韓國컴퓨터情報學會 論文誌
第16卷 第11號, 2011. 11.

2011-16-11-2-4

과학기술정보콘텐츠 통합관리시스템 구축을 위한 데이터 마이그레이션 모델 수립 및 적용 사례

신 성 호*, 이 민 호*, 이 원 구*, 윤 화 묵*, 성 원 경*, 김 광 영*

A Data Migration Model and Case Study for Building Management System of Science and Technology Contents

Sung-Ho Shin*, Min-Ho Lee*, Won-Goo Lee*, Hwa-Mook Yoon*, Won-Kyung Sung*, Kwang-Young Kim*

요 약

국내 데이터베이스 구축 시장은 작년 기준으로 3조 6,633억원 규모로 추정된다. 데이터베이스 산업의 지속적인 성장과 맞물려, 데이터 마이그레이션의 중요성도 날로 높아지고 있다. 고객 위치를 찾아주는 g-CRM이나 고객에게 상품을 추천하는 맞춤 설계 기능 등은 모두 고객 데이터베이스, 상품 데이터베이스, 지리정보 데이터베이스 등이 결합되어야 구현 가능한 서비스들이다. 스마트 비즈니스의 핵심 인프라도 통합되고 완전하며 신뢰할 수 있는 데이터베이스라고 할 수 있다. 이러한 데이터베이스의 중요성에도 불구하고, 데이터 마이그레이션 및 통합 과정의 효율적인 수행 방안이나 데이터 검증 방법에 대하여는 구체적인 연구가 부족한 실정이다. 본 연구에서는 다양한 타입으로 존재하는 과학기술분야 콘텐츠들을 대상으로 데이터 마이그레이션을 위한 모델을 설계하였고, 모델을 기반으로 실제 데이터 마이그레이션 작업을 수행한 결과를 제시하였다. 아울러 데이터 마이그레이션 결과에 대한 검증을 위해 데이터베이스의 완전성, 데이터값의 일관성, 관계의 일관성을 검증하였고, ANSI/ASQ Z1.4-2003에서 제시된 샘플링 검사 기법도 적용하였다. 결과적으로 모델 수립에 의한 체계적인 데이터 마이그레이션 수행은 데이터베이스 정합성 및 데이터값의 정확성에 영향을 미치고, 고품질의 데이터베이스를 유지하기 위한 필수 요소라 할 수 있다.

▸ Keyword : 데이터 마이그레이션, 과학기술정보 콘텐츠, 데이터 검증, ANSI/ASQ Z1.4-2003

Abstract

The domestic market of database in Korea is estimated to be over 3.663 trillion won. The data migration is getting to be more important along with the continuous growth of the database

• 제1저자 : 신성호 • 교신저자 : 김광영
• 투고일 : 2011. 10. 05, 심사일 : 2011. 10. 19, 게재확정일 : 2011. 10. 24.
* 한국과학기술정보연구원 정보기술연구실(Dept. of Information Technology Research, KISTI)

industry. g-CRM and private recommending function are examples of the service that can be given through coupling among customer database, product database, geographic information database, and others. The core infrastructure is also the database which is integrated, perfect, and reliable. There are not enough researches on efficient way of data migration and integrating process and investigation of migrated data though trends of database in IT environment as above. In connection with this issue, we have made a model for data migration on scientific and technological contents and suggest the result of data migration process adapting that model. In addition, we verified migration's exhaustiveness, migration's consistency, and migration's coherence for investigation of migrated data and database. From the result, we conclude data migration based on proper model has a significant influence on the database consistency and the data values correctness and is essential to maintain high qualified database.

▸ Keyword : Data Migration, Scientific and Technological contents, Data Verification, ANSI/ASQ Z1.4-2003

I. 서 론

2011년 8월 현재, 애플의 앱스토어에는 35만개가 넘는 앱이 등록되어 있고 지난 1월 기준으로 사용자 다운로드 수가 100억건을 돌파했다[1]. 구글, 삼성전자 등도 자체 앱스토어를 운영하고 있으며, 많은 앱이 올라오도록 유도하는 다양한 정책과 마케팅을 펼치고 있다. 스마트폰용 애플리케이션은 하나의 콘텐츠로 인식되고 있으나, 실제 내용을 보면 데이터베이스가 중요한 역할을 하고 있다. 지도 서비스와 관련된 애플리케이션은 지리정보 데이터베이스를 기반으로 구성된다. 미국에서 선풍적인 인기를 끌고 있는 Foursquare의 사례 역시 상점 데이터베이스 구축과 지속적인 업데이트가 전제되지 않으면 성립할 수 없는 비즈니스 모델이다. 고객 위치를 찾아주는 g-CRM이나 고객에게 상품을 추천하는 맞춤 설계 기능 등은 모두 고객 데이터베이스, 상품 데이터베이스, 지리정보 데이터베이스 등이 결합되어야 구현 가능한 서비스들이다. 결국 스마트 비즈니스의 핵심 인프라도 통합되고 완전하며 신뢰할 수 있는 데이터베이스라고 할 수 있다.

국내 데이터베이스 구축 시장은 작년 기준으로 3조 6,633조원 규모로 추정된다. 데이터베이스 산업의 지속적인 성장과 맞물려, 데이터 마이그레이션의 중요성도 날로 높아지고 있다[2]. 데이터 마이그레이션이란 새로운 정보시스템을 개발 혹은 개선의 필요로 인하여 구 정보시스템 혹은 그 외의 방법으로, 축적된 과거 자료를 새로운 정보시스템에서 운용 가능하도록 체계적으로 이동시키는 일련의 프로세스를 의미한다.

아울러 데이터 마이그레이션의 목표는 현재의 데이터베이스 시스템 운영 환경을 이해하고, 사용 중인 데이터베이스 관리 시스템의 특성을 파악하여, 최적의 데이터베이스 구조를 유지하며, 신 시스템이 최상의 성능을 발휘하도록 하는 것이다. 이러한 중요성에도 불구하고, 실제 데이터 마이그레이션 및 통합 과정의 효율적인 수행 방안이나 데이터 검증 방법에 대하여는 구체적인 연구가 부족한 실정이다.

본 연구에서는 과학기술분야 콘텐츠들을 대상으로 데이터 마이그레이션을 위한 모델을 설계하였고, 실제 데이터 마이그레이션 작업에 적용한 결과를 제시하였다. 아울러 데이터 마이그레이션 결과에 대한 검증을 위해 데이터베이스의 완전성, 데이터값의 일관성, 관계의 일관성을 검증하였고, ANSI/ASQ Z1.4-2003에서 제시된 샘플링 검사도 적용하였다.

II. 관련 연구

1. 국내 데이터베이스 산업분류 및 동향

국내 데이터베이스 구축 시장은 <그림1>과 같이 2009년 3조 4,175억 원 규모의 시장을 형성한데 이어, 2010년에는 7.2% 성장한 3조 6,633억 원의 시장을 형성하는 것으로 나타났다.

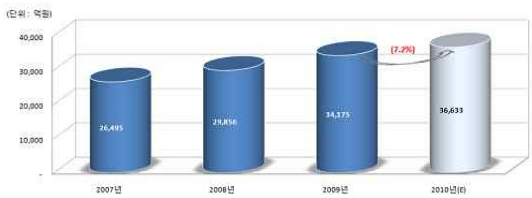


그림 1. 데이터베이스 구축 시장 규모 추이('07~'10년)
Figure 1. The size of DB development market('07~'10)

데이터베이스 구축 시장의 세부 부문별 규모는 <표1>에서 볼 수 있듯이, 자료처리 부문이 2010년 3,831억 원으로 전년 대비 3.8% 성장하였으며, 구축 시장의 89.5% 정도의 비중을 차지하고 있는 시스템구축 부문은 3조 2,802억 원으로 전년 대비 7.6% 성장하여 데이터베이스 구축 시장의 성장을 견인한 것으로 나타났다[2].

