

데이터 품질 평가에 관한 연구

A Study of the Data Quality Evaluation

정 혜 정*
Hye-Jung Jung

요 약

본 연구는 데이터 품질 평가란 관점에서 현재 진행되어지고 있는 국제 표준 ISO/IEC 25000 시리즈인 SQuaRE(Software product Quality Requirements and Evaluation) 프로젝트 중에서도 ISO/IEC 25012의 데이터 품질 평가 모델에 대한 연구이다. 데이터의 양이 많아지면서 사용자 관점에서는 정확한 데이터, 최신의 데이터, 사용하는 도구에 적합한 데이터, 보안성과 비공개성을 준수할 수 있는 데이터를 원한다. 저 품질의 데이터가 비즈니스에 미치는 영향을 평가하여 실제적으로 적용할 수 있는 관점에서 데이터 품질 관리에 대한 것을 연구한다. 이러한 관점에서 데이터에 대한 품질을 평가할 수 있는 평가항목을 제시하고 제시된 평가항목의 평가 방안에 대해서 제시한다. 본 연구는 현재 진행되어지고 있는 ISO/IEC 25012의 표준문서와 소프트웨어 품질관리 표준문서 ISO/IEC 9126-2를 기초로 하여 연구되었으며 현재 진행되어지고 있는 소프트웨어 품질 평가 모델 ISO/IEC 25000을 기반으로 하여 데이터 품질을 정량적으로 평가하는 방안을 제시한다.

Abstract

In this paper, We study on the Data Quality Model of ISO/IEC 25012 among the Software product Quality Requirements and Evaluation(SQuaRE) in ISO/IEC 25000 Series. Because of the increasing data, user require the accuracy data, recent data, suitable data for used tools, complied security and not open to be public. We research the data quality management in the point of application of be affect influenced low quality in business. We propose the testing items and we propose the method of the evaluation proposed testing items. We study on the basis of International Standards ISO/IEC 25012 and ISO/IEC 9126-2 and we proposed the testing method quantitatively on the basis of ISO/IEC 25000.

* Keyword : 데이터 품질 평가(Data Quality Testing), 품질 평가(Quality Evaluation), ISO/IEC 9126, ISO/IEC 25012, Testing Items

1. 서론

시장 경쟁이 활발해 지면서 기존에 축적된 데이터를 유용한 정보로 변환하여 기업의 의사 결정에 도움을 줄 수 있는 정보로 활용하는 것은 상당히 중요하다. 최근 들어 금융업계에서 도입한 리스크 관리 시스템(Risk Management System)도 복잡한 시장 경쟁 속에서 데이터 관리를 중심으로 기업의 최대 목표를 달성하기 위한 노력이다. 부정확한 데이터로 인하여 생길 수 있는 손실은 첫째, 데이터 품질 저하로 발생할 수 있는 비용적인 측면을 고려할 수 있으며 둘째, 고객 불만과

제품 관련 불만 사항 등을 들 수 있으며 셋째, 낮은 품질의 데이터로 인한 비즈니스 비용으로 회수 불가능한 비용이나 제품과 서비스에 대한 추가작업 비용 등을 들 수 있다. 이와 같이 데이터 품질에 문제가 발생할 수 있는 근본적인 원인은 첫째, 데이터의 속성 및 구조에 대한 표준화 연구가 미흡하고 둘째, 데이터에 대한 분산된 관리를 통합할 수 없으며 셋째, 데이터 구조 간 중복이나 불일치가 발생할 수 있으며 넷째, 데이터 구조의 표현에 있어서 용어 등의 표준화가 되어 있지 않다는 것이다. 물론 데이터에 대한 품질 향상을 위한 첫 번째 과제로는 데이터의 품질을 평가할 수 있는 표준화 연구라고 할 수 있으며 데이터의 구조와 데이터 품질, 데이터 성능, 데이터 보안, 데이터 관리 등에 대한 표준화 연구가 필요하다.

* 정 회원 : 평택대학 디지털응용정보학과 교수

jhjung@ptu.ac.kr

[2007/05/03 투고 - 2007/05/09 심사 - 2007/06/21 심사완료]

이터 연계라는 관점에서 표준화에 대한 연구가 진행되어야 할 것으로 보여진다. 데이터의 품질 관리라는 차원에서 데이터 품질 관리 항목을 선정하고 선정된 항목을 중심으로 데이터에 대한 품질을 측정하여야 만 측정된 결과에 따른 데이터 품질 개선이 이루어지고 개선된 사항을 통한 품질 개선 결과 평가가 가능할 것이다. 이러한 관점에서 가장 기초적으로 데이터 품질 관리를 위해서 평가 되어야 할 평가항목을 제시하고 평가방법을 제안한다. 본 연구의 2장에서는 소프트웨어 품질 평가와 관련한 데이터 품질 평가에 대한 사전연구를 소개하고 3장에서는 기존의 연구와 비교 분석을 실시한다. 4장에서는 데이터 품질을 평가하기 위한 평가 항목과 평가 방안을 제시하고 5장에서는 본 연구의 결론과 데이터 평가 방안을 통한 앞으로의 연구 과제에 대해서 제시한다.

