

초록 데이터베이스 구축에 있어서 학술지 전자출판 파일의 활용과 문제점: KoreaMed를 중심으로*

The Utilization of Electronic Journal Files in the Production of an Abstract Database: A case of KoreaMed System

이 춘 실(Choon-Shil Lee)**

초 록

본 연구에서는 인쇄본 학술지 발행과정에서 발생된 학술지의 전자출판 파일을 활용하여 신속하고 효율적으로 초록 데이터베이스를 구축하는데 있어서, 2001년 현시점에서 무엇이 가능하며 어떤 것이 문제점인지를, 현재 국내에서 영문으로 구축되고 있는 한국 의학학술지 논문의 초록 데이터베이스인 KoreaMed를 중심으로 분석하였다. 학술지 인쇄출판시 생성된 컴퓨터 파일을 이용하여 초록정보 데이터베이스를 구축하는 KoreaMed의 구축방법론은 데이터의 정확성, 입력작업의 신속성, 비용절감 등에 있어서 그 효율성은 입증되었으나, 학회의 절대적인 협조 없이는 매우 어려운 작업이다. 국내 학회와 인쇄소의 영세성으로 학술지 파일의 보관이 잘 안되어 있어서 학술지 과월호 파일의 수집이 수월하지 않으며, 새로 발행되는 호의 컴퓨터 파일 수집을 위해 학회와 지속적인 관계를 유지하는 것 또한 매우 어려운 것으로 나타났다. 전자출판 파일을 이용하여 데이터를 가공하는데 있어서 제일 까다로운 문제는 특수문자의 처리이며, 학술지마다 출판형식(layout과 format)과 파일 구성방식이 다르고, 사용하는 출판편집 소프트웨어가 다르며, 학술지 파일의 수록매체 등이 매우 다양한 것도 데이터 처리를 복잡하게 만드는 문제점인 것으로 나타났다. 앞으로 학술지 발행과 동시에 학술지 발행기관이 직접 XML로 가공하여 KoreaMed에 제출하는 등 KoreaMed 시스템을 신속하고 효율적으로 구축하기 위한 공조체제를 갖추기 위해서는, 학술지 전자출판 파일의 관리와 XML 파일 만들기 등에 대한 지속적인 홍보와 교육이 필요하다.

ABSTRACT

The study examined the current status and the use of electronic publishing files to produce a bibliographic database. In particular, it examined the problems faced in the production of KoreaMed, an abstract database of Korean medical journals. The methodology of KoreaMed to utilize the computer files which was produced in the process of publication of a print journal is found to be very effective. It assures the accuracy of data, accelerates the input speeds, and reduces the input costs. However, such a project can not be accomplished to a satisfactory level without the cooperation of the publishers involved. It turns out that many small publishers and academic societies hardly have saved the electronic publishing files of previous issues. Besides, it is hard to maintain the right channel to receive the files continuously. The input and processing of special characters are very delicate problems. In addition, the diversity of layout and formats of journals, of the electronic publishing software used, and of the storage media, makes the utilization of the electronic publishing files a very complex process. In order to operate the KoreaMed system more efficiently by requiring the publishers to submit XML files which meets the standard of the KoreaMed, it is necessary to educate and train personnels of journal publishers for the management of electronic publishing files.

키워드: 학술지 전자출판, 전자출판 파일, 초록 데이터베이스, KoreaMed, 의학 데이터베이스, 한국의학학술지

* 이 연구는 과학기술부 지원 '01 숙명여자대학교 연구기반 확충사업(01-N6-02-01-A-01)에 의하여 지원되었음.

** 숙명여자대학교 문헌정보학 전공 교수(cslee@sookmyung.ac.kr)

논문접수일자 2001년 11월 15일

게재확정일자 2001년 12월 5일

1. 서론

최근에 와서 국내에서 발행되고 있는 학술지는 모두 저자에게서 논문원고를 컴퓨터파일로 접수한 후 전자출판편집 소프트웨어를 이용하여 컴퓨터로 편집작업을 한 후 인쇄 출판되고 있다. 즉, 학술지 발행은 학술지 전자출판 파일의 생성을 의미한다. 따라서 전자학술지 원문 데이터베이스 구축이나 색인 및 초록 데이터베이스의 구축은 이러한 전자출판 파일로부터 직접 만들어질 수 있는 환경이 조성되어 있다.

본 연구에서는 학술지 발행에 사용된 학술지의 전자출판 파일을 활용하여 신속하고 효율적으로 초록 데이터베이스를 구축하는데 있어서, 2001년 현시점에서 무엇이 가능하며 어떤 것이 문제점인지를, 현재 국내에서 영문으로 구축되고 있는 한국 의학학술지 논문의 초록 데이터베이스인 KoreaMed를 중심으로 고찰하였다.

2. KoreaMed

(<http://www.koreamed.org>)

KoreaMed는 대한의학회 산하 대한의학학술지편집인협회(의편집, KAMJE)에서 구축 중인 영문 초록 데이터베이스이다(그림 1). 국내에서 발행되는 의학학술지 중에서, 의편집이 1997년부터 시행하고 있는 총 32개 항목으로 구성된 "의학학술지 평가"에서 평점 2.5 이상(5점 만점)을 받은 학술지에 발표된 논문을 대상으로 하고있는데, 2001년에 4종이 추가되어 10월 말 현재 KoreaMed 등재학술지는 총 73종이다. KoreaMed 학술지의 연도별, 호별 입력현황은 KoreaMed Journal Browser(<http://www.koreamed.org/JournalBrowser.php>)를 통해 실시간으로 확인 가능하다(그림 2).