표 1. 데이터베이스 구축 시장 현황('07~'10년)
Table 1. Variation of DB development market('07~'10)

(단위 : 억 원)

중분류	2007년	2008년	2009년	2010년(E)	증감률 (09-10)	CAGR (07-10)
자료처리	3,424	3,459	3,690	3,831	3.8%	3.8%
시스템구축	23,071	26,397	30,485	32,802	7.6%	12.4%
합계	26,495	29,856	34,175	36,633	7.2%	11.4%

2010년 3조 6,633억 원의 시장 규모를 형성하고 있는 데이터베이스 구축 시장은 2011년에는 4조 원대 시장 규모로 진입하면서 2013년에는 4조 7,011억 원의 규모에 이를 것으로 전망된다[2]. 이러한 데이터베이스 구축 시장의 성장은 시장 비중이 높은 시스템구축 부문의 영향이 크지만, 시장 규모가 작은 자료처리 부문의 시장도 디지털 자료에 대한 끊임없는 수요로 지속적인 상승이 이어질 것으로 기대된다.

데이터베이스 산업의 지속적인 성장과 맞물려, 데이터 마이그레이션의 중요성도 날로 높아지고 있다. IT 도입기에 체계성 없이 시스템을 개발해서 기업의 경쟁력 향상과 급변하는 사회 구조에 맞춰 IT 예산을 대폭 투자했던 기존과는 달리 최근에는 IT 투자를 최대한 줄이면서, 기업의 경쟁력을 강화하는 방향으로 가고 있다.

2. 데이터 마이그레이션 개념

데이터 마이그레이션에 대한 번역어로는 다양한 용어가 사용되고 있다. 정보시스템 구축 분야에서는 '(데이터) 이행'이란 표현을 주로 사용해 왔다. 문헌정보 분야에서는 자료관리 시스템과 관련하여 자료 변환이나 자료 전환, 변환이란 용어가 주로 쓰이고 있으며, 이 또한 통일되지 못하게 사용되고 있다[3]. 이 중 '이행'이 그나마 적절하다고 볼 수 있지만, 의미 전달이 불분명할 수 있으므로, 본 연구에서는 데이터 마이그레이션이란 용어를 그대로 사용하고자 한다.

데이터 마이그레이션의 의미는 <표2>에서 정리하였듯이, 크게 두 가지로 생각해 볼 수 있다[4]. 일반적인 사전 및 컴퓨터 과학 분야에서 이야기하고 있는 데이터 마이그레이션의 정의는 "과거의 시스템 및 운영 환경에서 사용되는 데이터를 더 나은 시스템 및 운영환경으로 이전하는 프로세스"이다. 또 하나는 기록보존 관점에서의 정의이다. 데이터 마이그레이션이란 시간이 흐름에 따라 기록매체, 포맷 시스템이 대체되거나, 유실되거나, 노후화 될 수 있는 상황에서, 기록물로의 지속적인 접근 위해 다음 세대의 운영 체제, 저장 매체, 포맷, 시스템으로 이전의 데이터를 이관하는 프로세스로 정의된다.

종합해보면, 데이터 마이그레이션은 어떤 목적에 의해 소스의 개체들을 목표 장소로 체계적으로 이동시키는 일련의 프로세스로 정의가 가능하다.

표 2. 데이터 마이그레이션의 정의
Table 2. Definitions of data migration

출처	데이터 마이그레이션 정의
whatis.com	하나의 운영 환경으로부터 다른 운영환경으로 이전하는 프로세스
wikipedia	데이터 저장 장치들 사이에서 정보를 이동하는 것
NTCA	과거의 데이터를 새로운 시스템으로 불러오는 프로세스
IFLA	주기적으로, 하나의 하드웨어/소프트웨어 환경으로부터 다른 것으로, 혹은 하나의 컴퓨터 기술로부터 이어지는 기술
ISO	시간이 흐름에 따라 시스템과 기록매체가 다른 것으로 대체되거나, 유실되거나, 노후화될 수 있는 상황에서 기록물의 지속적인 접근 가능성을 보장하기 위해 기록물을 하나의 시스템 또는 저장 매체로부터 다른 것으로 이전하는 프로세스

3. 관계데이터베이스의 XML 변환

3.1. 관계데이터베이스와 XML 스키마

관계데이터베이스테이블을 XML(Extended Markup Language) 파일로 변환하는 주요 목적은 전자상거래나 기업간(Business-To-Business)의 웹 응용에서 데이터베이스 파일을 효율적으로 이용하기 위한 것이다[5]. 이를 위해 관계형 구조의 테이블과 태그를 가지는 계층형 파일구조인 트리구조로 변형하여 각각의 요소를 XSLT(Extensible Stylesheet Language Transformations)를 이용하여 원하는 형태로 쉽게 편집 변환하여 이를 다시 HTML 형태로 웹상에 뿌려주고 반대로 웹상에서 XSLT 형태로 변환된 XML 파일이 전송되면 이를 다시 관계형 데이터 테이블로 변환하여 데이터베이스에 저장할 수 있도록 하는 것이다. 이렇게 함으로서 기존의 SQL 서버나 Access 등의 데이터베이스를 웹과 함께 더 효율적으로 HTML 형식으로 연동하여 사용할 수 있다. 관계테이블과 XML 문서의 다른 점은 XML과 데이터베이스는 설계 목적부터 다르다. 데이터베이스의 설계 목적은 데이터를 효과적으로 저장하고 데이터의 무결성, 백업, 복구 등의 관리를 담당하고, XML 문서는 사용자와 어플리케이션 사이와 사용자 간, 어플리케이션 간에 데이터 교환을 위해 효과적으로 사용된다.

3.2. 관계테이블의 XML 변환 툴

조직의 관계데이터베이스를 전자상거래, 기업 간(Business-To-Business), 웹 응용에서 사용하기 위해 XML파일로 변환하는 것은 XML 변환기능을 내장하고 있는 전문데이터베이스용 프로그램이면 모두 가능하다. 예를 들면, SQL Server 2008이나 Access 2007 등 데이터베이스 전문프로그램에 XML 변환 기능들이 지원되고 있다. Xpath, XDR Schema, XSL Tranformation, HTTP, OLE 데이터베이스 등과 같은 새로운 규칙을 가지고 XML 자료와 관계데이터베이스를 매핑하는데 사용한다. 그리고 XML 데이터의 검색, 삽입, 수정도 가능하다. 이들 전문 데이터베이스프로그램과 XML을 사용하여 비즈니스 데이터를 웹으로 이동하여 종업원, 고객, 파트너 등을 위한 종합적 업무자료로 이용하는 것은 아주 유용하다. 이 외에 <표3>에서 제시된 것처럼 다른 상업용 소프트웨어 중에서도 특정 파일 포맷을 XML 파일로 변환시켜 주는 소프트웨어가 많이 출시되어 있다.