2. 데이터 품질 평가 동향

최근 들어서 데이터에 대한 품질에 많은 관심을 가지고 있는 이유는 기관이나 기업의 각 업무가 정보화 되면서 기관별 기업별로 업무를 처리하는데 있어서 데이터 간에 심각한 중복성과 불일치성의 문제가 발생하고 있으며, 기관과 기업 간의 데이터 표준 및 데이터의 일관성 부족으로 인하여 데이터에 대한 호환이 이루어지지 않고 있다는 것이다. 또한 기관과 민간기업 고객 관계 관리(CRM: Customer Relationship Management) 시스템을 구축하면서 데이터를 축적하고 있으나 이러한 데이터에 대한 일관성의 부족으로 인하여 활용하는데 많은 어려움을 겪고 있다는 것이다. 사용자가 요구하는 데이터 품질에 대한 요구는 업무와 어느 정도 관계가 있는가?, 데이터는 얼마나 정확한가?, 데이터의 불일치가 어느 정도인가?, 데이터는 얼마나 업데이트가 잘 되어 있는가?, 데이터는 사용하는 도구에 얼마나 적합한가?, 필요한 데이터를 얼마나 잘 제어할 수 있는

가?, 데이터의 보안이 얼마나 잘 이루어지고 있는가? 등은 데이터의 품질 평가를 위한 평가항목에서 고려하여야 할 사항으로 보여진다. 소프트웨어 품질 관리라는 차원에서 제품에 대한 평가항목은 국제 표준 ISO/IEC 9126을 기반으로 하고 있으나 정량적이고 객관적으로 평가하기에 다소의 어려움이 있어 ISO/IEC 25000이라는 이름으로 새롭게 평가 방안을 연구하고 있다. 데이터 품질에 대한 연구는 현재 ISO/IEC 25012[7]라는 이름으로 전체 프로젝트의 한 파트로 연구되어지고 있으며 데이터 품질에 대한 국제 표준 문서의 연구 상황은 아래와 같다.

〈표 1〉 ISO/IEC 25012 연구 진행 상황*

NP	WD	CD	CD.2	CD.3	FCD	IS
		2006.5	2006.10	2007.4	2007.10	

표에서 제시한 것처럼 데이터 품질은 현재 FCD 단계의 진행 상황에 있으며 데이터 품질 평가란 관점에서 연구가 진행되어지고 있다. 일반적으로 기존의 연구[4-11]에서 데이터 품질에 대한 연구는 데이터를 처리하여 모델을 설정하는데 있어서 필요한 정보가 모두 존재하는지를 평가하는 완전성과 데이터의 의미가 누구에게나 동일하게 전달되는지를 평가하는 명확성과 데이터를 정확하게 기록하고 있는지를 평가하는 정확성과 중복 데이터 간에 동일한 값을 갖고 있는지에 대한 일관성과 필요한 형태로 데이터가 존재하는지를 평가하는 유용성에 대한 평가를 실시하였다. 그리고 소프트웨어의 품질 평가란 관점에서는 국제 표준 ISO/IEC 9126-2[1,2,3,4, 18, 19, 20]를 기반으로 하여 이루어졌으며 6가지 품질특성을 기초로 하여 평가하였다. 소프트웨어 기능을 평가하는 기능성과 고장을 관리하는 신뢰성과 소프트웨어 사용자에 대한 사용성 평가와 소프트웨어 유지관리를 위한 유지보수성과 소프트웨어 설치 등을 평가하는 이식성과 시간과 자원의 효율적인 측면을 평가하는 효율성으로 나누어 평가 메트릭을 제시하

