

## 데이터 품질진단을 위한 자동화도구 개발

고재환\* · 김동수\*\* · 한기준\*\*\*

### Development of Automated Tools for Data Quality Diagnostics

Jae-Hwan Ko\* · Dong-Soo Kim\*\* · Ki-Joon Han\*\*\*

#### ■ Abstract ■

When companies or institutes manage data, in order to utilize it as useful resources for decision-making, it is essential to offer precise and reliable data. While most small and medium-sized enterprises and public institutes have been investing a great amount of money in management and maintenance of their data systems, the investment in data management has been inadequate. When public institutions establish their data systems, inspection has been constantly carried out on the data systems in order to improve safety and effectiveness. However, their capabilities in improving the quality of data have been insufficient.

This study develops an automatic tool to diagnose the quality of data in a way to diagnose the data quality condition of the inspected institute quantitatively at the stage of design and closure by inspecting the data system and proves its practicality by applying the automatic tool to inspection.

As a means to diagnose the quality, this study categorizes, in the aspect of quality characteristics, the items that may be improved through diagnosis at the stage of design, the early stage of establishing the data system and the measurement items by the quality index regarding measurable data values at the stage of establishment and operation. The study presents a way of quantitative measurement regarding the data structures and data values by concretizing the measurement items by quality index in a function of the automatic tool program.

Also, the practicality of the tool is proved by applying the tool in the inspection field. As a result, the areas which the institute should improve are reported objectively through a complete enumeration survey on the diagnosed items and the indicators for quality improvement are presented quantitatively by presenting the quality condition quantitatively.

Keyword : Data Quality, Quality Diagnostic, Data Profiling, Automation Tools

## 1. 서 론

기업이나 기관에서 관리하는 데이터가 경영의사 결정에 유용한 정보로 활용되기 위해서는 정확하고 신뢰할 수 있는 정보의 제공이 필요하다. 급변하는 비즈니스 환경에서 데이터의 수집 빈도는 과거보다 늘어났으며 이러한 환경에서 기업에 낮은 품질의 데이터가 유입 될 경우 조직적인 측면에서는 의사결정에 대한 지원이 적절히 이루어지지 않아 전략 수립이 어려워지고, 업무측면에서도 적합한 데이터가 제공되지 않아 고객만족도나 현업의 정보시스템에 대한 활용도가 낮아지게 된다. 또한, 시스템 측면에서는 시스템 간 연계와 통합이 어렵게 되고, 시스템 마다 중복된 데이터로 인하여 관리비용도 증가하게 된다. 이제는 데이터를 자산으로 인식할 필요가 있으며, 데이터 품질 개선을 위한 추가적인 노력이 필요하다. 데이터의 품질진단을 수행하는 이유 중 하나는 현 데이터 품질 수준을 인식하고 품질진단 결과 현 문제점의 인지를 통해 추가적인 품질관리 활동을 수행하게 함으로써 고품질의 데이터를 확보하는데 있다.

그러나 대부분의 중소기업이나 공공기관에서는 정보 활용에 대한 투자는 지속적으로 이루어지고 있으나, 정보관리에 대한 투자는 아직 미진한 편이며, 저 품질의 데이터로 인하여 정보시스템의 관리 및 유지보수에 많은 비용이 소요되고 있는 실정이다.

또한, 공공기관의 정보시스템 구축시 시스템의 효율성을 향상시키고 안전성을 확보하기 위한 목적으로 정보시스템 감리가 지속적으로 수행되고 있으나, 데이터 품질에 대해서는 아직도 관리가 미흡한 수준에 있다.

한국데이터베이스진흥원에서 2010년 2월 25일 ~ 4월 29일 기간 동안 공공기관 및 민간 기업을 대상으로 조사한 데이터 품질에 대한 인식과 활동 수준 설문평가 보고 자료를 살펴보면 전반적으로 데이터 품질에 대해 필요성은 공감하지만 구체적인 활동이 미흡한 것으로 조사되었다.

본 연구에서는 정보시스템 감리의 설계단계와 종료단계에서 수검기관의 데이터 품질 현황을 객관적으로 진단할 수 있는 점검항목을 도출하고, 점검결과에 대한 신뢰성 확보와 정형화된 평가방안을 자동화도구와 함께 제시할 수 있도록 관련 연구를 수행하였다.

