

데이터 마이그레이션 한글 입력 개선 방안 연구

서강현*, 김치용**, 장종욱***

*동의대학교 소프트웨어융합학과

**동의대학교 게임애니메이션공학전공

***동의대학교 컴퓨터공학과

e-mail : rkdgusdl789@naver.com, kimchee@deu.ac.kr, jwjang@deu.ac.kr

A Study on Improving Hangul Input of Data Migration

Kang-hyeon Seo*, Chee-Yong Kim**, Jong-Wook Jang***

*Dept. of Software convergence, Dong-eui University

**Major of Game Animation Engineering, Dong-eui University

***Dept. of Computer Engineering, Dong-Eui University

요 약

데이터 마이그레이션은 기존 시스템에 존재하는 데이터, 즉 이행 대상 데이터를 추출하여 새로운 시스템 내의 지정된 형식으로 옮기는 과정을 보장한다. 마이그레이션의 효율성과 정상적인 작동을 보장하기 위해서는 데이터 형식에 상관없이 데이터의 이동이 가능해야 한다. 그러나 본 연구에서 사용된 DataLoader라는 데이터 마이그레이션 프로그램을 사용 시 데이터 형식이 한글인 경우 다른 시스템으로의 데이터 이동에 있어 오류 사항을 발견하였다.

따라서 본 연구를 통해 C#으로 제작된 DataLoader 프로그램 내부의 한글 데이터를 처리하는 소스 라인을 찾아 한글 데이터 처리 방식에 알맞게 수정하여 위와 같은 문제를 해결하고 데이터 이동의 정확성을 높이는 방법을 제시한다.

1. 서론

데이터베이스 산업의 지속적인 성장과 맞물려, 데이터 마이그레이션의 중요성이 높아짐에 데이터 마이그레이션이 수행하는 시간과 정확성 또한 중요해지고 있다[1].

이러한 수행 시간과 정확성에는 시스템 간의 교환되는 데이터의 종류 및 수가 많은 영향을 미친다는 것을 알 수 있다[2]. 하지만 본 연구에서 사용되는 Data Loader 프로그램을 사용하여 데이터를 교환할 시 한글 데이터의 경우 교환이 불가능한 것을 확인하였다.

따라서 본 연구에서는 한글 데이터의 입력 개선을 목표로 연구를 진행하였다.

2. 관련연구

(1) 데이터 마이그레이션의 개념

데이터 마이그레이션에 대한 번역어로는 다양한 용어가 사용되고 있다. 정보시스템 구축 분야에서는 ‘(데이터) 이행’이란 표현을 주로 사용해 왔다. 문헌 정보 분야에서는 자료관리 시스템과 관련하여 자료 변환이나 자료 전환, 변환이란 용어가 주로 쓰이고 있으며, 이 또한 통일되지 못하게 사용되고 있다. 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 크게 두 가지로 생각해 볼 수 있다. 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, 노후화될 수 있는 상황에서, 기록물로의 지속적인 접근을 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[1].

(2) 국내 데이터산업 부문별 시장 규모

2017년 데이터산업 시장 규모는 14조 3,047억 원으로 2016년 대비 4.0% 성장한 것으로 나타났다. 2010년 이후 연평균증가율은 7.5%로 매년 꾸준한 성장세를 유지하는 것으로 조사됐다.

부문별로는 데이터 솔루션 시장이 1조 6,536억 원, 데이터 구축·컨설팅 시장이 5조 8,565억 원, 데이터 서비스 시장이 6조 7,946억 원으로 나타났다. 부문별로 보면 데이터 서비스 시장이 47.5%로 가장 큰 비중을 차지했고, 이어서

데이터 구축·컨설팅 시장이 40.9%, 데이터 솔루션 시장이 11.6% 순으로 그 뒤를 이었다[3].