고 있다. 데이터의 품질 평가와 소프트웨어 품질 평가란 두 가지 차원을 모두 고려하여 현재 ISO/IEC 25012[7]의 국제 표준 프로젝트는 16개 품질 평가 항목을 통해서 데이터에 대한 품질 평가를 제시하고 있다. 현재 국제 표준에서 제시하고 있는 데이터 품질에 대한 평가는 첫째 일관성(Consistency)에 대한 평가로써 사용자가 사용하는 데이터가 전·후 관계 다른 데이터에서도 통일성 있게 사용되어지는지를 평가하는 것이다. 둘째, 현재성(Currentness)으로 데이터가 업데이트 되었을 경우 즉시 업무에 반영 되어지는지를 평가하는 것이며 셋째, 완전성(Completeness)으로 사용자가 데이터를 사용하는데 있어서 어느 정도 완벽한지를 평가하는 것이며 넷째, 정확성(Precision)으로써 얼마나 정확하게 데이터가 기록 수집되었는지를 평가하는 것이다. 다섯째, 정밀성(Accuracy)으로써 데이터가 얼마나 정확히 데이터의 속성을 표현하고 있는지를 평가하는 것이며 여섯째, 보안성(Security)은 데이터가 권위가 부여된 사용자만이 접근되어지고 해석되어지는지를 평가하는 것이다. 일곱째, 유효성(Availability)으로써 데이터가 회복되어지는 것을 평가하는 것이고 여덟째, 복구성(Recoverability)으로 성공적으로 복구되어진 파일을 측정하여 평가하고 있으며 아홉째, 이해성(Understandability)으로써 사용자가 적당한 언어, 심볼, 단위를 이해하고 표현할 수 있는지를 평가하는 항목이다. 열번째, 효율성(Efficiency)으로써 저장 공간의 효율성을 평가하는 것이고, 열한번째, 가변성(Changeability)으로써 데이터가 형태, 길이, 또는 할당 값을 변화할 수 있는 능력을 평가하는 것이고, 열두번째, 이식성(Portability)은 데이터가 하나의 플랫폼에서 다른 플랫폼으로 이식될 수 있는 능력을 평가하는 것이며, 열세번째, 능력성(Traceability)은 데이터가 원본과 변화된 것을 자동적으로 회계 감사할 수 있도록 하는 능력을 평가한다. 열네번째, 신뢰성(Credibility)은 데이터가 사실로써 고려되어지고 사용자에 의해서 믿어지는 정도를 평가하는 것이며, 열다섯번째 접근성

(Accessibility)은 데이터가 접근되고 활용되어질 수 있는 정도를 평가하는 것이며, 열여섯번째, 일반적 준수성(Regular Compliance)은 일반적인 규격 등에 접근한지를 평가하는 것이다. 위의 열여섯 가지는 현재 진행되어지고 있는 국제 표준에서 제시하고 있는 데이터 품질에 대한 평가 항목이다. 이와 같은 국제 표준 제정에서 연구되어지고 있는 데이터 품질에 대한 평가항목을 중심으로 3장에서는 현재 한국데이터베이스 진흥 센터에서 제시하고 있는 데이터 품질과의 실질적인 차이에 대해서 연구했다.

3. 데이터 품질 평가 연구 방향 비교

기관과 기업에서 운영하는 정보시스템의 범위가 확대되고 복잡도가 증가함에 따라 정확성, 일관성, 적시성 등과 같은 데이터 품질이 중요한 이슈가 되기 시작하면서 데이터 품질 관리는 상당히 중요한 연구과제가 되었다. 한국 데이터베이스 진흥 센터에서는 데이터의 품질 관리라는 측면에서 데이터 품질에 대한 데이터 관리를 네 가지 차원에서 제시하였다. 첫째, 표준 데이터로 정보 시스템에서 사용하는 용어, 도메인, 코드 및 기타 데이터 관련 요소에 대해 공통된 형식과 내용으로 정의하여 사용하는 표준 관련 데이터를 평가하기 위한 것으로 표준데이터의 관리를 통해서 불일치되는 문제점을 해결할 수 있도록 평가 모델을 제시하였다. 둘째, 데이터 모델에 대한 메타 데이터를 관리함으로써 데이터 구조에 대한 최신 정보를 유지하고 전사 차원의 데이터 모델 공유 및 재사용을 극대화하며, 체계적인 데이터 모델의 변경 관리를 가능하게 하는 모델데이터에 대한 평가방안을 제시하였다. 셋째, 관리 데이터는 데이터베이스를 효과적으로 운영·관리하는데 필요한 데이터를 의미하는 것으로 사용 관리 데이터, 장애 및 보안 관리 데이터, 성능 관리 데이터, 흐름 관리 데이터, 품질 관리 데이터 등을 포함하여 평가 방안을 제시하였다. 넷째, 업무 데이터는 기