## 2. 관련 연구

### 2.1 데이터 품질 평가를 위한 연구

데이터 품질을 측정하기 위해서는 먼저 품질평가를 수행하기 위한 기준이 제시되어야 한다. 소프트웨어의 제품관점에서 품질특성을 정의한 ISO9126의 품질기준은 기능성, 신뢰성, 사용성, 효율성, 유지보수성, 이식성으로 정의되어 있다. 한 예로 기능성에 대한 품질특성을 살펴보면 특정 조건 하에서 명시된 요구와 내재된 요구를 수행하기 위한 기능을 제공하는 S/W 제품의 능력을 말하는 것으로, 사용자의 요구를 측정하기 위해서 S/W가 무엇을 하는가에 관심을 두고 있다. 기능성의 세부 특성은 아래 표와 같이 적합성, 정확성, 상호운용성, 보안성, 준수성으로 정의하고 있다[5].

<표 1> ISO/IEC9126의 품질특성 중 기능성의 품질부특성

품질 부특성	내용
적합성 (suitability)	지정된 작업과 사용자 목적을 위한 적절한 기능들을 제공하는 능력
정확성 (accuracy)	올바른 효능 결과를 제공할 수 있는 능력
상호운용성 (interoperability)	하나 이상의 시스템과 상호작용 용수 있는 능력
보안성 (security)	권한에 따라 정보에 대한 접근을 차단하거나 허용하는 능력
준수성 (compliance)	기능과 관련된 표준, 관례 또는 규제를 고수하는 능력

김수경은 ISO9126의 품질모형을 기반으로 데이터 품질특성을 제시하였으며, 제시한 데이터 품질

특성은 객관성, 정확성, 완료성, 접근성, 관련성, 신용성, 이해성, 해석성, 간결한 표현성, 적시성, 일관된 표현성, 검증가능성, 조작 편의성의 13가지를 제시하였다[3].

김문영은 데이터품질 평가요소를 크게 데이터의 내용, 데이터의 구조, 데이터의 이동과 흐름관점에서 바라보았다. 데이터 내용에 대한 품질평가는 오류 데이터를 찾아내고 적합성 검사를 실시하는 데이터 내용에 대한 품질평가 활동으로 수행되고, 데이터의 구조에 대한 평가는 단순히 참조 무결성이나 유일성, 관계검증, 정규화 뿐 아니라 기업의 정보구조를 체계적으로 나타내기 위해 데이터를 사용자 및 전사적 관점에서 인식·분석하고 정의하였는지를 평가한다. 데이터의 이동과 흐름에 대한 평가는 데이터를 최종사용자에게 의미 있는 정보나 지식으로 변환하기 위한 과정으로, 적시성(Time-liness)과 편리성의 관점에서 평가하였다[2].

## 2.2 국제표준 연구현황(ISO/IEC25000소개)

국제표준 연구내용인 ISO/IEC25000는 기존의 ISO/IEC9126의 품질모델을 기반으로 하여 소프트웨어제품품질 요구사항 및 테스트 절차를 규정한 ISO/IEC12119(Software Packages-Quality Requirements and Testing)와 소프트웨어 제품을 평가하기 위한 방법 및 절차를 규정한 ISO/IEC14598

(Software Product Evaluation) 표준을 대체하는 새로운 국제표준이다. 이 표준은 개별적으로 관리되던 소프트웨어의 품질표준이 하나의 프레임워크로 통합되었다는 측면에서 의미를 갖고 있다. ISO/IEC 25000 시리즈의 구성은 계획과 관리, 측정 참조 모델, 품질 요구사항, 품질모델 매트릭스, 품질평가 매트릭스로 구성되며, 각각의 구성내용은 <표 2>와 같다[13].

정혜정은 데이터 품질평가 관점에서 국제표준 ISO/IEC25000 프로젝트 중에서도 ISO/IEC 25012의 데이터 품질평가 모델에 대한 연구를 수행하였으며, 데이터 품질에 대한 정량적 평가방안에 대한 연구를 수행하여 각각의 품질평가 항목과 측정방안을 제시하였다[7].

현재 진행 중인 국제표준 ISO/IEC25012의 데이터 품질에 대한 연구는 소프트웨어 제품관점에서 다루어지는 데이터 품질평가방안을 제시하고 있는 반면 국내 데이터 품질관리 지침연구에서는 데이터 자체 품질에 대한 평가방안을 다루고 있다는 점에서 연구의 기본 방향이 다르다고 할 수 있다[7].