구분	2010년	2011년	2012년	2013년	2014년	2015년	2016년	2017년(단)	증감률 '16-'17	CAGR '10-'17
데이터 솔루션	6,725	8,717	10,487	10,789	13,619	14,124	15,720	16,536	5.2	13.7
데이터 구축 컨설팅	37,407	43,180	47,715	49,985	53,730	55,280	55,850	58,565	4.9	6.6
데이터 서비스	42,242	43,218	47,317	52,258	57,329	64,151	65,977	67,946	3.0	7.0
전체	86,374	95,115	105,519	113,032	124,678	133,555	137,547	143,047	4.0	7.5

그림 1. 데이터산업 부분별 시장 규모

(3) 해외 데이터산업 현황

451 Research가 분석한 세계 데이터 시장은 8개의 세부 솔루션 시장으로 이뤄져 있다. 8개 솔루션의 전체 시장은 2018년 999억 달러 규모로 형성됐다. 2017년부터 2022년까지의 연평균성장률(CAGR)은 10.3%를 기록했으며, 2022년에는 1,461억 달러로 성장할 것으로 예상된다.

8개 세부 시장 중 가장 큰 시장은 운영 데이터베이스(Operational databases) 시장으로, 2018년 기준 약 400억 달러 이상으로 형성됐다. 리포팅과 분석(Reporting and analytics) 시장은 약 200억 달러, 데이터 분석 플랫폼(Analytic data platforms) 시장은 150억 달러 내외로 형성돼 각각 2, 3위를 차지했다[3].



그림 2. 2017-2022년 데이터 기반 전체 시장 규모

(4) Data Loader 프로그램

Data Loader 프로그램은 본 연구에서 사용되는 마이그레이션 프로그램으로 한 시스템과 연결을 한 후 시스템에 있는 항목들에 관하여 원하는 데이터들을 입력, 추가할 수 있으며, Data Loader 프로그램 자체는 C#을 통하여 제작이 되었다.

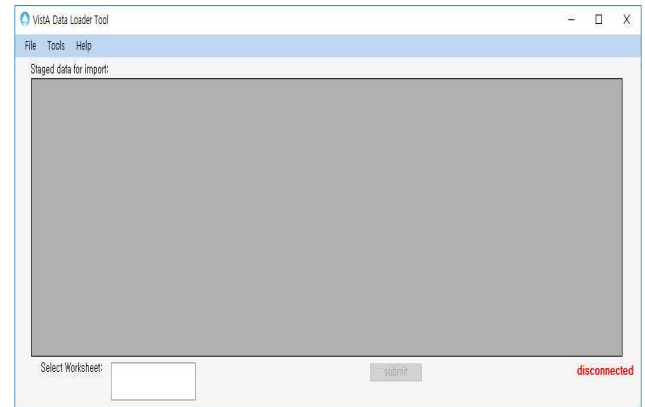


그림 3. Data Loader 프로그램 UI

추가하고 싶은 데이터들은 엑셀 파일에 입력을 한 후 Data Loader 프로그램에 해당 엑셀 파일을 업로드 한다. 그 후 'Submit'을 선택하면 엑셀에 입력된 데이터들이 지정된 시스템으로 입력되는 형식으로 진행된다.

