

메타데이터의 통합을 위한 스키마 매핑 및 데이터 변환 시스템

이민호*, 이원구*, 최윤수*, 윤화목*, 송사광*, 정한민*

Schema Mapping and Data Conversion System for Integrating Article Metadata

Min-Ho Lee*, Won-Goo Lee*, Yun-Soo Choi*,
Hwa-Mook Yoon*, Sa-Kwang Song*, Hanmin Jung*

요약

본 논문에서는 논문 메타 데이터 특성 분석 연구를 토대로 데이터 변환 방법들을 고안하고 스키마 매핑 및 변환 시스템을 구현한다. 빅 데이터 분석을 위해서는 다양한 시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변환하는데, 현재의 데이터 변환 시스템들은 구문 의존적 문제와 사용의 불편함을 가지고 있다. 본 논문에서 구현된 시스템은 논문 메타데이터 분야에 특화된 시스템으로, 사용하기 쉬운 스키마 매핑 인터페이스를 가지고 있으며 다양한 논문 데이터 구문을 변환할 수 있다. 또한 시스템에 등록되지 않은 새로운 스키마를 가진 데이터가 입력되더라도 시스템의 재 컴파일 없이도 된다. 본 시스템은 사용성 평가를 통하여 시스템 사용성 평균 점수로 89.25점을 받았다.

▶ Keywords : 데이터 통합, 메타데이터 변환, 스키마 매핑, 시스템 인터페이스, 사용성 평가

Abstract

We devise data conversion methods and implement schema mapping and conversion system based on the study on research paper metadata characteristics analysis. Data conversion in unified form from databases of various systems is necessary for big data analysis. Legacy data conversion systems have drawbacks of syntax dependent problem and inconvenience for use. The implemented system, which is specialized system for research paper metadata, has easy schema mapping interface and can convert data with various syntax. In addition to that, Recompiling of the system is not necessary even if new schema which is not preregistered in the system comes in.

• 제1저자 : 이민호 • 교신저자 : 이원구

• 투고일 : 2012. 09. 27, 심사일 : 2012. 10. 09, 게재확정일 : 2012. 10. 22.

* 한국과학기술정보연구원 소프트웨어연구실 (Dept. of Software Research, Korea Institute of Science and Technology Information)

We proved its usefulness by usability evaluation.

- ▶ Keywords : Data Integration Metadata Conversion, Schema Mapping, System Interface, Usability Test

I. 서 론

최근 우리는 데이터의 폭발적 증가로 인한 빅 데이터 시대로 진입하고 있다. 기업들은 정보화 단계를 넘어서 데이터 활용 단계로 들어서고 있으며, 이에 따라 데이터를 분석하여 문제점을 찾아 기업의 경쟁력을 확보하기 위한 데이터 마이닝이 점점 더 중요해지고 있다. 데이터 마이닝¹⁾은 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 것으로, 이를 위해서는 다양한 시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변환하여 저장 및 관리하여야 한다. 다양한 시스템으로부터 생산된 데이터는 구조와 형식이 서로 상이하여 이를 통합하기 위한 여러 가지 노력들이 있었다[1][2][3][4][5].

이러한 연구들에서 공통적으로 수행하는 데이터 통합 방법은 데이터의 구조와 표현법을 정의하는 공통의 스키마를 정하고 기존 시스템의 데이터베이스 스키마를 공통의 스키마와 매핑한 후 데이터를 변환하는 것이다. 하지만 요소가 많은 큰 스키마의 경우에는 수작업을 통해 매핑하기가 어렵거나 실수에 의한 오류가 발생하기 쉽다. 비록 데이터베이스 스키마가 데이터베이스를 설계한 사람에 따라 다르기는 하지만 같은 도메인에서는 비슷한 요소들이 많이 있는 특성이 있기 때문에, 데이터 특성을 분석하면 사용하기 쉬운 스키마 매핑과 변환 방법을 도출할 수 있다. 본 연구에서는 기존에 수행한 논문 메타데이터 스키마의 특성 연구를 기반으로 논문 메타데이터 스키마 매핑 시스템과 데이터 변환 시스템을 설계 및 구현한다[6].

2장에서는 메타데이터 및 스키마 변환과 관련한 연구들을 살펴보고, 3장에서는 기반 연구인 논문 스키마의 특성을 변환 관점에서 자세히 살펴보고 스키마 변환 시스템을 설계한다. 4장에서는 구현된 시스템의 사용자 평가 결과를 분석하고, 끝으로 5장에서 향후 개선 방향에 대해서 살펴본다.