## 2.3 국내 데이터 품질관련 연구

### 2.3.1 데이터 품질관리 프레임워크

데이터 품질관리란 조직 내외부의 지식 노동자와 최종 사용자의 기대를 만족시키기 위한 지속적

<표 2> ISO/IEC 25000 시리즈의 구성

프레임워크 구성	내용
ISO/IEC 2500n (Quality Management)	품질관리 부문은 소프트웨어 제품 요구 사항 명세 및 평가의 관리에 대한 책임을 지원하는 기능 또는 요구사항 및 지침 제공한다.
ISO/IEC 2501n (Quality Model)	품질 모델 부문은 표준 사용 내부, 외부 및 품질에 대한 특성을 포함하는 상세한 품질모델을 제공하고 품질 모델의 사용에 대한 실용적인 지침까지 제공한다.
ISO/IEC 2502n (Quality Measurement)	품질 측정 부문의 표준은 소프트웨어 제품 품질 측정 참조 모델, 품질 조치의 수학적 정의, 그들의 응용 프로그램에 대한 실용적인 지침을 포함한다.
ISO/IEC 2503n (Quality Requirement)	품질 요구 사항 부문의 표준은 품질의 요구 사항을 지정하는 데 도움을 주고 이러한 품질 요구 사항을 개발하는 소프트웨어 제품 또는 평가 과정에 대한 입력으로 품질 요구 사항을 도출하는 과정에서 사용할 수 있다.
ISO/IEC 2504n (Quality Evaluation)	품질 평가 부문은 평가자, 획득자 또는 개발자에 의해 수행 여부, 요구 사항, 권고 사항 및 소프트웨어 제품 평가에 대한 지침을 제공한다.

인 데이터 및 데이터 서비스 개선활동으로 기업내부 데이터 간 연계에서 데이터 일관성 유지를 위해 수행하는 활동으로도 볼 수 있다.

국내의 데이터 품질관리를 위한 연구 내용으로는 한국데이터베이스진흥원(KDB)의 품질관리 프레임워크가 있으며, 이것은 EA의 프레임워크를 벤치마크 하여 개발되었다. EA의 프레임워크는 View와 Perspective로 구분되며 View는 비즈니스아키텍처, 데이터아키텍처, 애플리케이션아키텍처, 기술아키텍처로 구성되고, Perspective는 관련된 조직 계층을 보여주며 개괄적, 개념적, 논리적, 물리적 관점을 보여준다. KDB의 데이터 품질관리(DQ) 프레임워크도 View와 Perspective로 구분되며, X축은 데이터 품질관리 대상의 기준이 되는 View를 보여주고, Y축은 조직 계층별 관점을 나타낸다. 품질관리 대상인 View는 데이터, 데이터구조, 데이터관리 프로세스로 구분되며 3가지 구성요소는 궁극적인 데이터 값이 품질을 향상시키기 위해 데이터구조에 대한 품질을 관리하고, 지속적으로 품질을 유지하기 위해서는 데이터관리 프로세스를 준수해야 하는 상호연관성을 가지고 있다[4].

데이터 품질관리 프레임워크에서 첫 번째 관리 대상 항목인 표준 데이터는 정보시스템에서 사용하는 용어, 도메인, 코드 및 기타 데이터 관련 요소에 대해 공통된 형식과 내용을 정의하여 적용하는 표준관련 데이터를 의미한다. 표준 데이터를 기반으로 한 데이터 표준화는 여러 시스템에 산재해 있는 메타데이터에 대한 한글(논리)명/영문(물리)명 명명규칙, 데이터 타입, 데이터길이 등에 대한 기준을 제시하여 데이터의 구조적 품질을 향상시키고자 하는 활동에 해당한다.

표준 데이터는 정보시스템의 데이터 품질확보를 위한 기본요소로 표준 데이터를 정의하고 관리함으로써 데이터 간의 불일치와 데이터 오류를 방지하고 데이터에 대한 이해도를 높일 수 있다.

품질관리 프레임워크에서 구조적 측면의 관리대상인 모델데이터는 데이터 모델을 운용 관리하는데 필요한 데이터를 의미한다. 여기에는 데이터참조

모델, 개념 데이터모델, 논리 데이터모델, 물리데이터모델에 대한 메타데이터 및 DBMS 객체정보가 포함된다. 데이터 모델에 대한 메타데이터 관리의 데이터 구조에 대한 최신정보 유지 및 전사차원에서 데이터 모델의 공유와 재사용성을 극대화하고 체계적인 데이터모델의 변경관리를 가능하게 한다[8]. 데이터의 품질은 데이터 값 자체에 대한 품질관리에서 출발하여 근원적인 문제해결을 위한 데이터구조에 대한 품질관리로 확장된다. 데이터의 구조적 측면에서의 세부품질 검증대상 항목으로 코드 인스턴스의 중복여부, 선분이력의 중복여부, 제약조건 위반여부, DBMS 객체생성의 완전성, 모델 정규화 등이 있다.