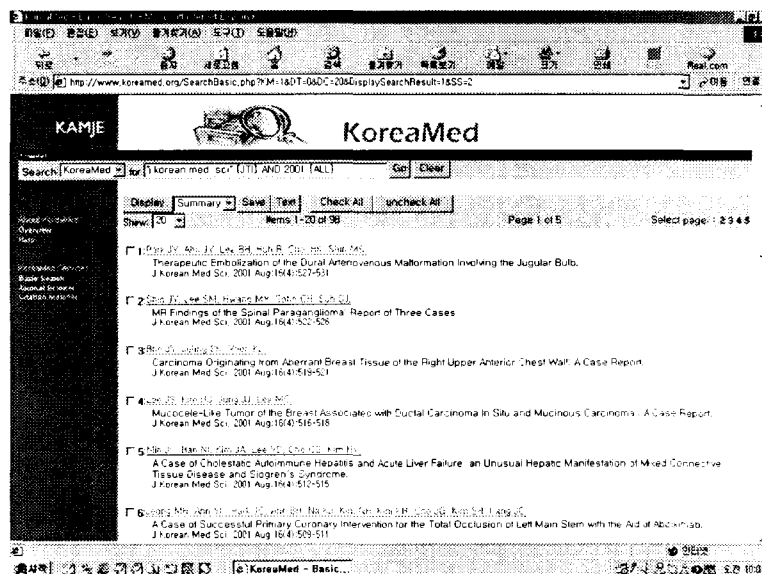


그림 1. KoreaMed 검색결과 화면 (Summary 레코드)

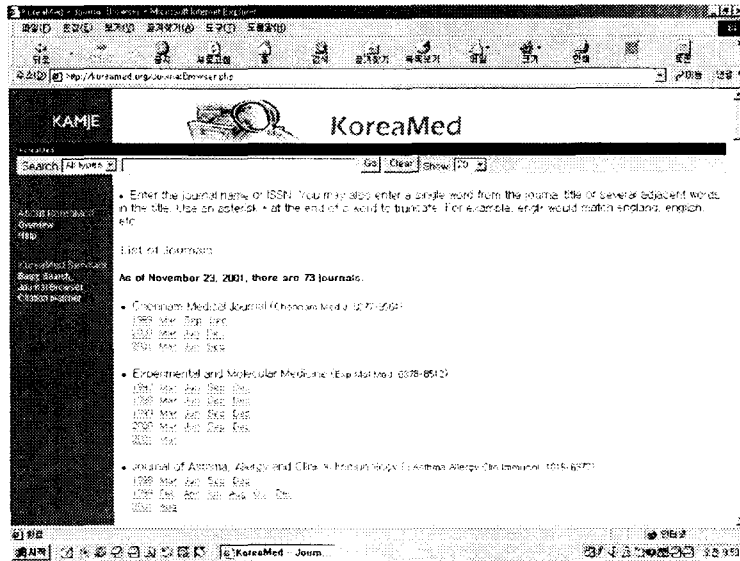


그림 2. KoreaMed 학술지 입력 현황: Journal Browser

의편집 가입학회 및 단체는 128개이고 (<http://www.kamje.or.kr/member>: 의편집 2001a), 의편집 가입 기관들이 발행하는 학술지는 132종이므로(의편집 2001b), 이 중의 절반정도가 현재 KoreaMed 학술지인 셈이다. 참고적으로 언급하자면, 국내에서 현재 발행되는 의학학술지는 220여종이고(의편집 1997), 대한 의학회 회원학회는 116개이다(대한의학회 2000a, 2000b, 2000c). KoreaMed 등재학술지 73종의 연간 발표논문수는 8,000편 정도

인데(표 1), 국내학술지에 발표되는 의학논문수는 연간 15,000편에서 20,000편 정도로 추정된다.

a. KoreaMed 시스템 구축 목표

KoreaMed의 구축이 논의되던 1997년 무렵에는 국내에 이렇다 할만한 의학학술정보 데이터베이스가 존재하지 않았기 때문에, KoreaMed 시스템의 구축 목표는 의학계 전체가 광범위하게 사용하는 한국 의학학술정보 데

표 1. KoreaMed 학술지 연간 발표논문수(추정)*

	1998년	1999년	2000년
응답학술지수	46종	50종	42종
논문수	4,841편	5,325편	4,723편
평균논문수/학술지	105편	105편	112편
추정논문수	7,682편	7,643편	8,209편

*의편집이 2001년 9월에 시행한 설문조사 결과를 토대로 한 추정치.

이터베이스의 구축이었다. 그 후에 한국의학도서 관협의회가 공동구축하는 의학관계 학술잡지 종합정보시스템(MEDLIS)에 “학술잡지 기사색인”(http://medunion.snu.ac.kr)이 추가되었고, 과학기술부 지정 전문연구정보센터 중의 하나인 의학연구정보센터(MedRIC)의 “국내 의학학술지 초록 검색”(http://kmbase.medric.or.kr/mjbrs.htm) 등이 개발되었다. 국회도서관의 『정기간행물 기사색인』 중에 의학학술지가 일부 포함되어 있을 뿐, 의학분야는 과학기술분야 정보 데이터베이스에서도 소외되고 있어서, 한국 의학학술지 논문에 대한 검색은 사실상 외국 의학학술정보보다 이용하기 더 어려운 것이 현실이다.

또한 KoreaMed는 한국 의학학술 정보의 국제화를 위하여, 한국 의학학술 논문에 대한 서지 및 초록 정보를 영문으로 배포하는 것이 중요하다고 판단하고 영문으로 데이터베이스를 구

축하고 있다. 국내 및 외국 이용자들의 이용 편의를 도모하기 위한 방안으로 국제적으로 널리 이용되고 있는 PubMed와 유사한 형식으로 시스템을 구현하고 있다.

데이터의 입력은 KoreaMed가 개발한 KoreaMed DTD에 맞추어 학술지 한호 단위로 저자명, 논문명, 기관명, 초록, 문헌유형, 저자키워드와 MeSH 용어 등의 영문정보를 XML로 가공하여 처리하고 있다.

b. KoreaMed 구축사업 연혁 및 실패원인 분석

1997년에 KoreaMed 시스템 구축이 시작되었으나, 그 동안 두차례에 걸쳐 서비스가 중단되고, 그 때마다 처음부터 다시 새로운 시스템을 설계하고 재구축 작업을 하였다(표 2).

KoreaMed 구축사업의 실패원인을 분석해보면, 첫째 검색엔진 개발 및 보수 유지의 문제

표 2. KoreaMed 구축사업 연혁

날 짜	연 혁
1997년 3월	대한의학회 산하 대한의학학술지편집인협의회(Kamje)에서 KoreaMed 사업추진 결정.
1997년 12월	KoreaMed 서비스 시작. 이원정보통신에서 운영.
1999년 2월	인포메드에서 운영(이원정보통신의 담당 프로그래머 이직).
1999년 5월	인포메드에서 서비스 중단(서버 해킹 당함).
1999년 6월	KoreaMed 운영 공모.
	한국학술정보에서 검색엔진 개발 및 데이터베이스 재구축 시작.
2001년 2월	한국학술정보에서 구축성과 없고, 계속 서비스할 능력이 없다고 통고하고 계약 해지. 특히, 사업성 없음 인정(2억원 이상의 경비 요청).
2001년 3월	의편협 KoreaMed 팀 구성. 검색엔진 개발 및 시스템 재구축 시작.
2001년 11월	1997년 이후 발표된 논문을 우선적으로 입력하고 있음. 최근호 입력에 최우선순위를 두고 있음.