II. 관련 연구

[6]에서는 학술 논문의 메타데이터 스키마를 Priscilla Caplan의 분류에 따라 구문, 의미구조, 내용 규칙 세 측면으로 나누고 특성을 살펴보았다. 세 가지 분류 중 구문은 메타데이터를 이루는 요소들을 어떻게 기계 가독 형식으로 인코딩해야 하는 가를 말하며, 의미구조는 메타데이터 요소 자체의 의미를 말하는 것으로 일반적으로 이름과 정의로서 의미를 표시한다. 내용규칙은 메타데이터 요소의 값이 어떻게 선정되고 표현되는지를 명시한다[7]. [6]의 연구는 10개 주요 해외 출판사의 논문 메타데이터 스키마를 통합 메타데이터의 설계나 상호운용을 위한 데이터 변환 시에 참고할 수 있도록 하기 위함이었다.

메타데이터 변환을 위한 연구로는 XSLT를 이용하여 XML 형식의 데이터를 변환하는 연구가 있었다[8]. 이것은 온톨로지와 어휘 유사도에 기반 하여 단말노드 매칭 후보를 선택하고 경로 유사도를 통해 최종 단말노드의 매칭을 찾는 스키마 매핑 방법이다. 문서의 논리적 구조를 포함하여 플랫폼 독립적이고 계층구조를 가졌다 XML의 특성을 이용한 것으로 콘텐츠 독립적이고 확장이 용이하기는 하나, XML 문서를 변환 대상으로 하는 구문 의존적이라는 단점이 있어 계층구조가 없는 단일 구조의 문서는 변환이 불가능하다. 메타 데이터 변환을 위한 또 다른 시스템으로는 스키마 매핑과 내용 규칙 변환을 위한 함수들을 설정하는 인터페이스를 제공하여 변환 프로그램 모듈을 자동으로 생성하는 시스템²⁾이 있다. 이것은 범용 시스템으로 스키마 매핑 인터페이스와 다양한 변환 컴포넌트를 가지고 있어서 사용자가 이것들을 이용하여 매핑하면 시스템이 자동으로 데이터 변환 프로그램 모듈을 생성해 준다. 사용자는 생성된 변환 모듈을 원하는 응용 프로그램에 넣어 컴파일하여 사용할 수 있다. 이 시스템의 장점은 다양한 구문 및 데이터를 입력으로 받을 수 있으며, 여러 가지 변환 컴포넌트를 통해 다양한 데이터 변환이 가능하다. 또한 사용자 정의 컴포넌트의 등록이 가능하여 시스템의 확장성

1) 데이터마이닝:
<http://ko.wikipedia.org/wiki/데이터마이닝>

2) Altova Mapforce :
<http://www.altova.com/mapforce.html>

이 매우 뛰어난 장점이 있다. 하지만 사용법이 복잡하여 속달에 많은 시간이 소요되며, 데이터 변환을 시스템이 직접 수행하는 것이 아니라 스키마 매핑 후 자동 생성된 프로그램을 모듈을 컴파일하여 변환 응용 프로그램을 작성하여야 하기 때문에, 새로운 스키마를 가진 메타데이터를 추가하려면 새롭게 자동 생성된 모듈로 응용 프로그램을 다시 컴파일 해야 하는 번거로움이 있다.

III. 논문 메타데이터 통합을 위한 스키마 변환 시스템

본 연구에서는 [6]의 연구를 기반으로 논문 메타데이터의 특성을 반영한 메타데이터 변환 시스템을 설계한다. 변환 시스템은 논문 메타데이터에 적합한 변환 방법과 인터페이스를 구현하여 사용하기 쉬우며, 메타 데이터 변환을 직접 수행함으로써 신규 스키마의 추가 시에도 재 컴파일하지 않도록 설계한다.

1. 구문 측면의 특성

[6]에서 연구한 10개의 주요 정보원이 사용하는 구문 형태를 살펴보면 XML, SGML, MARC 등을 일반적으로 많이 사용하나 정보원이 자체적으로 정의한 태그가 달린 텍스트도 사용됨을 알 수 있다. 따라서 논문 메타데이터 변환 시스템은 XML 파서, SGML 파서, MARC파서는 물론이고, 자체 정의한 텍스트를 파싱할 수 있도록 문서의 구조, 태그, 구분자를 입력 받을 수 있는 인터페이스를 갖는 파서를 포함하여야 한다. 우리는 XML, SGML의 파싱을 위하여 일반적으로 많이 사용되는 공개 소프트웨어인 Dom4J³⁾를 사용하고, MARC 파서와 태그가 달린 텍스트를 위한 파서는 자체적으로 개발하였다.