관이나 기업의 업무 및 비즈니스를 수행하는데 필요한 데이터로 정의하고, 일반적으로 데이터 흐름에 따라 소스 데이터, 운영 데이터, 분석 데이터로 구분하여 평가하는 방안에 대해서 제시하였다. 데이터 품질 관리 지침서에 표준데이터는 데이터에 관련된 공통된 형식과 내용으로 정의하고 있으며 사용하는 표준 관련 데이터를 표준데이터로 정의하고 있다. 데이터 품질 관리를 위해서 제시한 지침서에 첫째, 데이터 관리는 단어에 대한 일관성 평가를 위해서 표준 단어 사전을 이용하도록 제시하고 있으며 평가항목은 표준성, 참조 가능성, 일반성, 대표성을 평가하도록 하고 있다. 둘째, 표준 도메인 사전을 이용하도록 제시되어져 있으며 데이터 중 논리적, 물리적 유사한 유형의 데이터를 그룹화 하여 해당 그룹에 속하는 데이터의 유형과 길이를 정의하도록 지시하고 있으며 평가항목은 표준성, 유일성, 업무지향성을 제시하고 있다. 셋째, 표준 용어 사전에 대한 평가로 표준 단어를 조합하여 정의한 용어를 말하는 것으로 표준성, 일반성, 업무지향성을 평가하도록 제시되어져 있다. 넷째, 표준코드를 평가하는 것으로 재사용성, 일관성, 유일성, 정보분석성을 평가하도록 제시되어져 있으며 다섯째, 데이터 표준 요소를 평가하는 것으로 설계 및 구축에 필요한 요소를 추출하여 표준이 필요한 요소를 정의하고 그 요소에 대한 업무적 표준을 정의함을 평가하도록 제시되어져 있다. 이와 같이 현재 데이터 품질 관리 지침을 위한 연구 내용은 대체적으로 데이터 자체의 품질에 대한 품질 평가 방안을 데이터 모델에 적용하여 평가할 수 있도록 제시되어져 있으나 현재 진행되어지고 있는 국제 표준 ISO/IEC 25012의 데이터 품질에 대한 연구는 소프트웨어란 관점에서 다루어지는 데이터에 대한 품질 평가를 위한 방안을 제시하고 있다는 차원에서 연구 방향이 다르다고 할 수 있다. 본 연구의 4장에서는 ISO/IEC 25012에서 연구하고 있는 소프트웨어에서 다루고 있는 데이터에 대한 품질 평가를 위한 방안을 제시하려한다.

4. 데이터 품질 평가 항목

데이터에 대한 품질 평가를 위해서 6가지 품질 특성을 중심으로 각각의 평가항목을 제시한다. 일반적으로 데이터에 대한 평가를 위해서 사용자 관점에서 가장 많은 반응을 보인 것은 첫째, 데이터는 정확해야 한다는 것이며 둘째, 데이터는 가공이 용이해야 한다는 것이다. 이러한 측면에서 데이터의 품질평가를 위한 품질 특성을 기능성, 사용성, 효율성, 신뢰성, 이식성, 준수성이란 관점에서 나누어 평가항목을 제시한다[18,19,20].

4.1. 기능성 평가항목

기능성에 대한 평가 항목은 네 가지로 나누어 제시한다. 가변성과 접근제어성, 정확성, 보안성을 데이터 기능성에 대한 평가항목으로 제시하며 평가에 대한 내용은 아래의 표와 같다.

첫째, 가변성(Exchangeability (Data format based)))을 들 수 있으며 가변성은 시험 기간 동안에 데이터에 대한 포맷의 변환에 있어서 변환을 시도한 시험에 성공적으로 변환된 경우를 평가하는 것으로 평가방법은 아래와 같다.

〈표 2〉 가변성 평가항목

평가항목		가변성(Exchangeability)	
평가방법		Exchangeability=A/B	
평가항목	A	데이터 변환 시험 기간 동안 다른 소프트웨어를 가지고 성공적으로 데이터 포맷을 변환한 수	
	B	변환될 수 있는 전체 데이터 포맷 수	

둘째, 정확성(Precision)의 평가로써 데이터 필드의 각각의 숫자가 얼마나 정확하게 나타나는가를 평가하는 것으로 평가방법은 아래와 같다.

〈표 3〉 정확성 평가항목

평가항목		정확성(Precision)	
평가방법		Precision=A/B	
평가항목	A	요구된 수준에 만족하도록 정확성을 나타낼 수 있는 데이터베이스의 데이터 필드 숫자	
	B	데이터베이스의 전체 데이터 필드의 숫자	

셋째, 접근제어성(Traceability)에 대한 평가항목을 제시한다. 접근제어성에 대한 평가항목은 사용자가 시험 기간 동안에 접근을 시도하였으나 허용된 사용자를 제외하고 접근이 허용되지 않은 경우를 평가하는 평가 방법으로 아래의 표와 같이 평가한다.