표 3. XML 변환 소프트웨어들의 변환 유형
Table 3. Converting types of S/W converting to XML

소스파일형태	목표파일형태
PDF	XML
XLS	XML
데이터베이스F	XML
데이터베이스F	XML/CSV/TXT/HTMl/RTF/SQL
PDF	TXT/XML/RTF/XML/HTML
데이터베이스X	데이터베이스F/CSV/XLS/XML
DAT/데이터베이스F/M/ADT	TXT/CSV/HTML/XML/XLS/SQL
XLS/WK1/123/ODS	TXT/CSV/HTML/XML
CSV/TEXT	XML
PDF/HTML/TXT	DOC/XML/HTML/TXT
DOC/DOCX/XML/HTML	DOC/HTML/MHTML/RTF/TXT

그러나 위와 같은 두 가지 방법은 다음과 같은 문제점을 가진다. 데이터베이스 전문프로그램에 의한 XML 변환은 데이터 변환 시 적용되어야 할 별도의 규칙이 있거나, 1:1이 아닌 1:N, N:1, N:M의 변환이 필요한 경우에는 작업이 효율적이지 못하거나 데이터 마이그레이션이 불가능한 단점을 가지고 있다. 또한 소프트웨어 툴을 사용한 XML로의 변환 방법은 조작이 간단하여 파일 형태의 소스 데이터를 XML로 변환하는 작업에는 유용하지만, 관계데이터베이스 내의 데이터들을 파일 형태로 변환(1차)시킨 후, 다시 XML로 변환(2차)하는 과정을 거치므로, 관계데이터베이스 내의 데이터들을 바로 XML로 변환(1차)시키는 방법보다 검증 과정을 한번 더 거쳐야 하는 번거로움과 변환 과정에서 에러가 발생할 위험도 있다.

III. 데이터 마이그레이션 모델 설계

데이터 마이그레이션과 관련하여 표준처럼 사용되는 방법론이나 정해진 절차는 따로 없다. 대부분 데이터 마이그레이션을 위한 관련 툴 소프트웨어를 데이터베이스 시스템에 포함하여 공급하거나 별도의 컨설팅을 통하여 데이터 마이그레이션을 수행하기 때문에 많은 시간과 비용을 투자해야 하는 부담을 갖고 있다. 또한 데이터 마이그레이션 툴 소프트웨어를 제공하더라도 동일한 스키마 구조 상태에서 다른 데이터베이스 시스템간의 단순한 데이터 마이그레이션 방법과 절차만을 제공하고 있다. 따라서 데이터베이스 간의 데이터 마이그레이션 목적 및 목표를 달성하기 위한 체계적이고, 효과적인 모델 설계가 필요하며, 설계된 모델이 다른 조직이나 콘텐츠에 사용되기 위해서 데이터 마이그레이션 모델에 대한 검증 작업이 필요하다.

실무에서 주로 사용되는 데이터 마이그레이션 절차는 분석 및 설계, 데이터 추출, 데이터 정제, 저장, 검증 순으로 진행된다. 일부 문헌에서는 소스(Source) 데이터의 분석, 데이터 클린징, 데이터 추출 및 가공, 목표(Target) 시스템으로의 데이터 적재, 적재 데이터 검증, 데이터 활용 단계를 제시하고 있다[6].

데이터 마이그레이션이 빈번하게 발생하는 금융 비즈니스에서의 데이터 마이그레이션을 위해 일반적인 소프트웨어 개발 프로세스와 달리 분석, 설계, 구현, 검증의 4단계의 프로세스로 구성한 데이터 마이그레이션 아키텍처에 대한 연구 사례도 제시되고 있다[7].

이처럼 데이터 마이그레이션의 절차는 마이그레이션 수행 환경에 따라 조금씩 차이가 발생할 수 있기 때문에, 본 연구에서는 데이터 마이그레이션 절차를 포괄적인 개념에서 접근하여 <그림2>와 같은 데이터 마이그레이션 모델을 제시하고자 한다. 데이터 마이그레이션도 시스템개발 프로젝트와 유사한 측면이 있기 때문에, 분석단계(환경분석, 데이터분석), 설계단계(흐름설계), 구현단계(기능구현), 테스트(검증) 단계로 구분하였고, 세부적으로 10가지 세부 업무로 나누어진다.

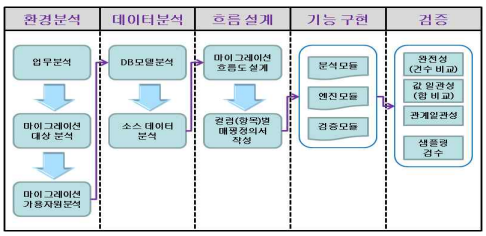


그림 2. 과학기술정보콘텐츠 데이터 마이그레이션 모델
Figure 2. A model of data migration for scientific and technological contents

1. 환경 분석

데이터 마이그레이션 수행을 위해서는 우선 현행 정보시스템 및 데이터베이스의 구조를 분석해야 한다. 정보시스템 구조 분석은 IT를 활용하는 각 업무 영역들에 대한 분석으로 볼 수 있다. 한 조직에서 수행되는 대부분의 업무들은 전사적으로 연관되어 있으므로, 전체 데이터 마이그레이션 대상 업무 영역을 파악할 필요가 있다. 업무 영역의 분석은 논리적으로는 업무 도메인 분석으로 해석할 수 있으며, 물리적으로는 각 업무 영역에서 활용하는 현행 정보시스템의 분석으로 볼 수 있다.

현행 정보시스템의 전체적인 구조 분석이 완료되면 데이터 마이그레이션 대상을 분석해야 한다. 마이그레이션 대상은 체계적으로 정보화되어진 데이터가 주를 이루나, 실제 체계적으로 정비되어 있지 않은 자료(이미지, 첨부파일 등)도 마이그레이션 대상에 포함되므로, 분석 시 항상 염두에 두어야 한다. 즉, 어떤 데이터 형식이 어떻게 현행 정보시스템에 저장되어 있는지를 파악하여야 한다. 또한 현행 정보시스템의 마이그레이션 대상 데이터들의 크기 계산도 함께 이루어져야 한다.

위 두 가지 항목의 분석이 완료되면, 정보시스템에서 활용 가능한 자원에 대하여 분석하여야 한다. 데이터 마이그레이션은 짧은 시간에 대용량 데이터를 이동하는 프로세스를 가지며, 마이그레이션을 위한 필요 자원에 대비해 항상 부족할 수 밖에 없다. 구축 정보시스템의 가용 가능 자원을 분석함으로써, 현행 정보시스템에서 분석한 마이그레이션 대상 자료를 부족한 정보시스템의 가용 자원을 최대한 활용할 수 있는 방안이 수립되어진다.

따라서 분석 단계에서의 과정이 제대로 수행되지 못할 시 실제 데이터 마이그레이션 수행 시 자원의 비효율적 남용 및 부족과 같은 현상이 발생하며, 이러한 것은 전체적인 데이터 마이그레이션 수행의 효율성을 떨어뜨리는 원인이 된다.

2. 데이터 모델 및 데이터 분석

분석 단계의 다음은 데이터 마이그레이션 대상의 모델 분석이다. 정보시스템은 변화하는 비즈니스에 의해 그 논리적, 물리적 모델 구조가 변경되어지며, 따라서 마이그레이션 대상 데이터들의 구조에도 변화가 발생한다. 이러한 변화를 파악하기 위해서는 마이그레이션 대상의 모델 분석이 필요하다. 데이터 마이그레이션은 데이터 모델의 주체는 아니나, 변경되어진 모델을 이해하지 못할 시 마이그레이션 과정에서 많은 시행착오를 가져올 수 있다. 즉, 데이터 매핑 정의 시 올바르지 않은 매핑을 진행할 수 있으며, 이러한 잘못된 매핑으로 데이터 마이그레이션을 수행하는 오류를 범하게 된다. 많은 프로젝트에서 데이터 마이그레이션을 단순히 데이터의 전송 관점에서 접근하므로 이러한 오류가 많이 발생하고 있다. 그러나 데이터 마이그레이션의 분석 단계에서 데이터 모델 변경에 대한 부분을 정확히 분석하고, 프로젝트 수행자들에게 공유되어진다면, 효율성을 더욱 극대화 할 수 있다.