2. 의미구조 측면의 특성

논문 메타데이터에서 공통적으로 사용되는 요소들은 일반적인 논문 검색 사이트의 간략보기 화면에서 볼 수 있는 요소들인 작성언어, 초록, 저자, 페이지, 출판일, 논문제목, 권호, ISSN, 저널제목 등이다.

하지만 공통적인 요소라 할지라도, 초록의 세분화, 축약제목의 존재, 상이한 출판일 의미 등의 차이점이 있고 각각의 정보원들이 가지는 추가적인 요소들도 있다. 의미구조가 다른 여러 요소들을 자동으로 매핑하는 연구들은 아직까지 높은

수준의 정확도를 보여주지 못하기 때문에 본 시스템에서는 사용자가 수동으로 매핑할 수 있도록 한다[9][10][11]. 대신에 사용자 편리를 위하여 스키마를 자동 파싱하여 요소를 제시하고, <그림 1>과 같이 사용자가 이를 편집하여 의미를 확장하고, 관리할 수 있는 기능을 넣었다.

입력된 스키마는 <그림 1>의 좌측에서 볼 수 있듯이 트리 형식으로 편리하게 볼 수 있으며, 특정 엘리먼트를 선택하면 해당 엘리먼트의 이름과 아이디, 속성 등이 우측에 출력된다. 우측 화면에서 속성을 추가 및 삭제할 수 있으며 부모 엘리먼트를 변경하거나 엘리먼트가 반복 출현 가능하도록 수정할 수도 있다.

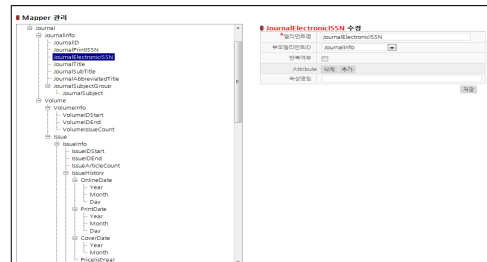


그림 1. 스키마 확장 인터페이스
Fig. 1. Schema Extension Interface

또한 <그림 2>와 같이 트리 기반의 스키마 표출 및 선과 점을 이용한 GUI를 통하여 쉬운 매핑이 가능하다. 스키마 매핑 인터페이스의 좌측은 입수 스키마를 트리 형태로 표현하며, 우측은 변환이 되는 목적 스키마를 트리 형태로 표현한다. 매핑이 되는 2개의 엘리먼트를 양쪽 트리에서 선택하면 인터페이스 상에서 선으로 연결이 되고 변환 방법을 설정할 수 있는 원형의 핸들이 연결선 중간에 생성된다. 핸들을 클릭하면 변환 방법 설정 인터페이스가 나타나는데 여러 가지 변환 방법의 선택, 정규표현식 설정, 규칙 테이블과 스크립트 파일의 선택이 가능하다. 변환 방법에 대해서는 3절에서 설명한다.

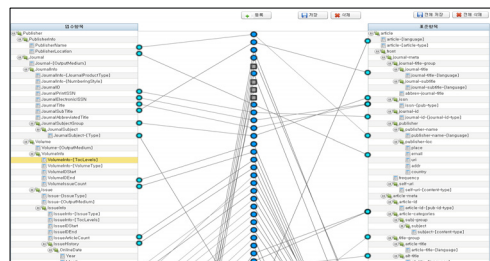


그림 2. 스키마 매핑 인터페이스
Fig. 2. Schema Mapping Interface

3) Dom4J : <http://dom4j.sourceforge.net/>

3. 내용규칙 측면의 특성

〈표 1〉은 선행연구에서 분석한 10개 스키마의 내용 규칙들에 데이터 변환 형태를 추가한 것이다.

논문 메타데이터 스키마의 요소들 중 페이지, 사용 언어, 논문 종류, 권호, 출판일 등에는 일정한 패턴이 존재하여 다음과 같은 네 가지 변환 형태로 분류할 수 있다.