### 2.3.2 데이터 품질관리 성숙모형

한국데이터베이스진흥원에서는 데이터베이스 운영 및 평가, 개선관점에서 데이터 품질관리 성숙모형을 제시하고 있으며, 데이터 품질관리 성숙수준을 도입 → 정형화 → 통합화 → 정량화 → 최적화의 5단계로 구분하여 데이터의 정확성, 일관성, 유용성, 접근성, 적시성, 보안성의 데이터 품질 기준별 관리 성숙수준을 측정하고 개선해 나갈 수 있도록 체계적인 접근방법을 제시하고 있다[9]. 이 품질관리 성숙모형은 데이터 품질관리 수준을 객관적으로 정량화하여 현재의 품질관리 수준을 파악하고 개선방향을 제시한다는 측면에서 의미가 있다.

## 2.4 품질진단 수행활동

### 2.4.1 데이터 품질진단을 위한 관리요소

데이터 품질을 관리하기 위한 업무에서는 운영 데이터 중 어느 부분에 오류가 있고, 그 원인이 무엇인지를 파악하는 활동이 품질진단의 핵심 업무에 해당한다. 데이터 품질진단은 해당 조직이 운영·관리하고 있는 정형·비정형 데이터의 품질을 측정하여 품질목표 수준이 유지되도록 지속적으로 개선하는 절차로 볼 수 있다[10]. 데이터 품질진단에서는 품질지표(DQI, Data Quality Index), 품질

핵심정보(CTQ, Critical to Quality), 업무규칙(BR, Business Rules)을 도출함으로써 데이터에 대한 품질수준을 정량적으로 측정하고 평가할 수 있다. 여기서 품질지표(DQI)란 데이터 사용 목적을 충족시키기 위한 기준으로 지속적으로 품질점검을 통해 관리되어야 할 평가기준을 말하며, 품질핵심정보(CTQ)는 6시그마에서 유래한 용어로 데이터 품질관리 관점에서 기업경영에 중요한 영향을 미치는 데이터 품질관리 대상 정보항목을 의미한다. 업무규칙(BR)은 기업에서 데이터 품질을 위해 관리되는 업무규칙으로 데이터의 값이 정확하기 위한 조건에 대한 표현에 해당된다.

데이터 품질에 대한 평가는 다면평가에 해당되며, 각 기업이나 기관은 운영에 중요한 의미를 가지는 요소가 무엇이지를 먼저 파악하고 이러한 요소를 구성하는 변수를 정확하게 정의하여야 한다. 동일한 변수를 선정하였어도 측정하는 방법은 기업이나 기관의 특성에 따라 달라질 수 있다[12].

<표 3>은 국내의 기관이나 기업에서 제시한 품질지표 항목을 나열한 것으로 이 표에서 공통적인 항목과 자동화도구를 통해 진단이 가능한 항목을 도출하였으며, 본 연구에서 수행할 자동화 도구를 활용한 품질진단 결과의 품질지표 항목으로 적용하였다.

<표 3> 기관별 데이터 품질지표 항목비교

품질 진단 보고서	건설링 기관	K통신사	진단가능성	
완전성	완전성 활용성 충실도 적시성	완전성	진단 가능	완전성
유효성	유효성	유효성		유효성
일관성	일관성 정합성	일관성 명확성		일관성
유일성	유일성	유일성		유일성
-	적수성 대사성	-		정확성
-	민원지수	-		진단 불가능

데이터 품질 기준으로 선정한 항목이 진단대상 기관의 업무특성과 일치하지 않을 수 있으나, 품질기준을 선택적으로 적용하거나 추가적으로 필요한 내용을 기본 점검항목에 포함시켜 진단대상 범위를 확장하거나 축소할 수 있도록 자동화도구 구현시 고려하였다.

## 2.5 데이터 품질평가절차 연구

### 2.5.1 데이터 품질진단 절차

데이터베이스진흥원에서는 데이터 품질진단 절차를 5단계로 나누고 각 단계별 수행내역을 체계적으로 제시하고 있으며 완전성, 유일성, 유효성, 일관성 등의 4가지 관점에서의 데이터 품질측정을 수행하여 원칙과 정책, 요구사항관리, 오너쉽, 데이터 표준, 데이터 구조, 데이터 흐름, 데이터 품질 측면에서의 진단결과를 리포팅하고 있다.