〈표 4〉 접근제어성 평가항목

평가항목	접근제어성(Traceability)	
평가방법	Traceability=1-A/B	
평가항목	A	사용자가 데이터에 접근한 경우 접근 기록 데이터베이스에 기록된 숫자
	B	시험 기간동안에 사용자가 데이터에 접근한 수

넷째, 보안성(Security)을 평가하기 위한 평가방안으로 시험 기간동안에 추적을 시도하고 추적에 성공한 횟수를 평가하여 보안성을 측정한다.

〈표 5〉 보안성 평가항목

평가항목	보안성(Security)	
평가방법	Security=1-A/B	
평가항목	A	공식적인 추적시험 동안에 성공적으로 추적에 성공한 숫자
	B	추적을 시도한 숫자

4.2.. 사용성 평가항목

데이터에 대한 품질 특성 중 사용성에 대한 평가항목은 총 8가지 측면에서 고려하였으며 평가항목은 표현일관성, 접근성, 조작용이성, 해석가능성, 현시성, 완전성, 이해성, 입출력이해가능성으로 나누어 평가한다. 첫째, 데이터에 대한 표현이 어느 정도 일치하는가를 평가하는 표현일관성(Consistency)에 대한 평가는 아래와 같이 제시한다.

〈표 6〉 표현일관성 평가항목

평가항목	표현일관성(Consistency)	
평가방법	Consistency=A/B	
평가항목	A	파일에서 다른 데이터와 표현이 일치하는 데이터의 수
	B	파일에 기록된 데이터의 수

둘째, 접근성(Accessibility)에 대한 평가는 사용자가 데이터를 사용하기 위해서 접근한 경우에 접근에 성공한 횟수를 평가하는 것으로 아래와 같이 제시한다.

〈표 7〉 접근성 평가항목

평가항목	접근성(Accessibility)	
평가방법	Accessibility=A/B	
평가항목	A	사용자가 성공적으로 접근한 데이터의 수
	B	이용가능 한 데이터의 수

셋째, 사용자 관점에서 데이터를 조작할 경우에 쉽게 조작이 가능한가를 평가하기 위한 조작용이성(Ease of handling)은 일정 수의 사용자를 중심으로 하여 데이터 조작을 시도하여 보고 그 중 조작이 용이하다고 평가되어지는 데이터 수를 세어서 평가한다.

〈표 8〉 조작용이성 평가항목

평가항목	조작용이성(Ease of handling)	
평가방법	Ease of handling=A/B	
평가항목	A	사용자가 데이터 조작을 쉽게 한 수
	B	사용자가 데이터 조작을 시도한 수

넷째, 해석가능성(Possibility of explain)에 대한 평가로써 데이터 파일에 사용된 언어, 심볼, 단위 등에 대해서 어느 정도 이해가 가능한가를 평가하는 것으로 시험에 사용된 데이터 파일의 언어, 심볼, 단위 등을 시험에 참여한 사용자가 어느 정도 이해하였는지 숫자를 세어서 평가한다.

〈표 9〉 해석가능성 평가항목

평가항목	해석가능성(Possibility of explain)	
평가방법	Possibility of explain=A/B	
평가항목	A	데이터 파일에서 이해된 언어(심볼, 단위)의 수
	B	데이터 파일에서 사용된 언어(심볼, 단위)의 전체수

다섯째, 현시성(Currentness)을 평가하는 것으로

현재 데이터는 빠른 수준으로 변화되고 있으므로 이러한 데이터에 대해서 어느 정도 빠르게 업데이트 되는지를 평가하는 것으로 조사된 데이터의 숫자 중에서 현재 요구하는 수준의 최신 데이터의 수가 어느 정도인지를 평가한다.

〈표 10〉 현시성 평가항목

평가항목	현시성(Currentness)	
평가방법	Currentness=A/B	
평가항목	A	현재 요구하는 수준의 최신 데이터의 수
	B	조사된 데이터의 수

여섯째, 완전성(Completeness)은 전체의 데이터 중에서 실제 파일에 각각의 데이터가 결손 없이 기록되어진 정도를 평가하는 것이다.

〈표 11〉 완전성 평가항목

평가항목	완전성 (Completeness)	
평가방법	Completeness=A/B	
평가항목	A	사용하려는 데이터가 실제 파일에 기록된 데이터 수
	B	전체 데이터 수

일곱째, 이해성(Understandability)에 대한 평가를 제시한다. 이해성에 대한 평가는 전체 데이터 수 중에서 메타데이터가 자동적으로 실제 구축된 데이터와 연결되어서 이해할 수 있는 정도를 평가하는 것이다.