표 1. 내용규칙과 변환 방법
Table 1. Content rules and Conversion Methods

내용규칙	설명	변환형태
페이지 표기	페이지 수를 표기하는 규칙	패턴변환, 불규칙 변환 (페이지 수의 경우)
언어 표기	논문이 기술된 언어를 표기하는 규칙	코드 변환
논문 형태	논문이 학술지인지 프로시딩인지 표기하는 규칙	코드 변환
원문 링크	원문 URL 표기 규칙	무 변환
권호 표기	권호 표기 규칙	패턴 변환
저자 형태	저자의 기여형태 (주저자, 교신저자, 공동 저자 등)	코드 변환
날짜 표기	연,월,일 표기 규칙	패턴 변환, 코드 변환
Class	Class rule for topic, code and etc.	Code Conversion
저자명 변환	저자명 표기 규칙	불규칙 변환

· 무 변환 : 원시데이터가 목적데이터로 변환 없이 그대로 복사되는 형태이다. 제목이나 서론 등과 같은 요소들이 이에 해당한다. 비록 무 변환이라 할지라도 UTF-8에서 EUC-KR으로의 변환 같은 문자 셋의 변환은 필요하다.

· 코드변환 : 국가코드나 언어코드 등과 같은 요소들은 국제 표준이 존재한다. 또한 관리 혹은 데이터 전송의 편리함 때문에 스키마를 정의할 때 코드를 부여함으로써 관리하는 요소들이 많다. 이러한 요소들은 코드 테이블간의 매핑을 통하여 쉽게 변환할 수 있다.

· 패턴 변환 : 데이터가 일정한 스트링 패턴을 가지고 있는

경우이다. 권호, 참고문헌 등의 요소가 패턴 변환에 해당한다. 예를 들어, 권호를 'vol. 1, no. 12'의 형태로 표현하는 스키마가 있다고 할 때, vol과 no은 고정 스트링이고 데이터 인스턴스에 따라 숫자만 변한다. 따라서 다음과 같은 정규표현식을 정의하여 위의 입력을 '12th edition, V.1' 형태로 변환 출력할 수 있다.

- input pattern : (vol. d{2}, no, d{2})
- output pattern : (d{2}th edition, V. d{2})
- 불규칙 변환 : 미묘하게 다른 의미를 가지는 요소는 코드나 패턴으로 정의하기가 어렵다. 예를 들어 원시 스키마에서는 페이지 번호를 뜻하는 요소인데, 목적 스키마에서는 페이지 번호를 뜻하는 요소는 없고 전체 페이지 수를 의미하는 요소만 있는 경우이다. 이렇게 의미가 약간 다른 경우나 의미는 같으나 참고문헌 표기와 같이 데이터 변환이 불규칙하게 일어나는 경우는 불규칙 변환으로 분류한다. 이것은 입력 파라미터 1개와 출력 파라미터 1개를 가지는 간단한 스크립트 프로그램 사용자가 스키마 매핑 시에 정의하도록 하고 데이터 변환 시에 해당 스크립트를 불러 변환한다. 〈그림 3〉은 논문집에서 특정 논문이 들어있는 페이지 범위를 'pp. 13 - 20'과 같은 형태의 요소로 입력되고, 목적 데이터베이스에서는 페이지의 수인 'page 8'이라고 변환할 때의 python 변환 스크립트 예이다. 1라인은 원시 데이터에서 'page' 요소의 값을 읽는다. 2라인에서는 '-'를 기준으로 시작페이지와 끝 페이지를 분리하고 3라인에서 페이지 수를 계산한다. 4라인에서 목적 데이터의 'count' 요소에 계산한 페이지 수를 저장한다.

```

1 : getattr(origin_doc, 'page')
2 : tokens=data.split('-')
3 : count=int(tokens[1])-int(tokens[0])+1
4 : setattr(target_doc, 'count', count)
    
```

그림 3. PageCounter 스크립트의 예
Fig. 3. Example of PageCounter Script

4. 변환 시스템의 설계

〈그림 4〉는 3절에서의 특성 분석 및 변환 방법을 구현한 데이터 변환 시스템의 구조를 나타낸다. 데이터 변환은 크게 스키마 매핑 및 변환 규칙 정의, 데이터 파싱, 요소 변환, 데이터 변환 단계를 거치게 된다.