다음으로 소스 데이터를 분석해야 한다. 소스 데이터에는 현행 정보시스템의 운영 과정에서 발생한 잘못된 데이터들과 목표 데이터로의 변환이 불가능한 데이터도 존재할 가능성이 있다. 이러한 데이터들을 분석 과정에서 도출하여, 목표 데이터로의 변경 및 데이터 정제 작업 또는 대상 제외 여부를 판단해야 한다. 만약 정제 작업의 대상이라면, 데이터 마이그레이션 수행 전에 반드시 현행 데이터를 변경해야 하며, 제외 대상이라면, 현행 데이터를 삭제해야 한다. 이 과정은 매우 번거롭고 까다로운 과정이며 의사결정이 필요한 부분이므로, 현행 정보시스템 운영자 및 현업 담당자의 확인 절차가 반드시 동반되어야 한다.

3. 흐름 설계

데이터 마이그레이션 분석 단계가 완료되면, 데이터 마이그레이션 수행을 위한 흐름을 설계해야 한다. 데이터 마이그레이션 흐름 설계는 구 정보시스템의 데이터가 신 정보시스템의 데이터로 변환되어 적재되어지는 과정을 흐름도로 설계하는 것이다. 일반적인 테이블 대 테이블 마이그레이션은 단순히 코드 변환의 과정만 설계하지만, 실제 신 정보시스템의 구축 목적이 단순 테이블 코드 값의 변경 차원을 벗어나, 업무 비즈니스의 변경 흐름까지도 반영되어야 한다.

데이터 매핑흐름도에는 목표 데이터를 기준으로 하여 필요 소스 데이터의 추출 및 변환/가공, 적재의 로직과 업무에 따른 비즈니스 흐름이 충분히 반영되어야 한다.

그 다음은 분석 단계에서 분석한 데이터를 기반으로 한 소스 데이터와 목표 데이터를 데이터 표준에 맞게 정의(테이블, 컬럼, 코드 파일 등)하고 해당 데이터들의 매핑(Data Mapping) 관계를 기술해야 한다.

먼저 목표 데이터에 대한 소스 데이터를 정의한다. 소스 데이터는 구 시스템의 테이블, Flat 파일, 스프레드시트 등과 같은 다양한 종류가 될 수 있으며(소스 원천 정의), 해당 데이터가 목적 테이블과 1:1 또는 M:1, 1:M, N:M 관계를 가질 수 있다(대상 컬럼 정의). 소스 데이터의 정의가 완료되었으면, 목적 테이블 데이터와의 관계에 대하여 기술한다. 데이터 마이그레이션에서 대부분의 데이터의 관계는 1:1의 단순 이동(Move)으로 표현되어 질 수도 있으나, 표준화 변경에 의한 코드 변환 및 데이터 변경, 파일 전송 등이 발생할 수도 있

다. 이러한 신, 구 데이터들의 관계를 데이터 마이그레이션 작업 이전에 정확히 명시하여야 효율적인 데이터 마이그레이션을 보장할 수 있다. 또한 형태가 다른 데이터 타입 간의 스키마 매핑이 존재하는지 분석이 되어야 한다. 선행연구에서 살펴보았듯이, 소스 데이터의 형태는 PDF, XLS, HTML, DB 등 다양한 형태로 존재할 수 있고, 목표 데이터의 형태도 동일하게 여러 가지가 될 수도 있다.

<표4>는 데이터 마이그레이션 매핑 정의에 필요한 기본적인 기술항목을 나열한 것이다. 원천 소스의 구분은 소스 항목에 기술하며, 매핑방법은 목표 항목에 기술한다.

표 4. 데이터 마이그레이션 매핑 정의 항목
Table 4. Migration mapping elements

분석항목	소스	목표
테이블명	○	○
컬럼명	○	○
컬럼타입	○	○
PK여부	○	○
컬럼길이	○	○
포맷	○	○
매핑방법	×	○

4. 구현 및 검증

설계서에 기초한 데이터 마이그레이션 구현은 기존의 상용 마이그레이션 도구를 구입하는 방법과 요구사항에 맞게 새로운 도구를 개발하는 방법이 있다. 기존의 상용 도구를 구입하여 사용하는 방법도 마이그레이션의 요구사항을 반영하기 위해서는 일부분 커스터마이징이 필요하다. 두 가지 방법 중 기간 및 예산, 요구사항 반영 가능 정도를 분석하여 신중히 선택할 필요가 있다.

일반적으로 데이터 마이그레이션 도구는 매핑정의서에 기술된 내용을 바탕으로 소스 데이터를 목표 데이터로 이관하는 작업을 수행한다. 소스 데이터베이스로부터 데이터베이스의 속성정보인 메타데이터를 생성하고 메타데이터로부터 해당 데이터베이스의 운영환경을 분석하여 분석 레포트를 생성하는 데이터베이스 분석 모듈, 소스 데이터베이스로부터 스키마, 데이터 및 비즈니스 로직을 순서대로 매핑에 의해 실제적인 변환을 수행하는 데이터 마이그레이션 엔진 모듈, 마이그레이션 후 데이터 검증 모듈로 구성된다.

데이터베이스 분석 모듈은 경우에 따라 구현 대상이 될 수도 있고, 구현하지 않을 수도 있다. 수작업으로도 데이터베이스 분석이 가능하며, 경우에 따라서는 수작업 분석이 더 필요할 수도 있기 때문이다. 엔진 모듈은 소스 데이터의 추출, 변환/가공, 목표 데이터 적재의 역할을 수행한다.

데이터 마이그레이션 수행 후, 마이그레이션 된 데이터에 대한 검증이 필요하다. 검증은 데이터 마이그레이션의 품질을 측정하는 방식이며, 애플리케이션을 통해 제공되는 결과물이 데이터 마이그레이션 전과 후 동일한 모습대로 보여지도록 보장한다. 성공적인 데이터 마이그레이션은 아래의 세 가지 조건을 충족시켜야 한다[8].

첫째, 완전성(migration's exhaustiveness)

마이그레이션의 대상이 되는 소스 데이터베이스의 모든 데이터들은 남김없이 목표 데이터베이스로 이동되어야 하며, 오직 거기에만 존재해야 한다.

둘째, 데이터값의 일관성(migration's consistency)

마이그레이션 된 데이터들은 목표 데이터베이스에 존재하는 요구사항과 제약조건을 충족시켜야 한다. 예를 들어, '성별'(sex)이란 속성이 있다고 가정할 때, 'U'(unknown)라는 속성값이 소스 데이터베이스에서는 허용되지만, 목표 데이터베이스에서는 허용되지 않는다면, 'U'값이 가지는 모호성 문제를 해결하거나, 목표 데이터베이스가 이 값을 허용하도록 조정해야 한다.

셋째, 관계의 일관성(migration's coherence)

소스 데이터베이스의 데이터 간 관계(dependencies)는 보전되어 목표 데이터베이스로 이전되어야 한다. 예를 들어, 소스 데이터베이스에 존재하던 고객 주문과 고객 간의 관계는 목표 데이터베이스에서도 동일하게 적용되어야 한다.

위 세 가지 조건은 마이그레이션 된 데이터에 대한 기본적인 검증이라 할 수 있다. 완전성의 조건을 충족시키기 위해서 소스 데이터베이스에서 마이그레이션 대상으로 파악된 전체 데이터 건수와 목표 데이터베이스에 마이그레이션 된 데이터 건수를 비교하여 확인하는 방법을 적용할 수 있다. 데이터값의 일관성 조건을 충족시키기 위해서는 소스 데이터베이스의 주요 필드의 값(수치)이 목표 데이터베이스의 해당 필드로 정확하게 전송되었는지 합계 값을 비교하여 확인하는 방법이 사용될 수 있다. 마지막으로 관계의 일관성은 목표 데이터베이스의 구조 설계 시에 충분히 고려되어야 할 사항이다. 하지만, 이 세 가지 방법은 데이터값 자체에 대한 검증적인 측면에서는 다소 부족하다. 따라서 본 연구에서는 위 세 가지 검증 방식 외에 마이그레이션 된 데이터의 정확도를 검증하기 위한 추가적인 검증 방안을 적용하였다.

