

효율적인 KSCI 체제 구축을 위한 XML 기반 모델 설계

이계준⁰ 조현양 최재황 손강렬
한국과학기술정보연구원
(kjlee, hycho, findit, krshon@kisti.re.kr)

Planning of XML Based Model for the Construction of Effective KSCI System

Kye-Jun Lee⁰ Hyun-Yang Cho Jae-Hwang Choi Kang-Ryul Shon

Korea Institute of Science and Technology Information

요 약

과학기술 논문의 수준을 평가하고, 국내 학술지 및 기관간의 연구능력 비교·분석의 척도로 사용하기 위한 KSCI(Korean Science Citation Index : 한국과학기술인용색인) 구축 및 활성화를 위하여 XML을 기반으로 하는 모델을 설계하였다. KSCI 데이터베이스는 인용한 논문과 인용된 논문들 사이의 관계를 정의하고 논문을 작성한 저자들에 대한 인력DB의 구축과 연계를 통하여 구성되어진다. 이러한 과정에서 발생되어지는 표준화 과정과 데이터베이스간의 연계를 효율적으로 추진하고 효율적인 KSCI 데이터베이스를 구축하기 위한 XML 표준을 설계하였다. 첫째, 데이터베이스들의 연계를 위한 모델을 설계, 둘째, 인용된 논문과 인용한 논문에서의 서지정보, 저널정보, 참고문헌정보에 대한 XML DTD를 정의, 셋째, 저자와 공저자들에 대한 인력DB 구축을 위한 XML DTD를 정의하였다.

본 논문은 KSCI데이터베이스 구축을 통해서 데이터에 대한 상호 교환, 공동 활용을 보다 효율적으로 수행하고 안정적인 체제 구축을 고려하여 모델을 설계하였다.

1. 서 론

개별 국가의 과학기술 수준을 종합적으로 측정·평가하기 위하여 다양한 통계자료 및 지표들이 사용되고 있으나, 국가별 과학기술력을 평가하기 위하여 활용되고 있는 지표들 중에서 가장 객관적인 지표로써 범용적으로 활용되고 있는 것은 연구결과의 발표장인 각 논문에 인용된 참고문헌의 분석에 근거한 논문의 인용률이다.

KSCI는 국내 연구자들의 연구업적을 보다 객관적으로 평가할 수 있는 도구의 마련과 국가 과학기술 경쟁력 확보라는 대 명제 하에 국내 학술지간, 기관간의 연구능력의 비교·분석을 위한 척도로 활용될 것이며, 국내 학술활동 활성화를 위한 기반 조성을 위하여 「한국과학기술인용문헌 DB」의 구축 및 기획연구가 수행되고 있다.

본 논문에서는 KSCI 구축 및 활성화를 위하여 기본 구성이 되는 첫째, 인용한 논문과 인용된 논문의 관련성 추론, 둘째, 참고문헌의 기술적 표현에 관한 부분, 셋째, 저자에 대한 인력DB와의 상호 연계를 반영 및 인용색인의 자동화, 인용색인의 검색 효율의 향상을 위하여 XML을 기반으로 하는 모델을 설계하였다.

모델의 구성요소는 첫째, 인용한 논문과 인용된 논문의 기본 정보를 표현하기 위한 DTD, 둘째, 참고문헌의 표현을 위한 DTD, 셋째, 인력DB 표현을 위한 DTD, 넷째, 상호 연계 및 관리를 위한 DTD등의 정의로 이루어진다.

본 논문은 KSCI 체제 구축을 위하여 XML 기반 모델을 설계하게된 연구 배경과 본문에서는 모델의 구성요소를 기본으로 DTD를 정의하고 설계하였으며 마지막으로 결론을 내는 것으로 구성되어 있다.