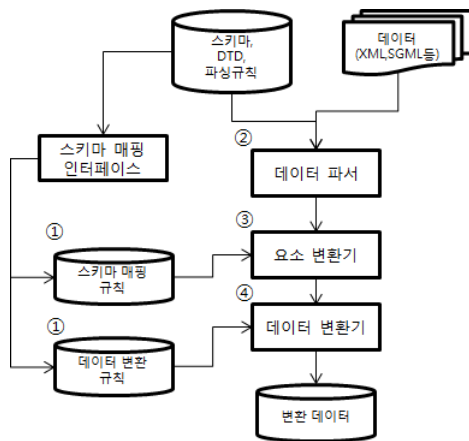


그림 4. 데이터 변환 시스템 구조
Fig. 4. Structure of Data Conversion System

1) 스키마 매핑 및 변환 규칙 정의

원시 데이터의 스키마 혹은 DTD를 스키마 매핑 인터페이스를 통하여 목적 데이터 스키마와 매핑을 시킨다. 이때 3절에서 분류한 데이터 변환 규칙을 설정한다. 설정된 스키마 매핑 규칙과 데이터 변환 규칙은 별도의 규칙 저장소에 저장된다.

2) 데이터 파싱

XML, SGML을 포함한 다양한 형태의 원시 데이터는 미리 저장된 스키마, DTD 혹은 사용자가 정의한 파싱 규칙에 의해 데이터 파서에서 파싱된다. 사용자 정의 파싱 규칙은 스키마나 DTD를 가지고 있지 않은 데이터를 파싱하기 위하여 사용자가 데이터 구조를 분석하여 직접 정의한 것이다.

3) 요소 변환

미리 정의된 스키마 매핑 규칙을 이용하여 요소 변환기는 데이터 파서에서 파싱된 데이터를 요소 변환한다. 이때 데이터의 값은 변하지 않고 스키마 구조상에서의 요소 위치 변환, 요소 병합 혹은 분리 등만 수행한다.

4) 데이터 변환

미리 정의된 데이터 변환 규칙에 의해 요소 변환 단계의 결과 값은 데이터 변환기에 의해 최종 변환 된다. 이때 데이

터 변환 형태에 따라 미리 정의된 코드 테이블, 패턴, 스크립트 등을 이용한다.

〈그림 5〉는 스크립트를 이용한 불규칙 변환을 자세히 설명한 것으로서, 시작 페이지와 끝 페이지를 범위로 갖는 페이지 요소를 페이지 수를 나타내는 요소로 변환하는 예이다. 데이터 변환 규칙 저장소의 내용에는 스키마명, 요소명, 스크립트명을 컬럼으로 가지는 스크립트 매핑 테이블이 있다. 변환되어야 할 요소가 불규칙 변환 특성을 갖는 경우, 스크립트 변환 테이블에서 요소명으로 검색하여 처리할 스크립트를 알아내고 해당 스크립트 파일을 로드한다. 이것은 데이터 변환기의 스크립트 로더 모듈이 수행하며 스크립트 인터프리터가 해당 스크립트를 수행하고 나온 결과를 데이터 변환기의 다른 모듈에게 넘긴다.

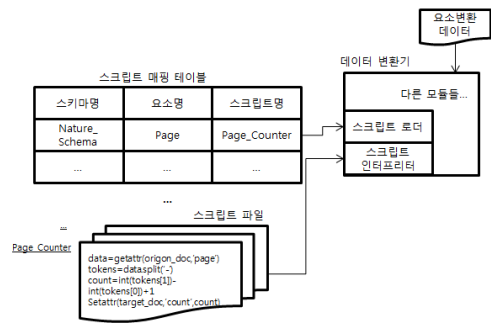


그림 5. 스크립트를 이용한 불규칙 변환
Fig. 5. Irregular Conversion using Scripts

IV. 사용성 평가

구현한 데이터 변환 시스템의 변환 정확도는 사용자가 얼마나 정확히 스키마 매핑과 변환 규칙을 설정하느냐에 달려있다. 요소가 적은 스키마를 설정하는 경우에는 사용자의 실수가 적고 쉽게 수정할 수 있는 반면에, 요소가 많은 스키마의 경우에는 비슷한 작업을 반복하게 되어 사용자의 설정 오류가 발생하기 쉽다. 따라서 사용성 평가를 통해 사용자가 얼마나 쉽게 시스템을 사용할 수 있는지를 측정하였다. 시스템 사용성 평가에 관한 여러 선행 연구를 조사하여 설문항목을 도출하였다. 항목 도출을 위해 조사한 사용성 평가 체크 리스트는 Steven Turner, Information & Design사⁴⁾, 박호환의 웹 사이트 사용성 평가 체크리스트이다[12][13].