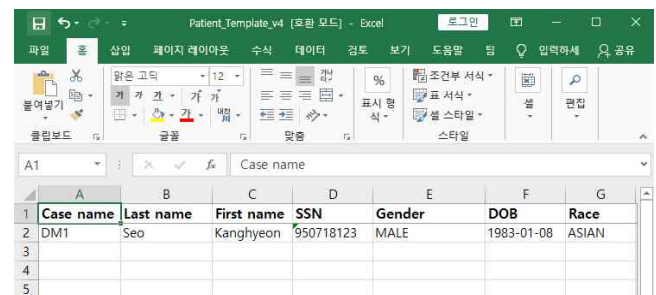


그림 4. 데이터 입력 엑셀 파일

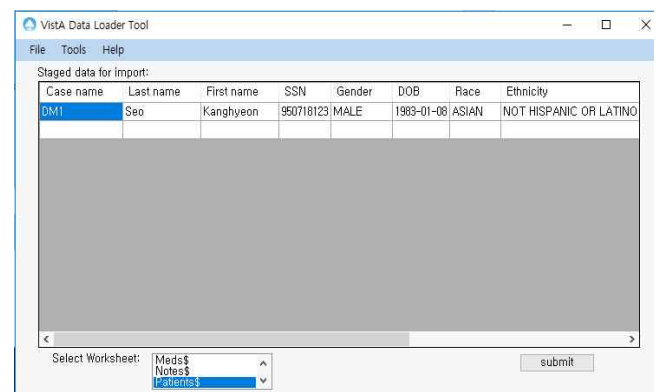


그림 5. Data Loader 엑셀 파일 업로드

3. Data Loader 한글 입력 개선 방안

현재 Data Loader 프로그램은 데이터를 입력하고 추가하는 과정에서 ASCII로만 동작이 가능하게 설정이 되어 있는 상태이다.

ASCII란 미국정보교환표준부호(American Standard Code for Information Interchange)로 영문 알파벳을 사용

하는 대표적인 문자 인코딩이다. 하지만 한글인 경우 ASCII로 인코딩할 경우 한글이 깨지는 현상을 확인할 수 있다. 한글 데이터를 손실 없이 사용하기 위해서는 UTF-8이 지원되는 환경에서 데이터를 입력할 필요가 있다.

UTF-8이란 유니코드를 위한 가변 길이 문자 인코딩 방식 중 하나로 한 문자를 나타내기 위해 1바이트에서 4바이트까지를 사용한다.

따라서 이에 대한 개선방안으로 C#으로 제작이 된 Data Loader 프로그램 내에 데이터를 처리하는 부분을 찾은 후 Unicode로 통신할 수 있도록 수정을 하는 것이다.

4. 구현

구현은 예상외로 생각보다 간단했다. Data Loader 프로그램 내부에 추가하는 데이터를 ASCII로 처리하는 구현은 다음과 같이 정의되어 있다.

```
IntPtr namePtr = (IntPtr) Marshal.StringToHGlobalAnsi(name);
IntPtr valuePtr = (IntPtr) Marshal.StringToHGlobalUni(value);
```

그림 6. Data Loader 한글 처리 소스 변경 전

그림 5와 같이 StringToHGlobalAnsi(name)으로 선언하면서 들어오는 데이터를 ASCII로 처리하는 것이다. 하지만 이런 식으로 데이터를 처리할 시 한글 데이터의 경우 바이트 수의 길이가 영어와는 차이가 있어, 한글이 깨지는 현상이 발생한다.

따라서 해당 소스 라인을 아래와 같이 변경하여 유니코드 형식으로 처리를 하면 한글 데이터라고 해도 1에서 4 바이트까지 처리가 되므로 한글이 깨지는 현상 없이 원활하게 데이터가 입력될 수 있다.

```
IntPtr namePtr = (IntPtr) Marshal.StringToHGlobalUni(name);
IntPtr valuePtr = (IntPtr) Marshal.StringToHGlobalUni(value);
```

그림 7. Data Loader 한글 처리 소스 변경 후

해당 라인을 수정 후 테스트를 위해 ‘서강현’이라는 데이터를 시스템 추가를 해보았다. 추가를 위해 우선 데이터를 입력하는 엑셀 파일에 ‘서강현’이라는 데이터를 입력 후 Data Loader 프로그램에 아래 <그림 8>과 같이 해당 엑셀 파일을 불러오는 작업을 수행한다.