(출처: 허선, 2001a. “코리아메드(KoreaMed)의 문제점.”)

점을 들 수 있다. 영문 저자명 검색이 제대로 안되는 등 (e.g., Lee CS를 구로 처리하지 않고, 두 단어의 OR로 검색처리함) PubMed 수준의 검색기능이 구현되지 못 하였고, 프로그래머의 이직으로 개발이 중단되고, 서버관리 소홀로 인한 서비스 중단사태 등이 일어났다. 둘째 사업성이 없다는 점이다. 학술지의 종수가 많고 입력비용이 과다하게 들어가는 것에 비하여, 국내에서 발생하는 의학 학술정보의 이용률이 매우 저조하고, 특히 유료로 사용할 실제적 필요성이 별로 없어서 수익성이 없는 사업이라는 것이다. 셋째, 학술지 수집의 문제점인데, 인쇄본 또는 인쇄용 파일 수집에 있어 각 학회의 협조를 받는 것이 여의치 않다는 것이다. 국내 의학 도서관의 국내학술지 소장상황이 별로 좋지 않은 것과 같은 맥락일 것이다. 넷째, 데이터 품질 관리의 미흡함인데, 기관명이 입력 안된 경우, 초록이 불완전한 경우 등 입력되어야 할 모든 데이터가 충실하게 반영되지 못하고, 무엇보다도 모든 데이터를 영문으로 입력요원이 손으로 쳐서 넣는 수작업 입력과정에서 오타가 많이 발생한 점이다. 이것은 모두 이러한 데이터베이스의 구축사업이 저예산 project 라는 데서 생겨난 문제일 것이다.

c. KoreaMed 시스템 구축을 위한 새로운 방법론 모색

이러한 실패요인 분석을 통하여 KoreaMed는 새로운 구축방법론을 모색하게 되었다. 즉, KoreaMed는 비영리기관이 사명을 갖고 구축할 수밖에 없으며, 데이터베이스 구축에 있어서 학술지 수집업무 등에 학회의 자발적 협조가 필요하며, 입력비용 절감과 신속하고 효율적인

데이터베이스 구축을 위하여 전자출판 파일을 활용해야 한다는 것이다. 그런데 저작권 문제와 학회, 출판사, web site나 CD-ROM 제작업체들의 이해관계 등으로, 의학계가 공히 인정하는 기구가 아니면 사실상 인쇄용 컴퓨터 파일 수집에 있어서 학회의 협조를 받기가 어렵기 때문에, 각 학회(학술지 편집인)의 협의기구인 의협회가 실질적 주체가 되어 학술지 전자출판 파일을 수집하고 시스템 구축 및 운영을 하는 것으로 결론지어졌다.

3. 학술지 전자파일 활용의 당위성

초록 데이터베이스 구축에 있어서 학술지 전자파일을 활용하는 것의 필요성은 정보의 효율적 관리라는 측면과 데이터베이스 구축 생산성의 향상이라는 측면으로 나누어 볼 수 있다.

a. 정보의 효율적 관리

학술지 정보의 효율적 관리란 그림 3에 보이는 것처럼, “컴퓨터에 한번 key-인(digital화된) 학술정보는, 동일한 정보를 되풀이하여 입력하는 낭비없이 (1) 인쇄출판 및 (2) 전자출판(web 출판, CD-ROM 출판 포함)을 하고 (3) 출판논문에 대한 색인 및 초록 정보 데이터베이스도 만드는 것”이라고 정의할 수 있다. 이때, 최초의 입력자는 물론 저자이다.

b. 데이터베이스 구축 생산성 향상

출판에 사용된 전자파일로부터 데이터를 직접 추출하게 되면 입력속도가 급증하여 데이터 가공 시간이 단축된다. 특히 초록 정보의 입력

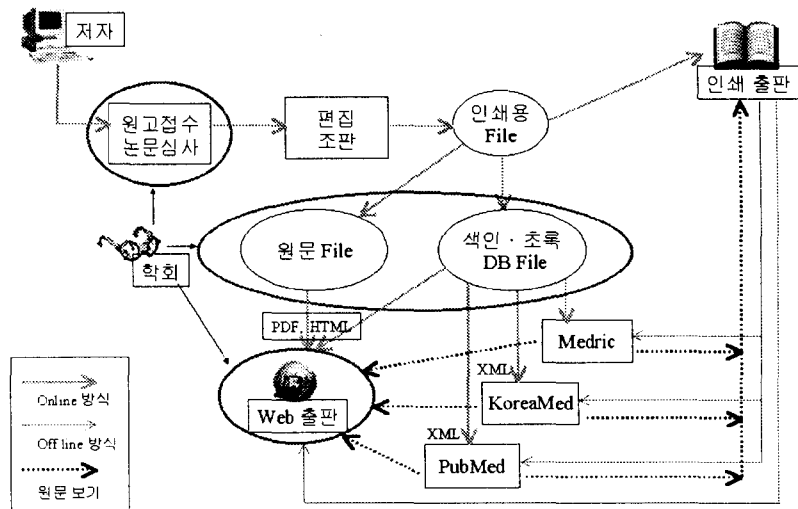


그림 3. 의학학술정보의 생산과 유통 흐름도

(출처: 이춘실, 2001b. "의학학술지 원문정보 및 초록정보의 효율적 관리.")

이 간단해진다. 현재 KoreaMed에서 논문 15편이 수록된 국내 의학학술지 한 호를 1인의 입력요원이 처리하는데 걸리는 시간은 1시간 정도이다(KoreaMed 입력요원 1인의 1일 처리량은 레코드 100건 정도).