〈표 12〉 이해성 평가항목

평가항목	이해성(Understandability)	
평가방법	Understandability=A/B	
평가항목	A	메타 데이터가 자동적으로 실제 데이터와 연결되어 이해할 수 있는 데이터의 수
	B	전체 데이터의 수

여덟째, 데이터의 이해가능성 중에서 특히 입출력에 대한 이해가능성(Understandable Input and Output)을 평가하는 평가항목이다. 데이터의 경우 입출력에 크게 좌우되어지므로 사용자에게 일정

수준의 입출력을 시도하도록 하고 이렇게 시도된 것 중에서 이해가 가능한 것의 수를 세어서 평가하는 것으로 한다.

〈표 13〉 입출력 이해가능성 평가항목

평가항목	입출력 이해가능성 (Understandable Input and Output)	
평가방법	Understandable Input and Output=A/B	
평가항목	A	사용자가 이해하는데 성공한 입출력 데이터의 수
	B	사용자가 이해하려고 시도한 입출력 데이터의 수

4.3. 효율성 평가항목

품질 특성 중에서 효율성에 대한 평가항목은 유용성과 효율성 두 가지를 제시한다. 첫째, 유용성(Availability)에 대한 평가는 백업이나 리스트아 되는 동안에 이용될 수 있는 파일의 수를 평가하는 것으로 아래와 같이 평가항목을 제시한다.

〈표 14〉 유용성 평가항목

평가항목	Availability(유용성)	
평가방법	Availability=A/B	
평가항목	A	사용자가 사용하기를 시도 하였을 때 사용자가 성공적으로 데이터 사용이 판명된 전체수
	B	관찰시간 동안에 사용자가 데이터를 사용하기 시도한 것으로 판명된 전체 숫자

둘째, 효율성(Efficiency)에 대한 평가를 제시한다. 효율성에 대한 평가는 실제 파일에서 데이터에 할당된 공간 중에서 최고로 사용된 공간을 측정하여 평가하는 것으로 한다.

〈표 15〉 효율성 평가항목

평가항목	효율성(Efficiency)	
평가방법	Efficiency(External)=sum(A-B)	
평가항목	A	실제 파일의 임의의 데이터에 대해서 최고로 사용된 공간
	B	실제 파일에 임의의 데이터에 대해서 할당된 공간

4.4. 신뢰성 평가항목

데이터 품질 특성 중 신뢰성에 대한 평가는 상당히 중요한 요인으로 보여지며 복구성과 신뢰성에 대한 평가를 통해서 신뢰성을 평가한다. 첫째, 복구성(Recoverability)은 백업이나 리스토아가 진행되어지는 파일 중에서 성공적으로 백업이나 리스토아가 진행된 파일수를 평가하여 평가항목을 측정한다.

〈표 16〉 복구성 평가항목

평가항목		복구성(Recoverability)
평가방법	Recoverability=A/B	
평가항목	A	성공적으로 백업/리스토아된 파일의 수
	B	백업/리스토아 진행되는 파일수

둘째, 신뢰성(Reliability)에 대한 평가는 신뢰성을 평가하기 위해서 신용 위험 평가 과정에 사용된 데이터의 수 중에서 내부 회계감사 과정에 포함된 신용 위험 평가에 의해서 이상으로 검출된 데이터의 수를 평가하여 평가를 시도한다.

〈표 17〉 신뢰성 평가항목

평가항목		신뢰성(Reliability)
평가방법	Reliability=1-A/B	
평가항목	A	내부 회계감사 과정에 포함된 신용 위험 평가에 의해서 이상으로 검출된 데이터 수
	B	신용 위험 평가 과정에 사용된 데이터의 수

4.5. 이식성 평가항목

데이터에 대한 이식성 평가는 변화성과 이식성을 평가항목으로 제시한다. 첫째, 변화성(Changeability)에 대한 평가는 업데이트를 시도하여 업데이트에 성공한 데이터의 수를 세어서 평가하는 것으로 한다.

〈표 18〉 변화성 평가항목

평가항목		Changeability(변화성)
평가방법	Changeability=A/B	
평가항목	A	성공적으로 업데이트가 된 데이터의 수
	B	업데이트를 시도한 데이터의 수

둘째, 이식성(Portability)에 대한 평가는 데이터가 다른 소프트웨어에 대체되어 사용되어지고 연속적으로 재사용을 시도하였을 경우 재사용에 성공한 데이터의 수를 세어서 평가한다.