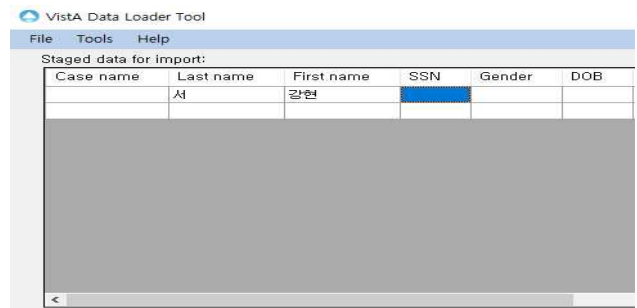


그림 8. Data Loader 한글 데이터 추가

엑셀 파일을 불러온 후 시스템으로 입력하기 위해 ‘submit’ 버튼을 선택하게 되면 아래 <그림 9>와 같이 데이터 입력에 성공했다는 창이 출력된다.

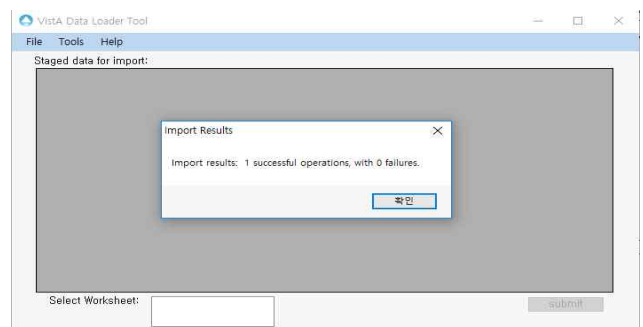


그림 9. Data Loader 데이터 추가 결과

Data Loader 프로그램에 지정된 시스템 내부에서 확인하면 아래 <그림 10>과 같이 ‘서강현’ 데이터가 추가되어 있는 것을 확인할 수 있다.



그림 10. 시스템 내에 추가된 한글 데이터

각 각 그림에 대한 자세한 설명과 진행 절차 등이 소개되어야 합니다. 그리고 본 논문에서 어떻게 이 그림이 필요한지를 알 수 있어야 합니다.

5. 결론

데이터 마이그레이션은 여러 가지 방법들 중에서 프로젝트 성격과 환경에 맞게 수행한다. 데이터의 변환 없이 직관적으로 바로 익스포트, 임포트 할 수도 있고, 능숙한 프로그램을 활용하여 배치프로그램을 만들어 데이터 마이그레이션을 수행 할 수 있다[4].

하지만 본 연구에서 다뤘던 Data Loader 프로그램의 경우 한글 데이터는 마이그레이션하는 과정에서 문제가 있었기 때문에 그 문제를 해결하는 것을 목표로 하였다. 따라서 2장에서는 데이터 마이그레이션의 정의 Data Loader

프로그램을 소개했으며 3장에서는 해결방안, 4장에서는 해결방안을 통한 구현을 해보았다.

본 연구를 통해 Data Loader의 한글 데이터 입력에 관한 문제는 해결이 되었으나, 기존 시스템에 존재하는 영어 데이터를 한글 데이터로 변경할 시에는 변경이 안 되는 문제를 발견하였다. 따라서 기존 시스템의 데이터 변경에 관한 문제를 해결하여 Data Loader의 성능을 향상시키는 것을 목표로 연구를 진행할 예정이다.

참고문헌

- [1] 신성호, 이민호, 이원구, 윤화목, 성원경, 김광영, “과학기술정보콘텐츠 통합관리시스템 구축을 위한 데이터 마이그레이션 모델 수립 및 적용 사례” 한국컴퓨터정보학회 논문지, 제16권 제11호 1호, 2011.11
- [2] 김성덕, “마이그레이션 사업 성공 요인에 관한 연구 : 최신 IT 플랫폼을 중심으로” 숭실대학교 학위논문(박사), 2015.8
- [3] “2018 데이터산업 백서”
- [4] 최종근, “데이터 마이그레이션을 위한 오픈소스 ETL 도구 평가”, 숭실대학교 학위논문(석사), 2011.6