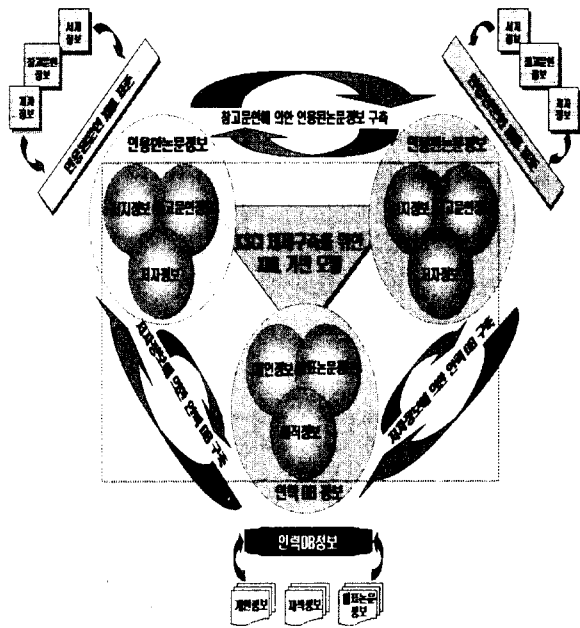
2. 연구배경

과학기술 논문의 수준을 평가하기 위한 객관적인 비교의 틀로서 널리 활용되고 있는 과학기술인용색인(SCI: Science Citation Index)은 영미권 중심이어서 국내 연구자의 연구 업적(SCI 전체 학술지의 0.37%가 국내 학술지임)을 평가하기에는 부적절하며, 국내 실정에 맞지 않는다. 이를 해결하기 위하여 KSCI의 구축을 통해서 국내 학술지간, 기관간의 연구능력 비교·분석의 척도로 활용하기 위한 체제를 구축하기 위해 연구가 활발히 이루어지고 있다. 그러나 KSCI를 구축하는데 있어 필요한 데이터의 상호 교환이나 공동 활용에 대한 방안이 미비한 실정이며, 인용된논문, 인용한논문, 참고논문, 인력DB 사이의 유기적인 결합 및 관리를 위한 표준이나 모델이 미비한 실정이다.

국내 과학기술 논문에 대한 KSCI DB를 구축하고 이것들에 대한 분석을 통해 과학기술 동향과 흐름을 파악하며, 새로운 평가 방법을 바탕으로 국내 과학기술 논문의 질적인 향상을 유도하기 위해 XML 기반 표준화와 모델화를 통하여 보다 효율적이고 안정적인 체제 구축을 위한 방향을 본 논문에서 제시하고자 한다.

3. 본 문

본 논문에서는 KSCI의 기본이 되는 인용한 논문정보와 인용된 논문정보 및 인력DB정보의 연계를 위한 설계를 하였다. “효율적인 KSCI 체제 구축을 위한 XML 기반 모델 설계”의 전체 구성은 [그림 3-1]과 같다.

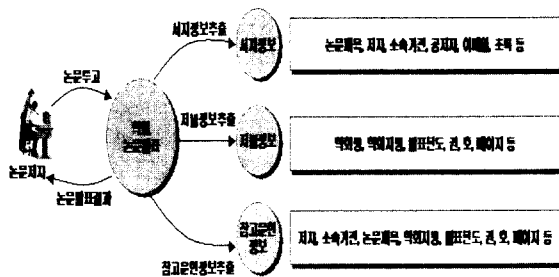


[그림 3-1] XML 기반 모델 설계의 전체 구성도

[그림 3-1]에서 보는 것과 같이 KSCI 체제구축을 위해서는 인용한 논문정보, 인용된 논문정보, 인력DB정보 사이의 유기적인 연계를 통하여 이루어진다.

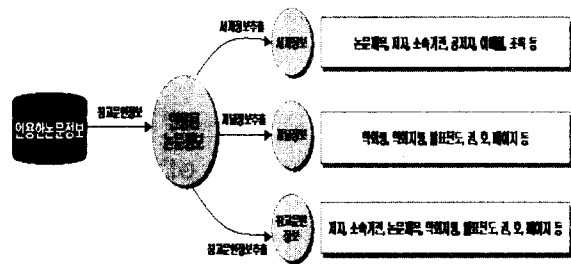
3.1 KSCI 기본 데이터 구성 요소

학회를 통해 발표된 논문들이 가지게 되는 정보를 보면 [그림 3-2]의 인용한 논문정보의 구성과 같이 이루어져 있다.



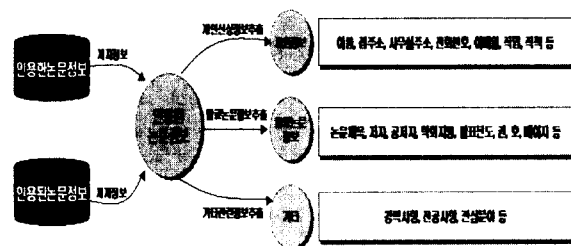
[그림 3-2] 인용한 논문정보의 구성

위 [그림 3-2]와 같은 방법으로 구축된 정보 중에서 참고문헌 정보를 대상으로 인용된 논문정보를 구축하기 위한 것은 [그림 3-3] 인용된 논문정보의 구성과 같이 이루어져 있다.