### 2.5.2 데이터 품질진단 기법

데이터 품질진단은 대상기업 또는 기관의 품질관리 수준과 데이터 품질에 대한 요구수준을 이해하고 현재의 품질 수준을 평가하기 위한 절차이다. 따라서 품질진단을 수행하기 위해서는 기업에서 중요하게 여겨지는 품질지표를 상황에 맞게 정의하고, 품질진단 대상에 대한 평가기준 제시를 통해 발견된 오류를 분석하고 개선해 나갈 수 있는 절차와 기법이 필요하다.

데이터 품질진단을 위한 기법에는 설문조사 기법과 데이터프로파일링 기법이 있다. 설문조사 기법은 정보품질평가(IQA) 설문조사 방법으로 데이터 품질에 대한 응답자의 주관적 평가와 응답자의 데이터 품질 프로세스, 프로그램, 사용기법 등에 대한 지식을 0에서 10까지의 등급으로 분류하고, 0은 '전혀 아님'을 10은 '전적으로'를 의미하는 10점 척도로 구성된다[12].

데이터베이스진흥원에서 수행한 2010년 데이터 품질관리 성숙수준 조사 자료에서도 공공기관 및 민간기업의 데이터 품질관리 성숙수준을 분석하기

위한 방법으로 조사표에 의한 온라인 설문조사기법을 활용하였으며, 조사항목으로는 응답기관 정보, 일반현황, 정확성, 일관성, 유용성, 접근성, 적시성, 보안성 등을 선정하여 각 영역별 수준을 측정하였다.

정보시스템 감리에서도 사실적인 자료나 성향과 관련된 자료를 얻기 위해 설문 기법을 활용할 수 있다. 설문기법을 사용할 때에는 설문내용이 유효하고 신뢰성 있는 응답을 얻을 수 있도록 설계하여야 하며, 설문의 개발은 다음과 같은 네 단계를 거쳐 작성된다[4].

- 가. 설문의 설계
- 나. 응답 척도의 설계
- 다. 설계 구조와 배치
- 라. 유효성 및 신뢰성의 확인

데이터 프로파일링은 주로 비정형 텍스트 데이터 및 비정형 콘텐츠의 메타데이터에 대한 품질진단에 활용되며 데이터 품질문제를 이슈화하고 개선점을 찾는 것이 수행의 주된 목적이다[10]. 데이터 프로파일링 분석 내용으로는 컬럼 분석, 날짜 분석, 패턴 분석, 코드 분석, 참조무결성 분석 등이 있다[6].

데이터 프로파일링 기법은 발견(Discovery)과 검증(Verification)이라는 절차로 구성된다. 주요 수행내역으로는 메타데이터를 통해 운영시스템의 구

조를 파악하고, 진단대상 컬럼을 선정한 다음, 분석 작업을 통해 데이터 값에 내재된 오류를 발견하여 업무담당자와 함께 최종 오류 데이터를 검증하는 절차로 수행된다. 프로파일링 기법을 통해 추출된 오류 데이터와 데이터 규칙은 향후 품질개선 활동의 기초 작업이 되는 동시에 데이터 품질유지를 위해 지속적으로 점검해야 할 업무규칙으로 관리된다.

효과적인 품질진단을 수행하기 위해서는 위에서 제시된 것과 같은 정형화된 데이터 품질진단 기법을 적용하여 진단절차를 수립하는 것이 중요하다.

## 2.6 기존 자동화도구에 대한 연구 조사

감리수행시 사업의 진행상황을 정확히 진단하고, 문제점을 지적하기 위해서는 객관적으로 감리의견을 뒷받침 할 수 있는 증적자료의 확보가 요구된다.

광송해는 감리자동화도구에 대한 연구논문에서 감리지원 소프트웨어를 소개하고 실제 감리원이 업무를 수행하는 과정에서 도움을 받을 수 있는 지원도구로써 <표 4>와 같이 데이터 품질 검사용 범용감리 소프트웨어와 보고서 작성용 고급언어, 유틸리티 소프트웨어, 전문가 시스템 등의 감리지원 소프트웨어를 소개하였다[1].