전자출판 파일로부터 데이터를 직접 추출하게 되면 학술지 인쇄본을 보고 수작업 입력을 하는 것에 비하여 입력비용이 당연히 절감된다. 또, 학술지 출판파일 원문에서 그대로 옮겨오고 다시 key-in 하는 것이 아니므로 오타가 거의 없다. 즉, 데이터의 정확성을 확보할 수 있기 때문에 수작업 입력에 비교하면 데이터의 품질도 지대하게 향상된다고 할 수 있다.

4. 학술지 전자파일 수집과 관련된 문제점

학술지 전자파일을 데이터베이스 구축에 활용

하는 것은 매우 바람직하고 유용하다. 그러나 실제로 KoreaMed에 등재된 모든 학술지의 전자파일을 수집하는 데에는 여러 가지 문제점이 있다. 첫째, 학회와 인쇄소의 영세성인데, 결과적으로 학술지 전자출판 파일의 보관이 잘 안되어 있고, 새로운 호가 발행될 때 계속해서 파일을 데이터베이스 구축기관으로 보내는 것이 잘 안 되는 이유이다. 둘째는 저작권의 문제이다.

a. 학회의 영세성

다른 분야에서도 마찬가지이지만 의학분야에서도, 회장 임기(2, 3년)에 따라 편집위원 및 학회 주관학교의 변경으로 학회 담당자의 연락처가 자주 변경되는 것이 현실이다. 학회 사무실을 학회장의 소속기관이 아닌 별도의 위치에 갖고있거나 편집 지원인력이 있는 학회가 50% 이상이라는 하지만(표 3), 학술지 편집 및 관리를 위한 전담 직원을 따로 두고 있는 학회가 그렇게 많은 것은 아니다. 대개는 학술지

편집위원장 또는 편집간사가 업무를 수행하고 있다. 그러므로 학회회장단이 바뀔 때마다 새로 바뀌는 담당자에게 업무의 연속성을 기대하기 어렵다.

b. 인쇄소의 영세성

국내 의학학술지를 5종 이상 발행하고 있는 아카데미아, 한국문화사 등 몇 개의 인쇄소를 제외하면 모두 1, 2종의 학술지를 발행하는 인쇄소들이다(표 4). 인쇄소가 e-mail로 그림이 삽입된 큰 파일을 보낼 수 있는 환경이나, 자체적으로 학술지 파일 보관본 CD-ROM을 만들

수 있는 환경은 최근에 와서야(1999년 이후) 가능해졌다.

c. 학술지 전자출판 파일 보관 미흡

학회와 인쇄소의 영세성으로 현재 학술지 출판파일을 일관되게 보관하고 있는 학회는 별로 없다. 이것은 또한 학회와 인쇄소 사이에서 학술지 파일 보관여부가 불분명하게 처리되고 있기 때문에 일어나는 현상이기도 하다. 2001년 학술지 파일이 없는 경우는 5개 학회에 불과하지만, 1998년 학술지 파일이 없는 학회는 27개나 된다(표 5). 응답하지 않은 학회는 대부분

표 3. 학회 운영현황*

학회(편집)사무실 별도 유지		편집업무 지원인력			
		없음	있음 ¹		
예	아니오		1명	2명	3명
41	31	30	38	2	2
(57.0%)	(43.1%)		(90.4%)	(4.8%)	(4.8%)

*의편집 가입학회 128개 중 72개 학회 응답.

¹지원인력이 있는 42개 학회 중에서의 비율

(출처: 허선, 2000. "국내 의학학술지 발행 및 편집실무 현황".)

표 4. 국내 의학학술지 발행 인쇄소 현황*

인쇄소	발행 학술지수	발행 의학학술지
의학문화사	9종	(대한외과학회지, 대한통증학회지 등)
광문출판사	8종	(소아과, 대한방사선종양학회지 등)
중앙문화사	7종	(대한이비인후과학회지, 순환기 등)
최신의학사	6종	(대한슬관절학회지, 대한안과학회지 등)
성문각	5종	(대한방사선의학회지, 대한초음파의학회지 등)
아카데미아	5종	(J Korean Med Sci, 대한병리학회지 등)
기타(20개)	40종	
무응답	4종	

*KoreaMed 학술지 73종

컴퓨터 파일을 보유하고 있지 않을 것으로 보이는 영세학회들이다. 특히 1998년 이전에 발행된 학술지의 출판파일을 가지고 있는 학회는 매우 적은 것으로 추정된다.

그러나 표 6에 보이는 것처럼 대부분 (83.6%)의 학회가 학회 web site를 운영하고 있다. 그 중에서 60% 정도가 학술지 원문파일이나 초록파일을 제공하고 있으며, 학회지 논문의 검색 기능을 제공하고 있는 학회도 60% 정도이다.

d. 계속적으로 파일 받기의 문제점

학회들이 파일제출 요구를 받았을 때만 일회성으로 파일을 제출할 뿐, 새로운 호를 발행할 때마다 계속해서 KoreaMed에 파일을 제출하기

위한 체재 및 준비가 안되어 있다. 이것은 학회와 인쇄소의 상호연계가 잘 안되고 있기 때문이기도 하다. 최종 파일을 학회에서 관리하는 것이 아니라 출판사에서 관리하는 경우가 많은데(허선 2001b), 출판사와의 접촉은 KoreaMed가 직접 할 수 없고, 학회에서 하는 것이 원칙이기 때문에 파일 받기에 걸림돌이 되고 있다.

e. 저작권 문제

KoreaMed는 학회지 원문파일을 입수하는 경우에도 원문 데이터베이스를 구축하는 것이 아니라, 필요한 서지사항 및 초록정보만을 추출하여 데이터베이스를 구축하는 것이기 때문에 법에 저촉되는 것이 없는 것으로 판단된다. 그러나, 학회 web site 구축 업체 또는 학술지 초

표 5. KoreaMed 학술지 전자출판 파일 보관 현황*

	1998년	1999년	2000년	2001년
파일 있음	27	38	44	49
모두	25	31	39	43
일부	2	7	5	6
파일 없음	27	16	10	5
발행안됨	1	1	1	1

*KoreaMed 학술지 73종 중 53종 학술지 응답(응답율 72.6%).
의편협이 2001년 9월중 실시한 설문조사 결과.

표 6. KoreaMed 학술지 Web site 현황*

	학회수
Web site 있음	61 (83.6%)
초록제공	37
원문제공	33
논문 검색기능 제공	35
Web site 없음	12 (16.4%)

*KoreaMed 학술지 73종 (2001년 10월 31일 현재).