[그림 3-3] 인용된 논문정보의 구성

인용한 논문정보와 인용된 논문정보와의 연계를 위한 인력DB정보의 구성은 [그림 3-4] 인력DB정보의 구성과 같이 이루어진다.



[그림 3-4] 인력DB정보의 구성

3.2 XML기반 DTD 설계

KSCI 기본 데이터 구성요소들을 기초로 하여 가장 기본이 되는 요소들을 대상으로 XML기반 DTD를 다음과 같이 설계하였다.

```
<?xml version="1.0" encoding="euc-kr" ?>
<!-- 저널정보를 나타내기 위한 ENTITY 정의 부분 -->
<ENTITY % BOOKINFO "%JOURNALNAME?: %PUBLICYEAR:, %VOLUME:, %NUMBER:, %PAGE:" >
<ENTITY % JOURNALNAME "(%KORJOURNAL?:%ENGJOURNAL?)">
<ENTITY % KORJOURNAL "(#PCDATA)" >
<ENTITY % ENGJOURNAL "(#PCDATA)" >
<ENTITY % PUBLICYEAR "(#PCDATA)" >
<ENTITY % VOLUME "(#PCDATA)" >
<ENTITY % NUMBER "(#PCDATA)" >
<ENTITY % PAGE "STARTPAGE CDATA #IMPLIED
ENDPAGE CDATA #IMPLIED" >
<!-- 서지정보를 나타내기 위한 ENTITY 정의 부분 -->
<ENTITY % AUTHORINFO "(%TITLE: %AUTHOR: %ORG:, %EMAIL?, (%UNIONAUTHOR: %UNIONORG:, %UNIONEMAIL?))*" >
<ENTITY % TITLE "(%KORTITLE: %ENGTITLE:)" >
<ENTITY % AUTHOR "(%FIRSTNAME: %LASTNAME:)" >
<ENTITY % FIRSTNAME "(#PCDATA)" >
<ENTITY % LASTNAME "(#PCDATA)" >
<ENTITY % ORG "(#PCDATA)" >
<ENTITY % EMAIL "(#PCDATA)" >
<ENTITY % UNIONORG "(#PCDATA)" >
<ENTITY % UNIONEMAIL "(#PCDATA)" >
```

```

<ENTITY % UNIONAUTHOR "(%FIRSTNAME; %LASTNAME:)" >
<ENTITY % FIRSTNAME "(#PCDATA)" >
<ENTITY % LASTNAME "(#PCDATA)" >
<!-- 인용한 논문정보와 인용된 논문정보를 나타내는 부분/-->
<ELEMENT KSCI ( SOCIETYINFO, REFERINFO, JOURNALINFO,
MANAGEINFO)+ >
<!-- 인용한 논문과 인용된 논문을 구분하기 위한 부분 -->
<ATTLIST KSCI CITING #FIXED "CITING"
CITED #FIXED "CITED" >
<!-- 서지정보를 나타내기 위한 부분 -->
<ELEMENT SOCIETYINFO (%TITLE; %AUTHORINFO;
ABSTRACT) >
<ELEMENT KORABST (#PCDATA) >
<ELEMENT ENGABST (#PCDATA) >
<!-- 참고문헌을 정보를 나타내기 위한 부분 -->
<ELEMENT REFERINFO (%AUTHORINFO; %TITLE;%BOOKINFO)+>
<ELEMENT PAGE (#PCDATA) >
<!-- JOURNAL 정보를 표현하기 위한 부분 -->
<ELEMENT JOURNALINFO (%BOOKINFO)+ >
    
```