정보시스템감리 수행시 지적사항에 대한 근거

<표 4> 감리지원 자동화도구 소개

종류	기능	단점
범용감리 소프트웨어	<ul style="list-style-type: none"> <li>◦ 데이터 품질검사</li> <li>◦ 시스템 과정들의 품질검사</li> <li>◦ 개체들의 존재여부에 대한 검사</li> </ul>	<ul style="list-style-type: none"> <li>◦ 응용 프로그램 처리 결과에 대해서만 검증가능</li> </ul>
고급언어(SQL)	<ul style="list-style-type: none"> <li>◦ 데이터를 보고서용으로 포맷변환</li> <li>◦ 데이터 다운로드 없이 통계조사나 보고서 작성 가능</li> </ul>	<ul style="list-style-type: none"> <li>◦ 관계형 데이터베이스에 한정</li> <li>◦ 감리원이 직접 조작하여 데이터를 출력하여야 함</li> </ul>
유틸리티 소프트웨어	<ul style="list-style-type: none"> <li>◦ 복제, 정렬, 디스크 검색, 디스크 포맷팅</li> <li>◦ 데이터 포맷 변경 및 보고서 작성 지원</li> </ul>	<ul style="list-style-type: none"> <li>◦ 유틸리티 소프트웨어의 기능 인지 및 활용을 위한 사전 이해 필요</li> </ul>
전문가 시스템	<ul style="list-style-type: none"> <li>◦ 감리원의 의사결정 지원</li> <li>◦ 가용 증거에 대한 판단기준 제시</li> </ul>	<ul style="list-style-type: none"> <li>◦ 전문가 시스템을 사용하는 감리원이 해당 분야에 대한 기본지식을 보유하고 있어야 함</li> </ul>

(fact finding)에 있어 감리인의 직관적 판단과 제한적인 대상검토로 인하여 감리의 객관성과 충분성(Audit coverage) 확보에 어려움이 있으며[11], 감리법인에서는 이러한 문제점을 극복하기 위한 감리수행 방안으로 <표 5>에 기술한 것과 같은 감리 분야별 활용 가능한 자동화도구를 감리제한서에 포함시키고 있으며, 실제 감리현장에서도 이러한 도구를 활용하고 있다.

<표 5> 감리지원 소프트웨어의 종류

분야	기능	소프트웨어
사업관리	사업관리 도구	◦ MS Project
응용 시스템	웹 사이트 구조적 완성도 점검도구	◦ MAXAMINE ◦ Linkbot
	프로그램 소스코드 품질 점검도구	◦ QAJ ◦ Resort for Java ◦ PMD
	웹 접근성 준수	◦ K-WAH 3.0
	웹 표준 분석도구	◦ W3G Validator
데이터 베이스	데이터베이스 모델 작성 도구	◦ ER-Win ◦ DA# Architecture
	SQL문에 대한 적정성 점검도구	◦ SQL Expert
	어플리케이션 개발 및 성능 튜닝 도구	◦ Orange for ORACLE ◦ TOAD
성능 및 보안	어플리케이션에 대한 성능/부하 점검	◦ WAST
	서버의 서비스 포트 및 취약점 점검	◦ Nessus
	소스나 실행파일의 보안취약점 점검	◦ Paros

데이터베이스 감리영역에서 주로 사용되는 자동화도구는 ER-WIN, TOAD/Orange, SQL Expert, 엑셀 등이 있으며, 이러한 도구들을 활용하여 데이터베이스 분야에서 점검되고 있는 항목들을 데이터 값과 데이터 구조 측면에서 분류하여 정리한 내용은 <표 6>과 같으며, 감리현장에서는 감리수행 기간을 고려하여 일부 항목에 대해서만 점검을 수행하고 있다.

<표 6> 감리에서의 데이터 품질 점검 항목

품질지표	데이터 값	데이터 구조
완전성	◦ 데이터 크기 ◦ 데이터 범위 ◦ 데이터 값 누락 여부	◦ 중요속성 누락 여부 ◦ 필수 속성의 설계반영여부
유효성	◦ 컬럼값의 범위 유효성 ◦ 데이터 포맷 유효성 ◦ 데이터 값의 업무규칙 준수	◦ 업무규칙을 관계로 표현
정확성	◦ 데이터의 오타자 여부 ◦ 실제 사실과 일치 여부 ◦ 레코드 중복	◦ 속성중복, 유일성 보장여부
유일성	◦ 데이터 값의 유일성	◦ 데이터 항목의 유일성
일관성	◦ 데이터 속성정의와 값의 일치여부 ◦ 데이터 제약조건과 값의 일치여부 ◦ 테이블정의와 값의 일치 여부 ◦ 동일 데이터의 상호일관성	◦ 참조무결성 ◦ 데이터 표준 정의의 적절성 ◦ 도메인 정의의 적절성 ◦ 코드 정의의 적절성 ◦ 기능, 의미, 성격이 동일한 데이터는 상호 동일한 용어와 형태로 정의