4). Usability evaluation checklist for web sites
<http://www.infodesign.com/au/ftp/WebCheck.pdf>

상기 3가지 리스트의 항목에서 계층적 분석방법(AHP: Analytic Hierarchy Process)을 통하여 크게 유효성, 구조적 및 시각적 명확성, 유용성, 지원성, 직관성의 5개 항목을 도출하고 각 항목마다 5개~7개 사이의 질의를 만들었다(표2).

표 2. 사용성 평가 설문 항목
Table 2. Usability Check List

평가 항목	설문내역
유효성	<ul style="list-style-type: none"> · 메뉴 항목 Label이 일관성을 가지고 체계적이며, 기능을 명확히 보여줄 수 있는가? · 모든 페이지에서 일관성 있는 네비게이션 구조를 사용하는가? · 사이트 내에서 현재 네비게이션 중인 위치를 쉽게 파악할 수 있는가? · 현재의 action을 정확히 알려줄 수 있는 인터페이스를 제공하고 있는가? · 스크린 상에서 메뉴나 내용목록(메뉴 바/툴바/네비게이션 바)의 위치는 적절한가? 또한, 메인 페이지와의 링크가 명확히 보여지는가?
구조적 및 시각적 명확성	<ul style="list-style-type: none"> · 화면상에 보여지는 모든 데이터는 의사결정을 위해 사용자가 필요로 하는 것인가? · 사이트의 전반적인 여백이 충분한가? (즉, 텍스트 영역은 "여백"을 가지고 있으며, GUI가 메뉴의 선택이 자유로울 정도의 여유를 제공하는가?) · 사이트 디자인과 레이아웃, 메뉴구조가 간단명료하고 이해하기 쉬운가? · 메인 사이트에 너무 많은 정보를 담고 있지는 않은가? · 표준 GUI 인터페이스 명세를 준수하여 메뉴가 구성되었는가?
유용성	<ul style="list-style-type: none"> · Action 및 data entry에 있어, '되돌리기' 기능을 제공하며, 심각한 결과를 초래할 수 있는 기능기에 대해 되돌리기 기능을 갖고 있는가? · 디자인이 사용자의 이용성을 저하시키지는 않는가? (즉, 지나친 사용성 고려가 사이트 성능에 영향을 미치지 않는가?) · 화면이나 대화상자에 입력할 때, 사용자가 필요로 하는 곳에 커서가 위치해 있으며, 시스템이 자동적으로 메타데이터 입력창을 채워주는가? · 시스템에 특별한 소프트웨어나 특정 하드웨어가 필요한가? · 화면에 나타나는 데이터를 쉽게 이해할 수 있으며, 화면 당 출력량, 배열, 색상, 하이라이트 사용이 적절한가? · 여러 화면들 사이에 일관성(표준화)이 있으며, 또한, 용어를 일관성 있게 사용하는가?
지원성	<ul style="list-style-type: none"> · 에러 메시지는 문제의 근본적 원인을 알려주고 있으며, 사용자가 처한 상황과 대책을 간단명료하고 친절하게 설명하고 있는가? · 에러 메시지는 에러를 수정하기 위한 조치를 제공하고 있는가? · 도움말 메시지가 간단명료하고 친절하며, 사용자가 처한 상황에 맞는가(Context-Sensitive)? · 사이트에 사용자의 피드백을 수용하는 기능이 있으며, 응답이 신속하며, 다양한 피드백 기능(E-mail, 게시판 등)이 제공되는가?