〈표 19〉 이식성 평가항목

평가항목		Portability(이식성)
평가방법	Portability=A/B	
평가항목	A	다른 소프트웨어에 대체되어지고 연속적으로 사용가능하도록 확인된 데이터의 수
	B	다른 소프트웨어에 사용되어지도록 연속적으로 재사용되도록 계획되어진 데이터의 수

4.6. 준수성 평가항목

데이터 품질에 대한 준수성(Regulatory compliance) 평가항목은 현재 사용하고 있는 데이터에 대해서 준수하여야 할 기본적인 표준 등을 잘 준수하고 있는지를 평가한다.

〈표 20〉 준수성 평가항목

평가항목		Regulatory compliance(준수성)
평가방법	Regulatory compliance=A/B	
평가항목	A	명세 되어진 것으로서 정확히 준수 되어진 숫자
	B	준수하도록 요구되어진 전체 숫자

본 연구에서는 데이터에 대한 품질을 평가하기 위해서 기능성 4개, 사용성 8개, 효율성 2개, 신뢰성 2개, 이식성 2개, 준수성 1개의 평가항목을 제시하여 총 19개의 평가항목을 제시하였다.

19개 시험 평가 항목에 대해서 직접적인 시험 결과를 산출하기 위하여 전국의 우편번호를 구축한 데이터베이스 수업용 공개 자료를 이용하여 확인하여 보았다. 구축된 자료는 총 40,084개이며 6개의 필드로 구성된 자료이다. 본 자료를 이용하여 위의 시험 항목을 접목시켜서 시험 평가하여 본 결과는 아래와 같다.

〈표 21〉 시험 결과 종합표

특성	부특성결과							
	가변성	정확성	접근 제어성	보안성				
기능성	0.9	0.9	1	1				
사용성	표현 일관성	접근성 용이성	조작 능성	해석가 능성	현시성	완전성	이해성	입출력이 해가능성
효율성	유용성 1	효율성 1						
신뢰성	복구성 0.9	신뢰성 0.9						
이식성	변화성 1	이식성 0.9						
준수성	준수성 1							

위의 시험 결과를 통해서 최종적인 데이터 품질 평가의 성적표는 각 항목별 점수를 합산하여 간단한 산술평균으로 산출하였으며 평가 결과는 아래와 같다.

〈표 22〉 품질 특성에 대한 집계표

품질 특성	기능 성	사용 성	효율 성	신뢰 성	이식 성	준수 성	평균
결과 값	95	94	100	90	95	100	95.67

위의 시험 결과 평균값을 기준으로 시험관은 제품의 품질에 대한 판정을 내리게 된다. 국내의 소프트웨어 시험 평가에 있어서 6가지 품질 특성에 대한 가중치 적용 문제를 계속적으로 연구하고 있으나 제품에 대한 분류 기준이 명확하지 못하여 아직 적용하지 못하고 있는 실정이다. 산업용 소프트웨어의 경우 신뢰성에 대한 중요성을 인식하고 가중치를 적용하여 평가하려는 준비를 하고 있으나 아직 적용되지 못하고 있으며 앞으로 계속적인 연구 실험 자료를 통해서 개선되어 져야 할 것으로 보여진다. 본 연구의 시험 평가 메트릭은 소프트웨어에 활용되어지는 데이터 품질이란 관점에서 국제 표준을 중심으로 평가 메

트릭을 제시한 것임으로 앞으로 계속적인 연구를 통해서 6가지 품질 특성에 대한 가중치 적용문제를 연구하여야 할 것이다. 데이터에 대한 중요도를 고려하여 본다면 시험 평가 결과 95점 이상을 시험의 합격 기준으로 하고 95점 미만의 경우 오류로 지적된 사항을 수정하여 재시험을 받을 수 있도록 경고 조치를 취하여야 할 것이다. 데이터에 대한 품질 평가는 본 연구에서 제시된 내용 외에 많은 평가 방안이 연구되어져야 할 것이며 데이터에 대한 품질 평가의 국내 표준에 대한 연구를 통해서 일관성 있는 평가가 이루어져야 할 것이다.