데이터값에 대한 검증을 위해서 마이그레이션 된 데이터에 대한 샘플링(Sampling) 및 육안검수(Manual inspection)를 수행하였다[9]. 샘플링 및 육안검수를 위해 ANSI/ASQ 표준을 참고할 수 있다. 이 표준은 동일 제품들이 반복되는

공정으로 제조되는 '로트(lots)' 또는 '배치(batches)' 개념에 기초하고 있다.

American Society for Quality에서 2003년 공포한 American National Standard - Sampling Procedures and Tables for Inspection by Attributes(ANSI/ASQ Z1.4-2003)은 계수형 샘플링 검사이다[10]. 이는 로트로부터 시료를 추출하여 검사하고 그 결과를 미리 정해둔 판정기준과 비교하여 로트의 합격 또는 불합격을 판정하는 절차이다. 계수형 샘플링 검사의 장점은 다음과 같다. 첫째, 검사대상이 많지 않으므로 많은 검사 인력이 필요치 않고 검사비용이 줄어든다. 둘째, 전수검사의 경우 검사대상의 과다로 인한 피로와 권태가 검사의 오류를 유발할 수 있으나 샘플링검사의 경우에는 그러한 오류가 줄어들 수 있다. 셋째, 검사에서의 불합격은 품질향상에 대한 동기부여가 된다.

단점은 첫째, 나쁜 품질의 로트를 합격시키고 좋은 품질의 로트를 불합격시킬 위험을 배제할 수 없다. 둘째, 효율적인 샘플링 검사를 계획하는 데 많은 시간과 노력이 든다. 이와 같은 단점이 있음에도 불구하고, 대용량 데이터를 적은 시간과 노력으로 검수해야 하는 경우에 효율적으로 사용될 수 있다. 또한 데이터가 흔히 제조업에서 말하는 제조품은 아니지만, 이것도 컴퓨터 시스템에 의해 만들어진 부산물이기 때문에 검증을 위해서 제조업에서 사용되는 샘플링 및 육안검수 방식을 적용하는 것이 타당하다[9].

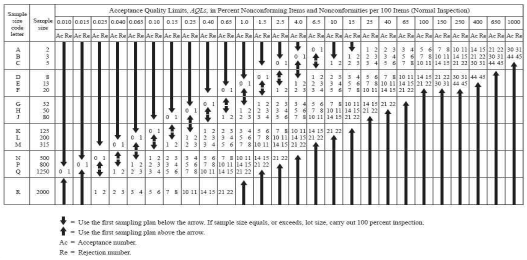
ANSI/ASQ Z1.4-2003에서 제시된 (단일) 샘플링 검사 절차는 다음과 같다. ① Lot 또는 Batch 크기 결정, ② 검사수준(Special inspection levels 또는 General inspection levels) 결정, ③ AQLs(Acceptance Quality Limits) 및 샘플 사이즈 결정, ④ 샘플링, ⑤ 검수, ⑥ 합격/불합격 판정

<표5>와 <표6>은 각각 샘플링 크기 결정 기준이 되는 코드표와 합격 판정 기준을 결정하는 마스터 테이블이다.

표 5. 샘플링 크기 결정을 위한 코드표
Table 5. Sample size code letters

Lot or batch size			Special inspection levels				General inspection levels		
			S-1	S-2	S-3	S-4	I	II	III
2	to	8	A	A	A	A	A	A	B
9	to	15	A	A	A	A	A	B	C
16	to	25	A	A	B	B	B	C	D
26	to	50	A	B	B	C	C	D	E
51	to	90	B	B	C	C	C	E	F
91	to	150	B	B	C	D	D	F	G
151	to	280	B	C	D	E	E	G	H
281	to	500	B	C	D	E	F	H	J
501	to	1200	C	C	E	F	G	J	K
1201	to	3200	C	D	E	G	H	K	L
3201	to	10000	C	D	F	G	J	L	M
10001	to	35000	C	D	F	H	K	M	N
35001	to	150000	D	E	G	J	L	N	P
150001	to	500000	D	E	G	J	M	P	Q
500001	and	over	D	E	H	K	N	Q	R

표 6. 일반 검수를 위한 단일 샘플링 검수 기준
Table 6. Single sampling plans for normal inspection(Master table)



이와 같은 데이터 마이그레이션 모델에 기초한 각 단계별 세부 수행 내용을 <표7>과 같이 정리하였다.

표 7. 데이터 마이그레이션 단계별 세부 내용
Table 7. Details by step for data migration

단계		세부 내용
환경분석	업무분석	· (논리) 업무 도메인 분석, · (물리) 업무별 현행 정보시스템 분석
	대상분석	· 현행 정보시스템의 데이터 형식(이미지, 첨부파일 등) 및 크기 분석
	자원분석	· 가용 가능 자원 분석(서버,스토리지 등)
데이터분석	모델분석	· 대상 데이터의 모델(구조) 분석
	소스분석	· 소스 데이터 분석(오류 유무 등) · 정제 필요성 판단(업무담당자와 협의)
설계	흐름설계	· 데이터 변환 및 적재 흐름도 설계 · 소스 데이터의 추출/변환/가공/적재 및 업무 비즈니스 흐름을 반영
	매핑정의	· 소스와 목표 데이터를 표준에 맞게 정의 (테이블, 컬럼, 코드파일 등) · 해당 데이터들의 매핑 관계 기술
구현		· 데이터 마이그레이션 도구 개발
검증		· 3가지 조건(완전성, 데이터값의 일관성, 관계의 일관성)의 충족 여부 검증 · 건수 검증, 합 검증, 샘플링 검증

Ⅳ. 데이터 마이그레이션 사례 연구

1. 디지털콘텐츠통합관리시스템

한국과학기술정보연구원(이하 KISTI)에서는 지금까지 국내외에 산재하는 핵심 과학기술정보를 연계한 NDSL, 국내 과학기술 학술정보 통합관리시스템인 OCEAN, 국내 460여 개 기관이 구독하는 저널 종합목록 기반 국가 학술지 분석(학술지 선정·평가 지원) 서비스인 WISECAT, 해외 전자저널 공동구매 컨소시엄이자, 관리시스템인 KESLI, 국내 과학기술정보의 인용색인 관리시스템이자, 서비스인 KSCI 등 정보

유통 관련 관리시스템들을 조직의 사정상, 그리고 <그림3>과 같이 관리 대상 데이터의 특성상 독자적으로 운영하여 왔다.

하지만, 국가 및 기관차원에서 데이터의 통합관리 필요성이 대두되고 있으며, 운영 및 관리 측면의 효율성을 극대화하고자 KISTI에서 운영·관리하고 있는 문헌데이터(학술논문, 학위논문, 특허, 연구보고서, 표준, 동향 등) 뿐만 아니라, 스마트 서비스(SMART service)를 위한 시맨틱 데이터를 고려한 체계적이며, 표준화된 디지털 콘텐츠 관리체제 구축이 요구되어졌다.

이에, 콘텐츠 관리자가 보다 편리하게·신속하게·체계적으로 문헌데이터 및 시맨틱 데이터를 관리할 수 있는 통합 관리시스템(DiCMS, Digital Contents Management System)을 구축 중에 있으며, 이를 위해 <그림4>와 같이 정보관리 표준(정보관리표준화)을 정의하고, 관리체계를 일원화하며, 분산된 관리 데이터베이스를 통합하였다[11].