록 및 원문 CD-ROM 개발 업체 등이 어떠한 형식으로든지 학회지 전자출판 파일이 유통되는 것을 금지하거나 문제제기를 하는 경우가 있다. 학회에서도 원문파일의 유출을 원하지 않기 때문에 학회들이 KoreaMed 데이터베이스에 들어갈 영문 서지정보 및 초록 정보만 추출하여 제출하려고 하지만, 마땅한 인력이 없고 시간이 많이 걸리기 때문에 실제로는 원문파일을 있는 그대로 제출하기도 한다. 또, 학회의 연락을 받고 실제로는 인쇄소에서 학술지 파일을 제출하기도 한다.

5. 전자출판 파일을 이용하여 데이터를 가공하는데 있어서의 문제점

전자출판 파일을 이용하여 데이터를 가공하는데 있어서 제일 까다로운 문제는 특수문자의 처리일 것이다. 또한 학술지마다 출판형식(layout & format), 출판편집 소프트웨어, 파일 구성방식, 학술지 파일의 수록매체 등이 매우 다양한 것도 데이터 처리를 복잡하게 만드는 문제점들이다.

a. XML editors 사용 불편.

XML editors(표 7)를 이용하면 만들어진 XML 문서가 유효한(valid) 문서인지 검사하는 것은 가능하지만, 특수문자의 entity 처리가 자유롭게 안되는 불편한 점이 있다. 또 입력과정에서 입력할 데이터 요소들을 매번 지정한 후 text 데이터를 입력하는 것은 불편하고 작업속도도 많이 떨어진다.

XML editors를 이용하여 입력하는 것보다는 차라리 Editpad나 Notepad 등의 일반 text 문서 편집기에서, 템플릿(template)을 이용하여 학술지 파일에서 데이터를 복사하여 붙이는(copy & paste) 것이 더 편리하다(그림 4). Explorer 5.0이상에서 well-formed 문서인지 검사가 가능하고, 기본적인 특수문자의 entity가 제대로 사용되었는지 여부도 확인 가능하다(그림 5). 그리고 KoreaMed 시스템은 XML 입력파일이 유효한지 오류가 있는지 검사한다(그림 6). 따라서 KoreaMed는 Editpad에서 XML 태그(tags)의 템플릿을 이용하여 데이터 입력작업을 하고, Explorer에서 well-formed 문서인지 검사한 후, KoreaMed 시스템에 올리기(uploading)한 후 데이터 파싱(parsing) 과정에서 레코드 오류를 검사하는 방법을 사용하고 있다.

표 7. KoreaMed에서 사용을 검토하여본 XML editors*

XML editors	개발회사
GeneXis XInstance	건지소프트
Tagfree 2000 XML Editor	(주)다산기술
XEditor	한국지식웨어
XMLwriter	Wattle Software

*KoreaMed에서 사용을 검토해 본 XML editors.

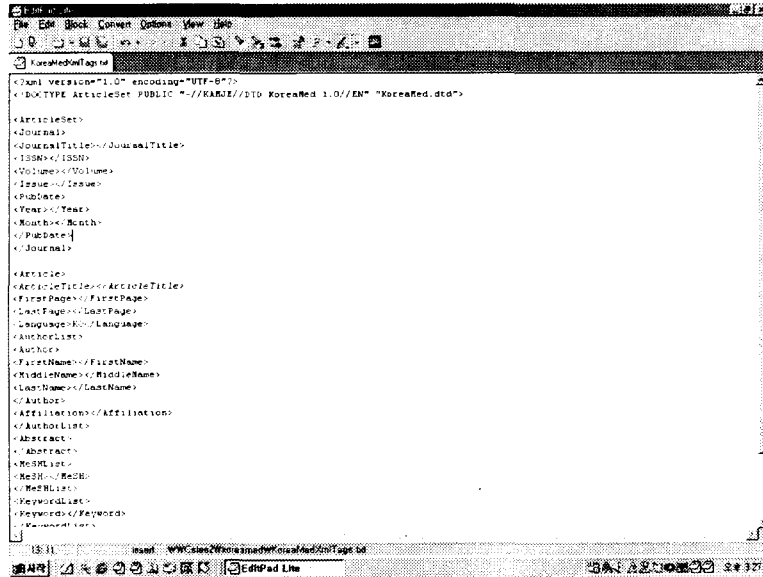


그림 4. KoreaMed XML template (Editpad)

