 거쳐 DOI번호가 확정된다. DOI등록은 특허번호로부터 자동등록됨과 동시에 DOI 번호가 할당된다. 특허 DOI는 특허문서의 발견과 위치제공, 그리고 원문연계 등의 기능을 지원한다.

특허DOI는 특허 문서에 대한 고유식별코드가

며, 문자와 숫자 16자리, 5개부분으로 구성된다. 특허번호를 발행한 발행국(2자리), 특허공개년도(4자리), 특허번호(7자리), 특허구분코드(2자리), 그리고 특허인지 실용신안인지를 의미하는 형태코드(1자리)로 구성된다. 이 식별체계는 INPADOC의 DOC_NUM 형식을 기준으로 각 국가의 특허번호형식에 기반하여 작성된다. 연도는 한국, 미국, 일본 공개특허에만 적용하고 나머지 국가의 경우 '0'으로 처리한다. 특허구분코드(KC)는 권리의 종류와 특허문서상태를 나타내며, 유럽특허의 구분코드로부터 참조하였다.

표 3. 특허고유식별체계 기술

정의	INPADOC과 각국의 특허데이터를 유일하게 구분해주는 기본키(primary key)															
용도	- INPADOC 및 각국 특허에 대한 Key Value - INPADOC과 개별국, 개별국과 INPADOC, 개별국과 개별국간의 연동시 연결고리															
생성 기준	INPADOC의 DOC_NUM형식을 기준으로 하여, 개별국의 PN형식을 고려															
형식	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
	U	S	2	0	0	1	1	2	3	4	5	6	7	A	1	P
	국가		년도				특허고유번호(PN)					KC		K		

표 4. 특허DOI 적용사례

	각국 특허번호	적용사례
한국	한국공개 1985-0012345	KR19850012345A1P
	한국등록 0032585	KR00000032585B1P
일본	일본공개 1986-002356	JP19860002356A1P
	일본등록 3229264	JP00003229264B2P
미국	미국공개 2001-1234567	US20011234567A1P
	미국등록 5678901	US00005678901B1P

IV. 특허 DOI를 이용한 특허원문연계체계

각국 특허청으로부터 입수한 각 국가 고유의 이미지포맷 및 데이터포맷으로부터 메타데이터와 원문이미지를 추출하여 데이터베이스로 구축한다. 각 국가마다 정해진 규칙에 따라 부여한 특허번호를 하나의 통일된 고유식별체계로 표준화하여 DOI를 생성하고, 특허정보 메타데이터와 해당 특허원문을 연계하는 시스템을 개발한다.

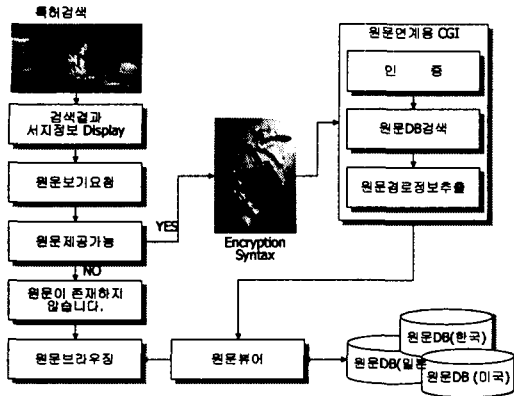


그림 1. 특허원문연계방법

각국 특허청에서 특허정보에 대한 식별기호로 부여하는 특허번호의 체계는 다양하다. 본 논문에서는 다양한 특허번호체계를 하나의 표준화된 고유식별체제로 변환하여 특허에 대한 DOI를 생성하여 원클릭으로 디지털화된 특허원문에 접근하는 방법에 관해 제안하였다.

특허서지정보와 특허원문을 연계함으로써 콘텐츠의 거래내역 확인이 가능하며, 거래내역의 확인을 통한 거래의 투명성을 확보할 수 있다.

참고문헌

[1] 한국특허청. <http://www.kipris.or.kr/>
 [2] 일본특허청. <http://www.jpo.go.jp/>
 [3] 미국특허청. <http://www.uspto.gov/>
 [4] Norman Paskin, "DOI : Current Status and Outlook", D-Lib Magazine, May 1999, <http://www.dlib.org/dlib/may99/05paskin.html>.

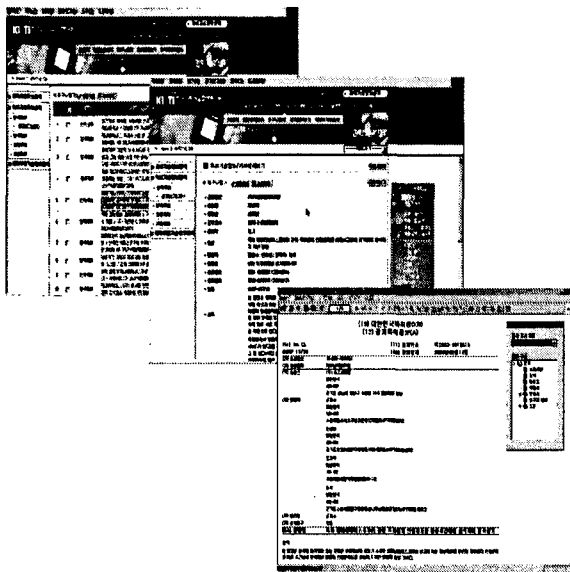


그림 2. 특허원문연계 구현사례

이용자가 검색엔진을 통해 검색된 데이터에 대해 이미지보기 아이콘을 클릭하면, 특허정보 메타데이터로부터 특허원문으로 이동하기 위해 특허원문에 대한 고유식별체계를 생성하여 원문이 있는 서버에 CGI 루틴을 호출한다. 원문을 보유하고 있는 서버에서는 전송된 고유식별체계를 파싱하여 원문의 경로정보를 찾아 이미지브라우저를 호출하여 제어권을 이미지 브라우저에 넘기고 해당문헌 전체를 클라이언트 컴퓨터에 다운로드한다.

V. 결론