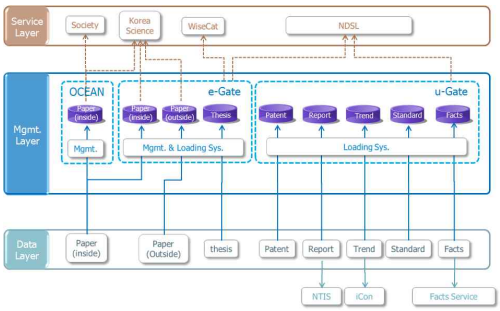


그림 3. 현행 분산형·이형 데이터베이스 관리 구조
Figure 3. Distributed and heterogeneous DB management(As-Is)

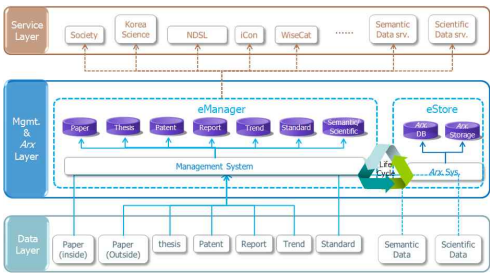


그림 4. 재구성된 통합형·동형 데이터베이스 관리 구조
Figure 4. Integrated and homogeneous DB management(To-Be)

또한 DiCMS 구축을 위한 주요 업무 중 하나는 기존 콘텐츠들을 관리하던 데이터베이스들을 연계하여 하나의 통합된 데이터베이스를 구축하는 것이다. 이를 위해 기존 데이터베이스들의 데이터를 통합된 데이터베이스로의 마이그레이션이 필요하였다.

2. 환경분석

DiCMS 데이터 마이그레이션을 위해 먼저 현행 정보시스템 및 데이터베이스의 구조를 분석하였다. KISTI에서 과학기술분야 정보유통의 대상은 <표8>과 같이 학술논문, 학위논문, 특허, 연구보고서, 동향분석, 표준, 사실정보 등 7개 콘텐츠로 나누어지며, 건수로는 약 8,900만건이다.

표 8. KISTI 과학기술 정보유통 콘텐츠 및 건수
Table 8. Science and technology contents and amount in KISTI

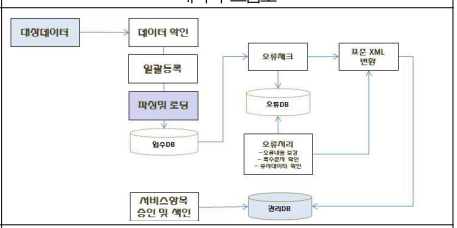
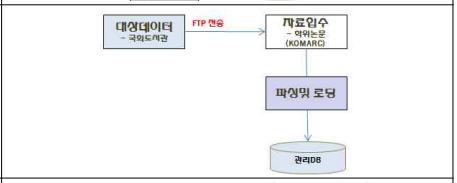
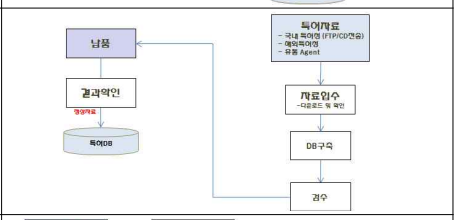
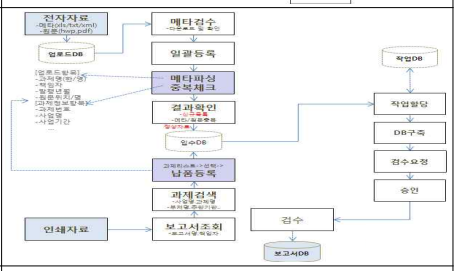
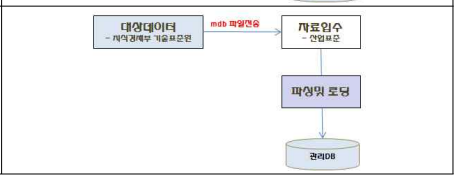
콘텐츠	구분	세부구분	건수
학술논문	학술논문	국내학술지	1,306,000
		해외학술지(영미권)	45,760,000
		해외학술지(중·일)	3,576,000
	학술회의	국내학술회의	311,000
		해외학술회의	8,459,000
	서지(종)	저널	84,000
		프로시딩	232,000
	소 계		59,728,000
학위논문	학위논문		1,338,000
특허	한국특허		3,598,000
	미국특허		7,114,000
	일본특허		8,993,000
	유럽특허		2,580,000
	국제특허		2,081,000
	소 계		24,366,000
연구보고서	국가연구개발보고서		132,000
	분석보고서		29,000
	소 계		161,000
동향분석	해외과학기술동향		149,000
	과학기술정책동향		18,000
	정보서비스 글로벌 동향		1,000
	소 계		168,000
표준	산업표준		60,000
사실정보	사실정보		3,807,000
합계			89,628,000

3. 데이터 모델 및 데이터 분석

데이터 마이그레이션 대상 콘텐츠인 6개 콘텐츠들의 데이터 흐름은 <표9>와 같다. 학술논문은 각 입수처로부터 데이터가 수집되면, 데이터 확인 후 일괄 등록하게 된다. 일괄 등록된 데이터는 파싱 과정을 거쳐 입수데이터베이스에 저장되고, 오류 체크 후, 변환 과정을 거쳐 서비스 전단계인 관리데이터베이스에 저장된다. 연구보고서는 입수처로부터 수집된 데이터를 메타파싱 후 데이터베이스 구축 과정을 거쳐서 관리데이터베이스에 저장된다. 특허, 학위논문, 동향, 표준은 비슷한

데이터 흐름을 보인다.

표 9. 과학기술정보 콘텐츠 별 데이터 흐름도
Table 9. Data flows by content

콘텐츠	데이터 흐름도
학술논문	
학위논문	
특허	
연구 보고서	
동향분석	
표준	

데이터 모델 분석 후 데이터 자체에 대한 분석을 수행하였다. KISTI에서 과학기술분야 정보유통의 대상 콘텐츠 중 데이터 마이그레이션의 대상은 사실정보를 제외한 6개 콘텐츠이며, 건수로는 약 8,600만건이다.

KISTI 과학기술분야 정보유통의 대상 콘텐츠를 물리적 수준에서 분석하면 <표10>과 같다. 6개 콘텐츠는 165개의 테이블에 저장되어 있고, 각 테이블의 전체 레코드의 총 건수는 약 5억2천만건 이상이다. 이는 데이터 마이그레이션 대상이 5억2천만건 이상임을 의미한다.