[표 3-1] 논문정보 요소를 표현하는 DTD 설계

```

<?xml version="1.0" encoding="euc-kr" ?>
<ENTITY % FIRSTNAME "(#PCDATA)" >
<ENTITY % LASTNAME "(#PCDATA)" >
<ENTITY % INFORMATION "%ADDRESS; %PHONE; %EMAIL;
%HOMEPAGE?" >
<ENTITY % ADDRESS "%ZIPCODE; %ADDR;" >
<ENTITY % ZIPCODE "(#PCDATA)" >
<ENTITY % ADDR "(#PCDATA)" >
<ENTITY % PHONE "%TEL; %HANDPHONE%; %FAX?" >
<ENTITY % TEL "(#PCDATA)" >
<ENTITY % HANDPHONE "(#PCDATA)" >
<ENTITY % FAX "(#PCDATA)" >
<ENTITY % EMAIL "(#PCDATA)" >
<ENTITY % HOMEPAGE "(#PCDATA)" >
<!-- 인력정보를 표현하기 위한 부분 -->
<ELEMENT WORKER (INTRO, PAPER)+ >
<!-- 개인에 대한 기본 정보 -->
<ELEMENT INTRO (NAME, JUMIN, HOME, OFFICE >
<ELEMENT NAME (KORNAME, ENGNAME, CHINANAME)>
<ELEMENT KORNAME (%FIRSTNAME:) >
<ELEMENT KORNAME (%LASTNAME:) >
<ELEMENT ENGNAME (%FIRSTNAME:) >
<ELEMENT ENGNAME (%LASTNAME:) >
<ELEMENT CHINANAME (%FIRSTNAME:) >
<ELEMENT CHINANAME (%LASTNAME:) >
<ELEMENT JUMIN (%PCDATA) >
<ATTLIST JUMIN FIRSTDIGIT CDATA #IMPLIED
LASTDIGIT CDATA #IMPLIED >
<ELEMENT HOME (%INFORMATION:) >
<ELEMENT OFFICE (%INFORMATION; OFFICENAME,
OFFICEWORK, ORDERS) >
<ELEMENT CHINANAME (%FIRSTNAME:) >
<ELEMENT CHINANAME (%LASTNAME:) >
<ELEMENT JUMIN (%PCDATA) >
<ATTLIST JUMIN FIRSTDIGIT CDATA #IMPLIED
LASTDIGIT CDATA #IMPLIED >
<ELEMENT HOME (%INFORMATION:) >
<ELEMENT OFFICE (%INFORMATION; OFFICENAME,
OFFICEWORK, ORDERS) >
    
```

```

<ELEMENT OFFICENAME (#PCDATA) >
<ELEMENT OFFICEWORK (#PCDATA) >
<ELEMENT ORDERS (#PCDATA) >
<ELEMENT PAPER (TITLE, JOURNALNAME, UNIONAUTHOR*, ORG,
PUBLICYEAR, VOLUME, NUMBER, PAGE) >
<ELEMENT TITLE (KORTITLE, ENGTITLE) >
<ELEMENT KORTITLE (#PCDATA) >
<ELEMENT ENGTITLE (#PCDATA) >
<ELEMENT JOURNALNAME (#PCDATA) >
<ELEMENT UNIONAUTHOR (#PCDATA) >
<ELEMENT ORGANIZATION (#PCDATA) >
<ELEMENT PUBLICYEAR (#PCDATA) >
<ELEMENT VOLUME (#PCDATA) >
<ELEMENT NUMBER (#PCDATA) >
<ELEMENT PAGE (#PCDATA) >
<ATTLIST PAGE STARTPAGE CDATA #IMPLIED
ENDPAGE CDATA #IMPLIED >
    
```

[표 3-2] 인력정보 요소를 표현하는 DTD 설계

4. 결론

KSCI 체제구축과 데이터의 상호교환 및 공동활용의 기능을 효율적으로 제공하기 위한 가용성·재사용을 고려하여 XML 기반 표준 모델 및 국내 과학기술계 학술지에 수록된 논문들에 대한 인용한 논문정보, 인용된 논문정보, 인력DB정보 사이의 유기적인 관계를 체계적으로 분석하여 KSCI 데이터베이스를 용이하게 구축할 수 있도록 모듈별 DTD를 설계하였다.

본 논문은 KSCI가 가지는 첫째, 국내 기관별, 분야별 과학기술 수준의 파악을 위한 객관적 자료로 활용하고, 둘째, 연구자들의 관심분야 파악과, 전후 인용현황 파악을 통한 관련분야 연구진행 추세 분석 등의 용이함을 제공하고, 셋째, 과학적인 산출물의 정량적인 측정 및 비교를 통하여 현재의 경쟁력 및 상황 파악과 향후 경쟁 결정에 중요한 자료 제공 등의 강점을 활성화하며, 안정적 체제를 구축하기 위한 기반이 되는 것을 목적으로 한다.

5. 참고문헌

- [1] 과학기술부, 2001. "한국과학기술인용문헌 데이터베이스 구축 및 활용에 대한 기획연구"
- [2] 조현양, 2000, 이용기반 데이터베이스 추국 방안에 관한 연구", 정보관리학회지, 17(2),
- [3] Extensible Markup Language (XML) 1.0. <http://http://www.w3.org/XML/>
- [4] XML Path Language (XPath) Version 1.0. <http://www.w3.org/TR/xpath>
- [5] Namespaces in XML, <http://www.w3.org/TR/1999/REC-xml-names-19990114/>
- [6] Document Object Model (DOM) Level 2 Core Specification, <http://www.w3.org/TR/DOM-Level-2-Core/>