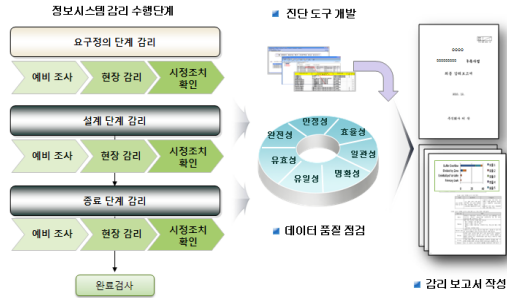
### 3. 자동화도구 제안

정보시스템 감리는 전자정부법 시행령에 의해 2011년 7월부터 요구정의단계, 설계단계, 종료단계의 3단계 감리를 수행하고 있으며, 감리원은 검사 기준서를 바탕으로 과업수행결과에 대한 적합/부적합 판정을 수행하고 있다. 감리수행시 데이터 품질에 대한 진단은 [그림 1]과 같이 설계가 완료된 단계와 DB 구축이 완료된 단계에서 진단업무수행이 가능하다.

본 장에서는 정보시스템 감리업무 중 데이터베이스 분야의 설계와 구현내역에 대한 완전성, 유효성, 정확성, 유일성, 일관성 측면의 품질을 진단하고 진단결과 리포팅 할 수 있는 데이터 품질진

단 도구의 주요기능과 개발 과정에서 수행한 주요 설계내역에 대해 제시한다.

가 적용, 원격 운영 DB의 접근방안 제시 등의 내용을 포함하였다.



[그림 1] 자동화도구 개발 개요

### 3.1 자동화도구 개발을 위한 설계

#### 3.1.1 프로그램 개발을 위한 요건정의

연구대상 자동화도구가 범용성과 효과성을 확보할 수 있도록 설계 및 구현단계에서 고려해야할 개발 요건을 <표 7>과 같이 도출하였으며, 기능 및 비기능 측면에서 도출한 개발 제약조건 내용으로는 품질측정 결과 정량화, 종합적인 품질진단 결과제시, 정보시스템 설계단계에서의 진단항목 도출, 진단대상 운영 DB의 안정성 확보, 업무규칙 추

<표 7> 자동화도구 구현시 고려사항

일련 번호	개발 제약조건	제약 유형
1	◦ 품질지표에 따른 점검항목을 정량화하여 산출할 수 있어야 한다.	기능
2	◦ 각 점검된 내역을 취합하여 진단대상 기관의 품질지표를 종합적으로 제시 한다.	기능
3	◦ 정보시스템의 설계 단계에서 진단할 대상 점검항목을 포함하여 설계에 반영한다.	기능
4	◦ 진단 수행시 운영 DB에 부하가 가지 않도록 안정성을 고려하여야 한다.	비기능
5	◦ 품질속성 확장을 위해 기본적인 점검항목에 추가적인 업무규칙 적용이 가능해야 한다.	기능
6	◦ 원격(Remort) 운영 DB에 쉽게 접근할 수 있는 방법을 마련해야 한다.	비기능