표 10. 콘텐츠 별 데이터 마이그레이션 건수
Table 10. Amount of data migration records by content

콘텐츠명	테이블 구성		
	구분	테이블수	건수
학술논문	공통/이용자/기관	18	3,559,000
	전자정보관리	37	69,517,000
	e-Gate	87	424,010,000
학위논문	학위논문	1	1,338,000
특허	한국/일본/미국/유럽/국제	6	24,244,000
연구 보고서	국가연구개발보고서	6	132,000
	분석보고서	6	29,000
동향분석	기술/정책/글로벌	3	169,000
표준	산업표준	1	60,000
총계		165	523,062,000

4. 데이터 마이그레이션 설계

데이터 마이그레이션 대상이 되는 소스 데이터와 목표 데이터 간의 매핑 관계를 데이터 컬럼 및 코드 수준까지 파악하여 매핑 정의서를 작성한다. 소스 데이터와 목표 데이터 간의 매핑 관계는 크게 네 가지로 정의될 수 있다. ① 대상 데이터베이스의 테이블 하나가 변환될 데이터베이스의 하나의 테이블로 매핑되는 경우(1:1 관계) 데이터 컬럼의 연관관계 및 데이터 변환관계를 명기, ② 대상 데이터베이스의 테이블 다수가 변환될 데이터베이스의 하나의 테이블로 매핑되는 경우(N:1) 데이터 컬럼의 연관관계 및 데이터 변환관계를 명기, ③ 대상 데이터베이스의 테이블 하나가 변환될 데이터베이스의 여러 테이블로 매핑되는 경우(1:N)에는 하나의 대상 테이블과 변환될 하나의 테이블의 다수의 묶음으로 간주하고 매핑 정의서를 명기, ④ 대상 데이터베이스의 테이블 하여러 개가 변환될 데이터베이스의 여러 테이블로 매핑되는 경우(N:M)에는 여러 개의 대상 테이블과 변환될 여러 개의 테이블의 다수의 묶음으로 간주하고 매핑정의서를 명기. 본 연구에서는 지면의 한계로 인해 각 테이블에 대한 개별 매핑정의서는 생략하고, 가장 복잡한 N:M 유형의 데이터 매핑 구조를 <그림 5>에서 예시적으로 표현하였다.

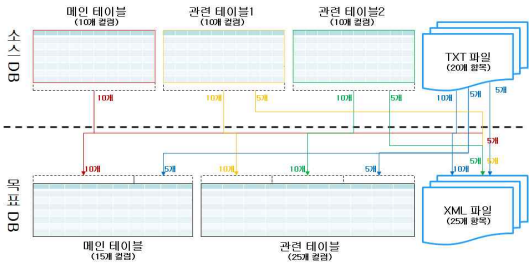


그림 5. N:M 유형의 데이터 매핑 구조 예시
Figure 5. N:M mapping example among various source data

작성된 매핑정의서를 기반으로 규칙(Rule)을 작성하되 데이터 마이그레이션 프로그램이 인지할 수 있도록 XML 기반의 규칙을 작성하였다. XML 기반의 데이터 변환 규칙 정의는 다음과 같은 장점이 있다. ① 복잡하고 어려운 컴퓨터 언어를 모르는 사용자라도 XML 언어의 구조만 파악하면, 손쉽게 데이터 변환 Rule을 작성 및 변경이 가능하다. ② 변경이 어려운 컴퓨터 언어가 아닌 XML 언어로서 변환 Rule을 관리함으로써 다양한 환경에 대한 접근성 및 이식성 확보가 가능한 장점도 있다. ③ 데이터 변환 및 이관/전송툴에 대한 가독성이 용이하고, PK에러, FK에러, Not Null 에러 등 다양한 에러 검출 시 Rule File 점검을 통해 에러 원인을 쉽게 파악하여 대처할 수 있다.

XML을 기반으로 한 Rule File은 다음과 같은 정보들을 포함한다. ① 물리적 데이터베이스 접근 정보를 확보/확인하고 규칙에 반영한다. ② 매핑정의서에 의해 소스 항목과 목표 항목 간 매핑 정보를 확인하고, 각 항목 간 변환 룰이 있는지 파악하여 반영한다. ③ 생성할 XML 파일/디렉토리에 대한 네이밍 룰에 대한 정의도 규칙에 반영한다. ④ 오류 및 모니터링 관련 항목에 대한 내용도 규칙에 반영한다. <그림6>은 실제로 데이터 마이그레이션 시 사용된 Rule File의 예시이다.

그림 6. 데이터 마이그레이션 룰 파일
Figure 6. Data migration rule file

5. 데이터 마이그레이션 구현 및 검증

데이터 마이그레이션을 위한 데이터 분석 및 마이그레이션 설계가 완료된 후, 데이터 마이그레이션을 수행하였다. 본 연구에서는 데이터 마이그레이션을 위해 소스 데이터베이스에서 목표 데이터베이스로 데이터를 변환 및 전송하는 프로그램을 개발하였다. 먼저 프로그램이 실행되면(그림7), 변환 규칙 정보를 저장하고 있는 Rule File의 동적클래스를 로딩(그림8)하여 소스 데이터베이스와 목표 데이터베이스 간 매핑 관계대로 데이터 이동 경로를 세팅(그림9)한다. 세팅 과정이 에러 없이 완료되면, 본격적으로 소스 데이터베이스의 데이터를 읽어서, Rule File의 규칙대로 변환하여 목표 데이터베이스로 데이터값을 저장하게 된다.

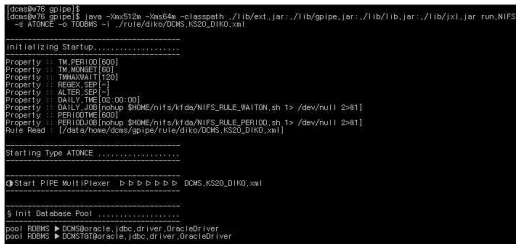


그림 7. 프로그램 구동
Figure 7. Program starting

그림 8. 동적 클래스 로딩
Figure 8. Loading dynamic class

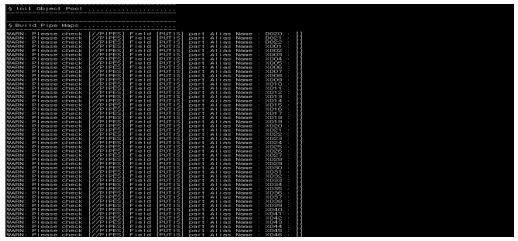


그림 9. 동적 클래스 로딩 후 빌드
Figure 9. Building after loading dynamic class

데이터 마이그레이션이 완료된 후, 검증 작업을 수행하였다. 데이터베이스의 완전성, 데이터값의 일관성, 관계의 일관

성을 검증하였고, 세 가지 요구조건에 대해서는 문제가 없는 것으로 나타났다.

샘플링검수에 의한 검증을 위해서 ANSI/ASQ Z1.4-2003에서 제시된 (단일) 샘플링 검사 절차를 따랐다. 동일 콘텐츠 내의 각 테이블 당 데이터 건수를 가중치로 하여 콘텐츠 별 샘플링 건수를 결정하였다. 콘텐츠별 샘플링 건수는 <표11>과 같다.

표 11. 샘플링 건수
Table 11. Sampling size by content

콘텐츠명	테이블 구성		샘플링건수
	구분	총건수	
학술논문	공통/이용자/기관	3,559,000	12,945
	전자정보관리	69,517,000	30,500
	e-Gate	424,010,000	75,800
학위논문	학위논문	1,338,000	1,250
특허	한국/일본/미국/유럽/국제	24,244,000	7,050
연구 보고서	국가연구개발 보고서	132,000	4,315
	분석보고서	29,000	1,125
동향분석	기술/정책/글로벌	169,000	1,315
표준	산업표준	60,000	500
총계		523,062,000	134,800

검수 데이터의 샘플링은 시스템에 의해 랜덤하게 이루어졌고, 육안 검수의 결과는 <표12>와 같다. 허용 오류 건수는 <표6>의 마스터 테이블에서 AQLs 1.0의 수준에서 계산된 것이며, 일부 오류가 있더라도 허용 오류 건수를 넘지 않으면, 데이터 마이그레이션 행위가 적절했다고 인정할 수 있다. 검수 결과, 콘텐츠별로는 모두 정상 판정이 나왔다. 테이블 수준에서는 일부 허용 오류를 초과하는 경우도 발견되었지만, 초기 데이터 마이그레이션 검증 기준이 콘텐츠 수준이었으므로, 데이터 마이그레이션은 적절히 수행된 것으로 받아들일 수 있다. 하지만, 데이터에 오류가 발견되었으므로, 해당 오류에 대해서는 데이터 마이그레이션 절차를 다시 확인하고, 오류를 바로 잡아주는 작업이 필요하다.