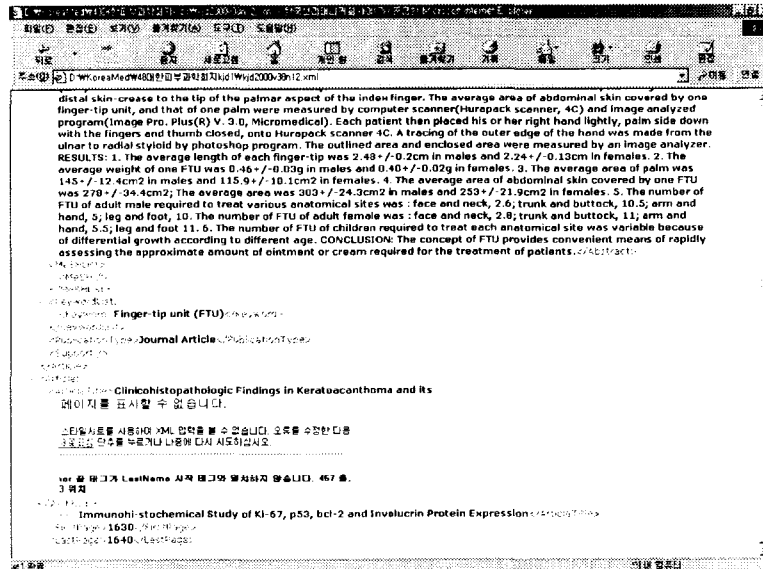


그림 5. Explorer에서 XML well-formed 문서 검사화면

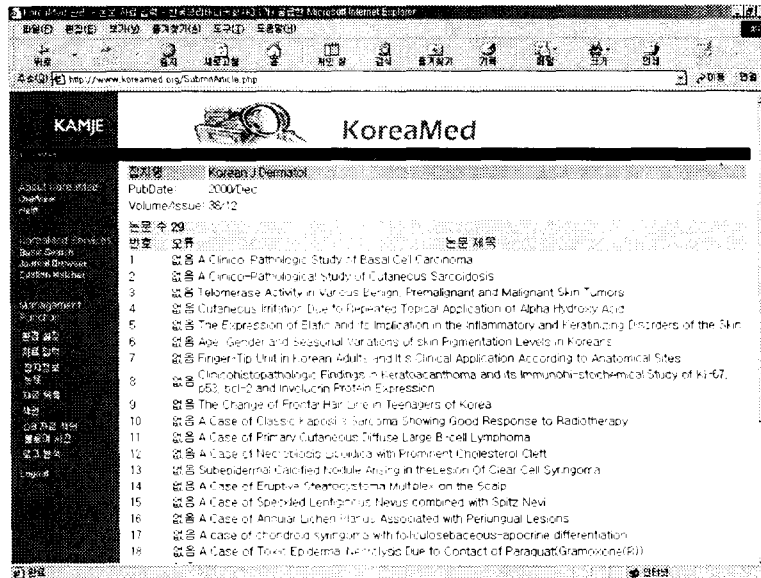


그림 6. KoreaMed XML data parsing 오류 검사 결과

전자출판 파일에서 데이터를 일일이 복사하여 붙이는 불편함을 해소하기 위해, 학술지 텍스트 파일에서 필요한 데이터 필드를 자동으로 추출하고 XML로 변환하는 프로그램을 생성하는 것을 고려할 수 있다. 그러나 학술지마다 논문 title page 형식이 다르기 때문에 이것은 간단한 일이 아니다. 문헌정보학 분야 학술지, 학위 논문이나 기술보고서를 대상으로 여러기관의 형식을 아우르는 DTD를 개발하고자한 연구 논문 등이 있으나(안영선, 이은철, 오삼균 2000; 윤소영 1996; 한종엽 2000), 의학학술지 70여종 이상의 공통분모를 추출하는 것은 불가능한 일일 것이다. 또 조만간에 국내 학술지들이 본문을 XML 태그로 구조화하여 전자출판할 것이라고 기대하기도 어렵다.

b. 특수문자 및 부호 처리
 특수문자 및 부호의 처리는 XML 데이터 처

리에 있어서 제일 까다로운 문제이다. KoreaMed는 <, >, & 등 XML 예약 문자를 <, >, & 등으로 entity 처리하는 것 외에는 모든 특수 문자를 풀어서 입력하고 그대로 출력하는 방식을 사용한다 (표 8). 따라서 α, β, γ 등의 그리스 문자는 alpha, beta, gamma 등으로 풀어쓰고, $\mu\text{g}, \mu\text{L}$ 등의 단위는 microgram, microliter 등으로 풀어 쓴다. 위첨자, 아래첨자 등은 PubMed와 마찬가지로 화학기호일 경우는 무시하고, 그 외에는 괄호 안에 표시한다.

c. 파일 내용(데이터요소)

KoreaMed 구축에 필요한 정보만 추출하여 제출하는 학회도 있으나, 원문 그대로 제출하는 학회가 더 많다. 그림이 삽입되어 있는 경우에는 파일 사이즈가 매우 커서 파일 주고받기와 입력작업의 효율이 떨어진다.

표 8. 특수문자 처리 예: KoreaMed vs PubMed

특수문자	KoreaMed		PubMed	
	입력	display	입력	display
<	<	<	<	<
>	>	>	>	>
&	&	&	&	&
α	alpha	alpha	&agr;	alpha
μg	microgram	microgram	microgram	microgram
\pm	+/-	+/-	±	+/-
m^2	m2	m2	m²	m2
10^{-6}	10(-6)	10(-6)	10⁻⁶	10(-6)
10_6	10(6)	10(6)	10<inf>(6)</inf>	10(6)
height ^{2.239}	height(2.239)	height(2.239)	height^{2.239}	height(2.239)
H ₂ O	H2O	H2O	H2O	H2O

d. 학술지 출판편집 소프트웨어(파일 형식)의 다양성

국내 의학학술지 출판편집 소프트웨어는 한글과 QuarkExpress가 주로 쓰이고 있다(표 9). 한글은 PC 파일이며, QuarkExpress는 Mac 파일인데, Mac 파일은 텍스트 파일로 변환하여 PC에서 KoreaMed XML 파일 만들기 작업을 하고 있다.

e. 학술지 파일구성 방식의 다양성

학술지 별로 파일을 구성하는 방식이 매우 차이가 있는데, 한 호 전체가 한 파일인 경우, 한 호에 실린 각 논문이 하나의 파일인 경우도 있다. 페이지 정보를 파일명으로 사용하는 경우도 있으며, 각 저자명, 또는 논문명의 keywords를 파일명으로 사용하는 경우도 있는데, 이런 경우에 목차 없이 페이지 정보를 입력하는 것은 불가능하다.

표 9. 국내 의학학술지 편집, 출판 소프트웨어*

편집 소프트웨어	학회
한글 (PC)	27
QuarkExpress (Mac)	10
PageMaker (PC)	2
모름	6
무응답	27

*의편협 가입학회 128개 중 72개 학회 응답.
(출처: 허선, 2000. "국내 의학학술지 발행 및 편집실무 현황".)

f. 학술지 파일 저장매체 및 제출 방법의 다양성

Floppy diskette, CD-ROM, zip disk 등에 학술지 파일을 저장하여 퀵서비스 또는 우편으로 보내는 경우도 있으며, e-mail로 보내는 경우, KoreaMed가 지정한 파일서버로 FTP하는 학술지도 있다. 심지어 학회 web site에서 downloading하라고 하는 학회도 있다.