5. 결론 및 향후 연구과제

본 연구를 통해서 데이터에 대한 품질을 평가하여 보았다. 정량적이고 객관적인 평가를 위해서 19개의 평가 메트릭에 적용하여 평가를 실시하였으며 이러한 평가를 통해서 데이터에 대한 품질 향상을 기대하게 되어질 것이다. 데이터에 대한 품질 향상을 위해서는 국제 표준 ISO/IEC 25012[7]의 연구 내용을 중심으로 객관적인 평가가 가능한 평가항목을 구성할 수 있도록 계속적인 연구가 진행되어져야 할 것이다. 앞장에서 제시한 19개 항목과 함께 데이터에 대한 품질을 평가하기 위해서는 데이터가 치우치지 않고 편견 없이 작업을 처리하는데 있어서 공평한 정도를 나타내는 객관성에 대한 평가와, 데이터가 사용자 업무 수행에 적용가능하며 도움이 되는 정도를 평가하는 관련성에 대한 평가, 데이터가 유익하며 데이터의 사용을 통하여 이익을 가져다주는 정도를 평가하는 이익성의 평가, 데이터가 설명이 잘되어 있고 정확함이 확인되며 쉽게 출처와의 연관성을 알 수 있는 정도를 평가하는 추적가능성에 대한 평가, 적절하게 데이터를 수집하는데 필요한 비용이 타당한가를 평가하는 비용효율성에 대한 평가, 데이터의 출처가 다양하고 많은 정보를 주는가를 평가하는 다양성에 대한 평가 등은

데이터 품질 평가에 있어서 고려되어져야 할 사항으로 보여진다. 데이터에 대한 품질 평가 항목은 데이터의 쓰이는 용도에 따라서도 상당히 다양할 수 있으므로 정확한 평가를 위해서 계속적인 연구가 병행되어져야 할 것이다. 이러한 연구는 데이터의 신뢰성[11-16]을 증가시킬 수 있으며 정확한 정보를 얻는 기초 과정이 되어질 것이다.

참 고 문 헌

- [1] ISO/IEC 9126, "Information Technology - Software Quality Characteristics and metrics - Part 1, 2, 3.
- [2] ISO/IEC 12119, "Information Technology - Software Package - Quality requirement and testing".
- [3] ISO/IEC 14598, "Information Technology Software Product Evaluation-Part 1,2,3,4,5,6.
- [4] ISO/IEC 25000 (Software and System engineering: Software product Quality Requirements and Evaluation(SQuaRE) - Guide to SQuaRE), 2005. 10, FDIS
- [5] ISO/IEC 25001(Software engineering: Software product Quality Requirements and Evaluation(SQuaRE) - Planning and management), 2005. 5, CD.
- [6] ISO/IEC 25010(new) Software engineering: Software product Quality Requirements and Evaluation(SQuaRE) - Quality model
- [7] ISO/IEC 25012(new) Software engineering: Software product Quality Requirements and Evaluation(SQuaRE) - Data Quality model
- [8] ISO/IEC 25020(Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) - Measurement reference model and guide)
- [9] ISO/IEC 25021(Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) - Measurement primitives)
- [10] ISO/IEC 25030(Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) - Quality requirements)
- [11] ISO/IEC 25040(Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) - Evaluation reference model and guide)
- [12] Stephen H.Kan, "Metrics and Models in Software Quality Engineering", Addison-Wesley, 2003.
- [13] Dirk Meyerhoff & Begona Laibarra, Rob Van Der Pouw Kraan, Alan Wallet., 'Software Quality and Software Testing in Internet Times', Springer Press, 2002.
- [14] Langberg, N. & Singpurwala, N.D, "A Unification of some Software Reliability Model", SIGM Journal on Scientific and Statistical Computation, pp.781-790, 1985.
- [15] Littlewood,B & Sofer,A., "A Bayesian Modification to the Jelinski-Moranda software reliability growth model", IEE/BCS Software Engineering Journal, pp.31-41, 1983.
- [16] Littlewood, B & Verral, J. L., "A Bayesian reliability model with a stocastically monotone failure rate", IEEE Trans. Reliability, pp.108-114, 1974.
- [17] Morris, C. N., "Parametric empirical Bayes inference: Theory and application". J.American Statistical Association, vol 78, pp.47-65, 1983.
- [18] 한국데이터베이스 진흥센터, "데이터 품질관리 지침(Ver 2.1)", 2006.
- [19] 윤영미, "생체인식 소프트웨어 품질 평가 모듈에 관한 연구", 한국정보처리학회 논문지 D 제13권 5호, pp. 699-708, 2006.
- [20] 이하용, "의료기기 내장형 소프트웨어 품질 평가 시험 모듈 개발", 한국정보처리학회 논문지 D 제13권 2호, pp.259-270, 2006.

● 저자 소개 ●



정 혜 정(Hye-Jung Jung)

1988년 경북대학교 통계학과 졸업(학사)

1991년 경북대학교 대학원 통계학과 졸업(석사)

2004년 경북대학교 대학원 통계학과 졸업(박사)

1995~현재 평택대학교 디지털응용정보학과 교수

관심분야 : 소프트웨어 신뢰성 공학, 소프트웨어 품질 평가, 소프트웨어 품질 평가에 대한 표준화연구, 소프트웨어 용어 표준화 연구, etc.

E-mail : jhjung@ptu.ac.kr