표 12. 오류 허용 수준과 검수 결과
Table 12. Acceptance number and result of inspection

콘텐츠명	구분	샘플링건수	허용오류 기준	오류건수	결과
학술논문	공통/이용자/기관	12,945	231	0	정상
	전자정보관리	30,500	550	21	정상
	e-Gate	75,800	1,330	355	정상
학위논문	학위논문	1,250	21	1	정상
특허	한국/일본/미국/유럽/국제	7,050	119	0	정상
연구 보고서	국가연구개발 보고서	4,315	77	0	정상
	분석보고서	1,125	28	2	정상
동향분석	기술/정책/글로벌	1,315	27	3	정상
표준	산업표준	500	10	0	정상
총계		134,800	2,393	382	정상

V. 결론

본 연구의 목적은 다양한 데이터 타입으로 존재하는 과학기술분야 콘텐츠들을 대상으로 데이터 마이그레이션을 위한 모델을 설계하고, 모델을 기반으로 실제 데이터 마이그레이션 작업을 수행한 결과를 제시하는 것이다. 이를 위해 2장에서는 국내 데이터베이스 산업분류 및 동향, 데이터 마이그레이션의 개념, 다양한 데이터 타입들 간 데이터 변환에 대해서 이론적으로 고찰을 하였고, 3장에서는 선행연구를 기반으로 효율적인 데이터 마이그레이션 모델을 설계하였다. 4장에서는 DiCMS 적용 사례를 통해서 제시된 데이터 마이그레이션 모델에 대한 검증을 수행하였다.

본 연구는 다음과 같은 시사점을 준다.

첫째, 데이터 마이그레이션 수행에 앞서 데이터 마이그레이션 목적 및 목표를 달성하기 위한 체계적이고, 효과적인 모델 설계가 필요하다. IT 실무에서 프로젝트 성으로 데이터 마이그레이션을 수행할 때, 시간 단축을 위해 절차 중심의 최소한의 업무만 제시되는 경우가 많다. 이는 데이터 마이그레이션의 중요한 과정들이 무시될 수가 있고, 결과적으로는 시간이 더 걸리는 경우도 발생할 수 있다. 따라서 이와 같은 실수를 줄이기 위해서 체계적인 모델 수립이 필요하다.

둘째, 데이터 마이그레이션의 검증을 위해서 기계적인 정합성 검수 뿐 아니라 샘플링에 의한 육안 검수도 필요하고,

ANSI/ASQ Z1.4-2003 방식도 데이터 정합성 검수의 한 사례가 될 수 있다. 데이터의 품질을 보증하기 위해 전수 육안 검사가 가장 좋겠지만, 현실적으로 불가능하므로, 샘플링 검사가 필요하다. 분야는 다르지만, 샘플링 검사라는 공통점이 존재하므로, 제조업의 품질검사의 한 방법을 참고할 수 있다.

셋째, 데이터 정합성 검수를 통해 본 연구에서 제시한 모델에 기초한 데이터 마이그레이션은 기계적 및 수작업 검수 기준을 통과하였고, 데이터 마이그레이션 모델로서 어느 정도 적절함을 알 수 있다.

본 연구는 DiCMS의 사례만을 조사한 것이어서, 다른 데이터 마이그레이션 사례에 대한 비교가 어렵다는 한계점을 가지고 있다. 현재 국내에서 금융권 중심으로 데이터 마이그레이션이 많이 일어나고 있지만, 구체적인 방법론에 대해서 공개하는 경우가 드물다. 또한 다양한 데이터 타입의 형태로 존재하는 과학기술정보 콘텐츠의 특성이 있으므로, 금융권 데이터와 비교하는 것이 무리일 수도 있다. 향후에는 본 연구에서 제시된 모델을 다른 사례에도 적용하여, 보편적으로 적용 가능한 데이터 마이그레이션 모델이 될 수 있도록 추가적인 연구가 필요할 것으로 판단된다.

참고문헌

[1] http://www.apple.com/kr/pr/library/2011/01/22 Apples-App-Store-Downloads-Top-10-Billion.html

[2] Environment changing of database industry, DBGuide, KDB, Sep. 2010

[3] S.H. Park, "Development of Guidelines for Migration of Electronic Records in Records Management System", MS paper, Myongji University, Aug. 2010

[4] D.Y. Kwon et al, "A Study on Migration Strategy for Long-term Preservation of Electronic Records", Journal of the Korean Society of Archives and Records Management, Vol. 9, No. 2, pp. 19-40, Dec. 2009

[5] W.T. Woo, "A Case Study on the Web Publishing of XML Document and the Transformation of RDB File into XML Document", Journal of Decision Science, Vol. 13, pp. 75-98, 2005

[6] Larry P. English, "Improving Data Warehouse and Business Information Quality", Wiley Computer Publishing, 1992

[7] S.H. Lee, "An Efficient Data Migration Method for Financial Business", MS paper, Soongsil University, Jun. 2010

[8] Validation of Data Migration, REVER-S21, REVER, Aug. 2008

[9] Data Migration Validation, Drug Discovery & Development magazine, Vol. 10, No. 2, pp. 28-31, Feb. 2007

[10] American National Standard - Sampling Procedures and Tables for Inspection by Attributes, American Society for Quality, 2003

[11] W.G. Lee et al, "A Research on Digital Content Management for SMART Service", Proceedings of Korea Computer Congress 2011, Vol. 38, No. 1(B), pp. 307-310, Jun. 2011

저 자 소 개

신 성 호

2000 : 경북대학교 경영학과 경영학사.

2002 : 경북대학교 경영학과(MIS전공) 경영학석사.

2002년~현재 : 한국과학기술정보연구원 선임연구원.

관심분야 : 데이터통합, 데이터품질, IS 평가.

Email : maximus74@kisti.re.kr

이 민 호

2000년 : 충남대학교 대학원 컴퓨터공학과 공학석사.

2006년 : 충남대학교 대학원 컴퓨터공학과공학박사수료.

2001년~현재 : 한국과학기술정보연구원 선임연구원.

관심분야 : 정보검색 및 추출, 정보보호, 분산시스템.

Email : cokeman@kisti.re.kr

이 원 구

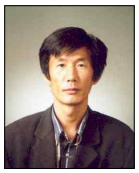
2000년 : 한남대학교 대학원 컴퓨터 공학과 공학석사

2005년 : 한남대학교 대학원 컴퓨터 공학과 공학박사

2005년~현재 : 한국과학기술정보연구원 선임연구원

관심분야 : 데이터베이스, 지식관리, 과학데이터

Email : wglee@kisti.re.kr



윤 화 묵

1997년 : 공주대학교 대학원 전자계산학과 공학석사

2008년 : 배재대학교 컴퓨터공학과 공학박사

현재 : 한국과학기술정보연구원 책임연구원

관심분야 : 데이터베이스, 정보검색, 온톨로지

Email : hmyoon@kisti.re.kr

성 원 경

1989년 : 연세대학교 대학원 언어학과 인문학석사

1996년 : 프랑스 파리7대학교 전산언어학 공학박사

2004년~현재 : 한국과학기술정보연구원 책임연구원

관심분야 : 데이터베이스, 지식관리, 과학데이터

Email : wksung@kisti.re.kr

김 광 영

2001년 : 부산대학교 전자계산학과 공학석사

2011년 : 충남대학교 문헌정보학과 인문학박사

2001년~현재 : 한국과학기술정보연구원 선임연구원

관심분야 : 정보검색, 개인화 검색, 아카이빙

Email : kykim@kisti.re.kr