6. 개선 사항 및 앞으로의 추진 방향

a. 학회에서 XML 파일 제작후 Korea-Med에 제출 유도

학술지의 새로운 호가 발행될 때마다 각 학회가 KoreaMed DTD에 맞게 데이터를 입력한 XML 파일을 KoreaMed에 제출하게 된다면, KoreaMed 시스템은 좀 더 신속하게, 효율적으로, 그리고 최소의 비용으로 구축될 수 있을 것이다. 현재 MEDLINE은 각 학술지의 발행자가 PubMed DTD에 맞추어 XML 데이터를 납본하도록 하고 있으며, 출판된 인쇄본이 NLM에 도착하기 전에 XML 파일을 FTP하도록 요구하고 있다(PubMed 2001a). FTP된 학술지 XML 파일은 48시간 이내에 MEDLINE에서 검색이 가능하다. 따라서 학술지 인쇄본이 구독자에게 배포되기 적어도 보름 전에 이미 그 학술지 논문의 검색이 가능하다 (이춘실 2001a). 현재 국내의 MEDLINE 학술지 중 *Journal of Korean Medical Science*와 *Yonsei Medical Journal*이 XML 데이터를 MEDLINE에 납본하고 있다.

b. 각 학회의 원문으로 Link하기

KoreaMed는 영문 초록 데이터베이스로서 원문을 제공하지는 않는다. 그러나 PubMed에서처럼 각 학회(구독 대행기관, 도서관)의 전자 학술지 web site로 link를 제공할 예정이다(그림 7). 각 학회가 회원여부, 구독여부(유, 무료 이용)를 임의로 통제하여 허용된 이용자만 원문이용이 가능하도록 한다면, 저작권 관련 문제의 시비는 없을 것이다.

PubMed에서는 위에 설명한 학술지 XML 파일의 MEDLINE 납본이 학술지 web site의 원문으로 LinkOut을 위한 전제조건이다(PubMed 2001b). 현재 국내 MEDLINE 학술지 중 *Journal of Korean Medical Science*가 PubMed에서 LinkOut 되고 있으며, *Yonsei Medical Journal*은 LinkOut을 추진중이다.

논문 검색후 그 논문의 원문을 가진 기관으로 링크하는 KoreaMed의 방법은, 현재 국내에서 한국학술정보의 KISS나 누리미디어의 DBpia와 같은 상용(commercial) 데이터베이스가 자체 데이터베이스에 원문파일을 보유하고 원문 서비스를 하는 방법과 여러 면에서 차이가 있다. KoreaMed 이용자는 70종 이상의 주요 한국의 학학술지에 대한 통합검색을 하고 나서, 한번의 클릭으로 학술지 web site나 상용 데이터베이스에서 제공하는 원문을 이용할 수 있게 된다. KoreaMed 학술지의 원문을 보유하고 있는 학회나 업체 또한 자신들이 보유하고 있는 원문의 이용을 촉진시킬 수 있다는 점에서, 이 방법은 관련된 기관이 각자의 독립성과 권한을 유지하면서 상호 보완적으로 유기적인 관계를 형성할 수 있는 매우 유용한 것이라고 할 수 있다.

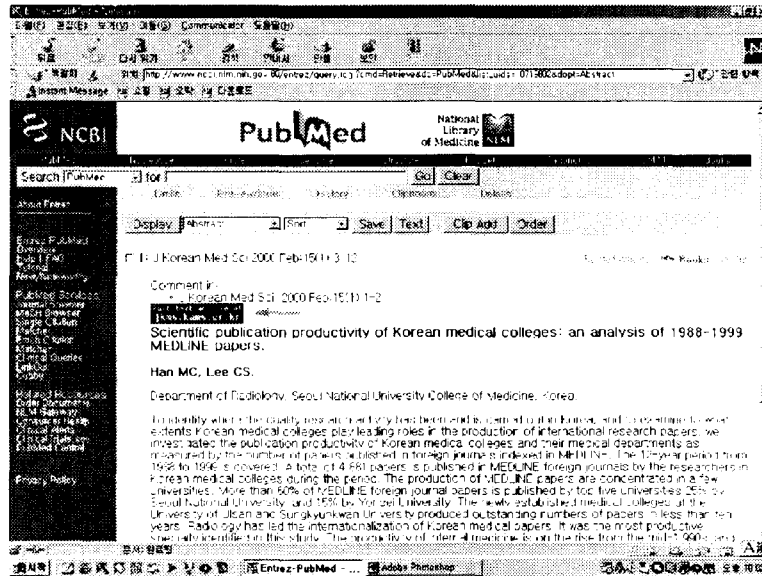


그림 7. *Journal of Korean Medical Science*로 LinkOut 해주고 있는 PubMed 초록레코드 화면 (화살표가 가리키는 icon을 click하면 학회 web site가 제공하는 바로 이 논문의 초록 및 원문 화면으로 link됨.)

c. 컴퓨터 파일이 없는 학술지, 과월호의 처리
 학회가 컴퓨터 파일을 갖고 있지 않는 과월호는 인쇄본의 title page를 scan한 후 OCR을 통하여 영문저자명, 영문논문명, 영문초록 등을 text로 입력하는 방안을 검토 중이다. 입력 속도, 정확도 등이 단순 입력방안보다는 효율적일 것으로 판단된다. 그런데 학술지의 인쇄본을 입수하는 것도 학술지 컴퓨터 파일의 입수만큼 어려운 문제일 것으로 예상된다.

d. MeSH 용어 부여 및 MeSH 검색시스템 구현

KoreaMed 레코드에 논문의 중심주제어(keyword)를 입력하는데 있어서 가장 중요한 결정은, MEDLINE처럼 MeSH 용어를 사용할 것인가의 문제이다. 국내 의학학술지 투고 규정들은 논문의 키워드로 MeSH 용어를 사용

할 것을 요구하지만, 실제로 저자들이 부여하는 주제어는 MeSH 용어가 아닌 경우가 대부분이다(이춘실, 문혜원 2000). KoreaMed에서는 MeSH 필드와 저자키워드 필드를 따로 두고, KoreaMed 학술지이면서 MEDLINE 학술지인 학술지는 저자가 논문에 제시한 키워드와 함께 MEDLINE이 각 논문에 부여한 MeSH 용어를 그대로 채택하여 입력하고, 그 밖의 학술지는 저자의 키워드를 입력하고 있다. 그러나 주제어 검색에서는 두 필드를 통합 검색하도록 처리하고 있다.

KoreaMed의 모든 레코드에 MeSH 용어를 올바르게 부여하기 위해서는 MeSH 용어 전문가가 필요한데, 국내에 MeSH 용어 전문가가 많지 않기 때문에 MeSH 용어 전문가의 양성이 현안으로 대두되고 있다. 그래서 MeSH 용어 전문가가 MeSH 용어를 부여할 것인지, 가칭

KoreaMed Thesaurus(KMT)를 독자적으로 개발할 것인지 등, KoreaMed 시스템의 의학주제용어를 일관되게 사용하는 방법에 대하여 현재 의편집에서 검토중이다.