	<ul style="list-style-type: none"> · 데이터 입력 화면과 대화상자 내의 필드는 적절한 기본 값을 가지고 있는가? · 다양한 정렬방법(sort)을 지원하는가? · 검색에 대한 검색결과는 보기 편리하게 나열되고, 결과의 수를 조절할 수 있으며, 그 결과는 신뢰할 만큼 정확한가?
직관성	<ul style="list-style-type: none"> · 메뉴명은 작업 도메인과 일치(카테고리 의미를 쉽게 파악할 수 있도록)하며, 메뉴선택이 논리적인 방법(순차적으로 구성)으로 정렬되어 제공되고 있는가? · 시스템의 상황을 한 눈에 알 수 있고, action을 취할 수 있을 정도의 가시성이 확보되었는가? · 선택된 아이tem(메뉴)이 뚜렷하게 구별되어 지는가? · 모든 화면에 표준 포맷(가관차원의 표준 디자인)이 적용되고 있으며, 관련 필드들이 한 화면에서 보여질 수 있도록 화면 구성이 자연스러운가? · 사이트의 용어사용(페이지 타이틀, 전문용어)에 있어서 일관성을 유지하며 문법적인 오류가 없는가? · www.w3.org/TR/WAI-WEBCONTENT에 의거한 체크리스트나 가이드라인 등 현재의 웹 표준을 준수하고 있는가? · 다양한 플랫폼(UNIX, Windows, Mac)과 브라우저에서 사용가능한가?

평가는 논문정보 유통기관의 메타데이터 관리자 6명과 대학 도서관 사서 4명을 대상으로 시스템을 사용한 후, 평가 설문 항목에 대해 답변하는 방식으로 수행하였다. <그림 6>은 사용성 평가 결과를 히스토그램으로 나타낸 것이다. 시스템 사용성 평균 점수는 산술 평균으로 89.25점이 나왔으며, 표준편차는 4.58이었다. 일반적으로 85점 정도면 사용성이 양호하다고 할 수 있다.

평가자들이 긍정적으로 평가한 항목들은 스키마 매핑 인터페이스의 단순함과 일관적인 조작방식을 들었으며, 기존 시스템들은 제공하지 못하였던 다양한 구문의 처리와 재 컴파일 불편성의 해소도 높이 평가하였다.

- 긍정적 평가항목
 - 인터페이스가 직관적이며 단순함
 - Label의 일관성
 - 조작방식의 일관성
 - 다양한 구문 처리
 - 간편한 데이터 변환 방법

평가자들이 부정적으로 평가한 항목들을 살펴보면 대부분 가시성 확보 부족, 에러 메시지 처리와 네비게이션 관련 항목이었다. 또한 요소가 많은 스키마의 매핑 시 스크롤을 자주 해야 하는 점이 불편한 점으로 평가되었다.

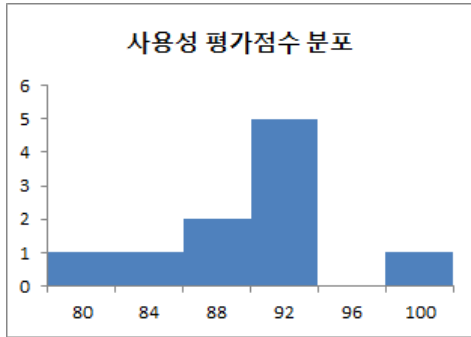


그림 6. 사용성 평가 점수 분포
Fig. 6. Usability Score Distribution Diagram

· 부정적 평가항목

- 시스템 상황의 가시성 확보 부족
- 현재 action에 대한 네비게이션 제공 부족
- action 및 data entry에 대한 되돌리기 기능
- 네비게이션 가시성 확보 부족
- 처한 상황에 맞는 에러 메시지 제공 부족
- 디자인 사용성이 이용성을 저하함.

V. 결론

데이터의 폭발적 증가와 데이터 활용을 통한 이익 창출 요구로 데이터 통합 및 변환이 요소 기술로 매우 중요해지고 있다. 데이터 통합을 위해서는 데이터를 표현하는 스키마의 이질성을 통일하고 데이터 표현 방법을 변환하여야 한다. 본 연구에서는 논문 메타데이터의 특성 분석을 통하여 기존 연구가 가진 데이터의 구분 의존적인 문제와 사용하기 어려운 단점을 해결한 스키마 매핑 및 변환 시스템을 제안하였다. 데이터 특성 분석을 통해 네 가지 변환 방법을 분류하고 이를 구현하였으며, 사용성 평가를 통해 사용하기 쉬움을 입증하였다. 향후 더 편리한 데이터 변환을 위하여 자동 스키마 매핑 연구를 수행할 예정이다.

참고문헌

[1] Margaret St. Pierre, "Issues in Crosswalking Content Metadata Standards", NISO White Paper, 1998.
 [2] Erhard Rahm, and Phillip A. Bernstein, "A survey of approached to automatic schema

matching", The VLDB Journal, Vol.10, pp.334-350, 2001.
 [3] Metadata Registry, <http://metadata-stds.org/11179/>
 [4] Md. Sumon Shahriar, and Jixue Liu, "Towards the Preservation of Referential Constraints in XML Data Transformation for Integration", International Journal of Database Theory and Application, Vol.3, No.2, 2010.
 [5] Won-Goo Lee, Hwa-Mook Yoon, and Won-Kyung Sung, "An Efficient Management System on Digital Contents: Literature Data", Pceedings of the International Conference on Convergence Content, Nara, Japan, 2010 December.
 [6] Min-Ho Lee, Won-Goo Lee, Hwa-Mook Yoon, Sung-Ho Shin, and Jae-Jeol Ryou, "Comparison and Analysis of Science and Technology Journal Metadata", Vol. 11, No. 9, pp. 515-523, 2011.
 [7] Priscilla Caplan, and Dong-Geun Oh, "The Understanding of Metadata", Tae-il publishers, 2004.
 [8] Kyung-Ho Lee, and Jun-Seoung Lee, "XMLSchema Matching based on Ontology Renewal for XML Document Conversation", KIISE Journal, Vol. 33, No. 7, pp. 727-740, 2006.
 [9] Md. Sumon Shahriar, and Jixue Liu, "Constraint-Based Data Transformation for Integration: An Information System Approach", International Journal of Database Theory and Application, Vol. 3, No. 1, 2010.
 [10] Jae-Hong Kim, and Sang-Jo Lee, "An Algorithm for Ontology Merging and Alignment using Local and Global Semantic Set", IEEK Journal, Vol. 41, No. 4, pp. 23-30, 2004.
 [11] Tae-jin Ha, "Realization of XSLT-based Schema Mapper using RCP", thesis paper, Seoul National University of Science and Technology, 2009.
 [12] Turner, Steven, "The HEP test for grading web site usability, Computers in Libraries", Complete coverage of library information technology, Vol.10, pp.37-39, 2002.

[13] Ho-Wan Gwak, Ji-Eun Gwak, Su-jin Kim, and Jeong-Mo Lee. "Usability Survey on Domestic Website Design: Questionnaire Survey and Heuristic Evaluation", Cognitive Science, Vol.1, pp.33-45, 2000.

컴퓨터공학과 공학박사
 현 재 : 한국과학기술정보연구원
 책임연구원
 관심분야 : 데이터베이스, 정보검색
 Email : hmyoon@kisti.re.kr

저 자 소 개



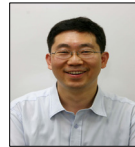
이 민 호
 2000 : 충남대학교
 컴퓨터공학과 공학석사
 2006 : 충남대학교
 컴퓨터공학과 박사수료
 2000 : 데이콤 중앙연구소 연구원
 현 재 : 한국과학기술정보연구원
 선임연구원
 관심분야 : 빅데이터, 정보보호
 Email : cokeman@kisti.re.kr



송 사 광
 1997 : 충남대학교
 통계학과 이학사
 1999 : 충남대학교
 컴퓨터공학과 이학석사
 2011 : KAIST 전산학과 공학박사
 현 재 : 한국과학기술정보연구원
 선임연구원
 관심분야 : 텍스트마이닝, 시맨틱웹
 Email : esmallj@kisti.re.kr



이 원 구
 2000 : 한남대학교
 컴퓨터공학과 공학석사
 2005 : 한남대학교
 컴퓨터공학과 공학박사
 현 재 : 한국과학기술정보연구원
 선임연구원
 관심분야 : 빅데이터, 과학데이터
 Email : wglee@kisti.re.kr



정 한 민
 2003 : 포항공과대학교
 컴퓨터공학과 공학박사
 현 재 : 한국과학기술정보연구원
 책임연구원, 과학기술연합
 학원대학교 겸임교수
 관심분야 : 시맨틱웹, 자연어처리
 Email : jhm@kisti.re.kr



최 윤 수
 1993 : 충남대학교
 컴퓨터공학과 공학사
 1995 : 충남대학교
 컴퓨터공학과 공학석사
 현 재 : 한국과학기술정보연구원
 선임연구원
 관심분야 : 빅데이터, 정보검색
 Email : armian@kisti.re.kr



윤 화 목
 1992 : 서울산업대학교
 전자계산학과 공학사
 1997 : 공주대학교
 전자계산학과 공학석사
 2008 : 배재대학교