KoreaMed는 또한 MeSH browser나 KMT browser의 구현을 통하여 레코드 검색 시스템과 연동하는 것도 계획중이다.

e. KoreaMed 이용촉진 방안 모색

현재 Googlebot, Inktomi Slurp 등 20여 개 검색엔진의 robot들이 정기적으로 KoreaMed를 방문하고 있으나, 주요 검색엔진에 site 등록을 계속하여 추진하고 있다. 또, 국내외 각 학회와 의과대학 도서관 web sites에서 link를 제공하도록 요청하고 있다.

7. 결론

학술지 인쇄출판시 생성된 컴퓨터 파일을 이용하여 초록정보 데이터베이스를 구축하는 KoreaMed의 구축방법론은 데이터의 정확성,

입력작업의 신속성, 비용절감 등에 있어서 그 효율성은 입증되었으나, 학회의 절대적인 협조 없이는 매우 어려운 작업이다. 그런데, 국내 학회와 인쇄소의 영세성으로 학술지 파일의 보관이 잘 안되어 있어서 학술지 과월호 파일의 수집이 수월하지 않으며, 새로 발행되는 호의 컴퓨터 파일 수집을 위해 학회와 지속적인 관계를 유지하는 것 또한 매우 어려운 것으로 나타났다. 전자출판 파일을 이용하여 데이터를 가공하는데 있어서 제일 까다로운 문제는 특수문자의 처리이며, 학술지마다 출판형식(layout과 format)과 파일 구성방식이 다르고, 사용하는 출판편집 소프트웨어가 다르며, 학술지 파일의 수록매체 등이 매우 다양한 것도 데이터 처리를 복잡하게 만드는 문제점인 것으로 나타났다. 앞으로 학술지 발행과 동시에 학술지 발행기관이 직접 XML로 가공하여 KoreaMed에 제출하는 등 KoreaMed 시스템을 신속하고 효율적으로 구축하기 위한 공조체제를 갖추기 위해서는, 학술지 전자출판 파일의 관리와 XML 파일 만들기 등에 대한 지속적인 홍보와 교육이 필요하다.

참 고 문 헌

국회도서관. 국내학술잡지 (『정기간행물 기사 색인』).
<http://www.nanet.go.kr:7777/dl/SimpleSearch.php>
대한의학학술지편집인협의회 (의편집).
<http://www.kamje.or.kr>
대한의학회. <http://www.kams.or.kr>

대한의학회. 2001a. 대한의학회 정회원학회. 『대한의학회 회보』, 59: 5-6.
대한의학회. 2001b. 대한의학회 준회원학회. 『대한의학회 회보』, 60: 5-6.
대한의학회. 2001c. 회원학회분류 『대한의학회 회보』, 61: 5.
서정욱 외. 1997. 우리나라 의학학술지의 현황.

- 『대한의학학술지편집인협의회 정기총회 자료집』. 1997년 2월 26일. [서울: 서울중앙병원].
- 안영선, 이은철, 오삼균. 2000. 학위논문의 XML DTD 설계에 관한 연구. 『정보관리학회지』, 17 (4): 113-129.
- 윤소영. 1996. 『학술지 논문기사의 문헌구조 분석을 통한 DTD 개발』. 석사학위논문, 淑明女大 大學院. 문헌정보학과.
- 의편협. 2001a. 회원학회지 및 편집인 현황. 『대한의학학술지편집인협의회 소식』, 15: 4.
- 의편협. 2001b. 단체회원별 학술지 평가결과 (1-9차). 『대한의학학술지편집인협의회 소식』, 15: 3.
- 의학연구정보센터 (MedRIC). 국내 의학학술지 초록 검색.
<http://kmbase.medric.or.kr/mjbrs.htm>
- 이춘실. 2001a. *Journal of Korean Medical Science* 원문정보와 PubMed의 LinkOut연동. 『대한의학회보』, 61: 1-3.
- 이춘실. 2001b. 의학학술지 원문정보 및 초록정보의 효율적 관리: *Journal of Korean Medical Science*의 경우. 『대한의학학술지편집인협의회 workshop』. 2001년 2월 23일. [서울: 서울대학교 임상의학연구소].
- 이춘실, 문혜원. 2000. 한국 의학 학술논문의 저자선정 주제어와 MeSH 용어의 비교 분석. 『정보관리학회지』, 17(3): 109-124.
- 학술잡지 종합정보시스템 (MEDLIS). 학술잡지 기사색인.
<http://medunion.snu.ac.kr/>
- 한종엽. 2000. 연속간행물 기사 DB의 서지 데이터 요소 표준화 연구. 『한국문헌정보학회지』, 34 (3): 109-132.
- 허 선. 2000. 국내 의학학술지 발행 및 편집실무 현황 『대한의학학술지편집인협의회 workshop』. 2000년 3월 28일.
- 허 선. 2001a. 코리아메드 (KoreaMed)의 문제점. 『대한의학학술지편집인협의회 workshop』. 2001년 2월 23일. [서울: 서울대학교 임상의학연구소].
- 허 선. 2001b. 코리아메드 영문초록 입력을 위한 회원학회의 자료 전달방법. 『대한의학학술지편집인협의회 workshop』. 2001년 7월 6일. [서울: 서울대학교 임상의학연구소].
- GeneXis XInstance.
<http://www.geonji.co.kr/geneXis/>
- Journal of Korean Medical Science*.
<http://jkms.kams.or.kr>
- KoreaMed. <http://www.koreamed.org>
- KoreaMed Journal Browser.
<http://www.koreamed.org/JournalBrowser.php>
- PubMed. <http://www.ncbi.nlm.nih.gov/PubMed/>
- PubMed. 2001a. Information for Publishers re: XML Tagged Data.
<http://www.ncbi.nlm.nih.gov/entrez/query/static/publisher.html>. Updated: July 3, 2001.

PubMed. 2001b. Linkout.

<http://www.ncbi.nlm.nih.gov/entrez/linkout/>. Updated November 19, 2001.

PubMed. 2000. SGML entity list.

<http://www.ncbi.nlm.nih.gov/80/entrez/query/static/entities.html>. Updated: July 26, 2000.

Tagfree 2000 XML Editor.

<http://www.tagfree.com/>
XEditor. <http://www.kware21.com>
XMLwriter. Wattle Software.

<http://www.xmlwriter.com/>

Yonsei Medical Journal.

<http://www.ymj.org/>