<표 8> 품질지표별 점검항목 구분

DQI	점검항목	대상
완전성	◦ 관계정의 내역을 DB에 LOAD하여 DBMS에 생성된 외래키 제약 수와 비교(구조)	테이블
	◦ 테이블 정의 내역을 DB에 LOAD하여 NOT NULL, PRIMARY KEY, FOREIGN KEY 생성 개수 비교(구조)	컬럼
	◦ 설계(ERD/테이블정의서)의 스키마와 DBMS 스키마를 비교하여 데이터 형이나 컬럼 생성 개수, NULL 정의내역 비교	컬럼
	◦ 테이블 전체 레코드 수 대비 주요 컬럼의 데이터 수 비율	컬럼
	◦ 코드정의와 공통코드 테이블에 정의된 코드 컬럼의 범위유효성	컬럼
유효성	◦ CRUD 매트릭스 점검-사용되지 않는 테이블(구조)	테이블
	◦ MAX, MIN 값 검증	컬럼
	◦ 포맷 검증 ex) 컬럼 값이 날짜유형일 경우에는 유효날짜 값을 가져야 한다	컬럼
	◦ 통계 값 및 설계 속성 검증	컬럼
정확성	◦ BR 규칙 도출 및 점검	컬럼
	◦ Analyze 일자의 최신성	테이블
	◦ 통계 값 및 설계 속성 검증	컬럼
유일성	◦ 컬럼값의 유일성 확인을 위해 PK가 없는 테이블 점검	테이블
	◦ FK 컬럼을 제외하고 동일 컬럼의 중복 정의 정도 검증(구조)	컬럼
일관성	◦ 동일 영문 컬럼에 대한 한글컬럼의 유일성	표준
	◦ 동일 영문 컬럼에 대한 도메인 정의의 유일성	표준
	◦ 동일 한글 컬럼에 대한 영문 컬럼의 유일성	표준
	◦ 표준용어를 loading하여 DBMS 또는 ERD에 정의한 영문 컬럼/한글 컬럼의 표준용어 적용여부 점검	표준
	◦ 참조무결성(RI) 오류 검증(구조)	테이블

주) ※(구조)는 구조적 측면의 품질 점검임.

자동화도구를 이용한 품질진단을 수행하기 위해



서는 품질지표 항목별 데이터 값 또는 데이터 구조 측면에서의 품질 진단항목 도출이 필요하다. 이와 관련하여 <표 8>에서는 데이터 값과 데이터 구조, 데이터 표준화 측면에서 품질진단과 관련된 점검항목을 도출하였다. 완전성과 관련된 품질진단 항목으로는 설계내역과 구현내역 간의 완전성 비교를 통해 정량적인 품질지표 산출이 가능하며, 유효성에 대해서는 데이터 값의 범위 유효

성을 진단함으로써 유효성에 대한 품질지표 값을 정량적으로 계산해 낼 수 있다. 또한, 정확성과 관련하여서는 업무규칙이 정확하게 지켜지고 있는지 준수율을 측정한다.

<표 9>에서는 점검 항목을 표준화, 테이블, 컬럼의 3가지 관점에서 분석하여 품질지표별 점검 항목을 도출하였으며, 자동화도구 개발을 위한 데이터 베이스 설계와 응용 프로그램 설계에 반영하였다.

<표 10>에서 제시한 컬럼 유형별 점검내역은 점검대상 데이터 항목별 점검유형을 제시한 것으로 분석대상 컬럼에 대한 데이터 프로파일링을 실시할 경우 컬럼 속성에 따른 점검요건에 해당하는 내용으로 자동화도구에서 프로파일링 진단기능 구현 시 개발요건으로 고려하였으며, 현장에서 진단시에도 선정된 핵심 품질항목(CTQ)의 도메인 특성을 고려하여 업무규칙(BR)을 도출하는데 참고할 수 있는 내용이다.

<표 9> 품질지표별 점검내역

분석관점		점검항목
표준화	표준적용의 적정성	표준 미준수 컬럼 조회
	컬럼정의 일관성	영문/한글 기준 점검 결과
테이블	업무분석의 적정성	CRUD 매트릭스 분석 및 점검 결과 조회
	테이블 스키마 검증	통계정보 생성 여부 컬럼 개수 NOT NULL 정의 수 PK 생성 여부 컬럼 명칭 미 매칭 수 컬럼 TYPE 미 매칭 수
	참조무결성 검증	참조무결성 점검 오류
컬럼	컬럼 공통 분석	총 건수(Record Count) 개별 건수(Distinct Count) 널 건수(Null Count) 스페이스 건수(Space Count) 제로건수(Zero Count) 개별 건수 퍼센트(Distinct Percent) 널 퍼센트(Null Percent), 스페이스 퍼센트(Space Percent) 제로퍼센트(Zero Percent)
	패턴분석	비정형 데이터의 패턴을 분석하여 표준 패턴에 위배된 도메인 값들을 확인
	날짜분석	날짜 유형의 데이터에 대하여 위배된 도메인 값들을 확인
	값의 범위 유효성 분석	개별코드, 통합코드에 대하여 코드정보와 코드값들을 확인 해당 컬럼의 데이터의 Min, Max값 및 해당 값의 리스트를 확인하여 값의 허용 범위를 확인

<표 10> 컬럼 유형별 점검내역

점검 항목	점검 대상 데이터 항목 예	점검유형
번호	주민등록번호, 여권번호, 계좌번호, 사업자 등록번호	패턴 점검 체크비트 점검
금액	금액, 세금, 가격, 단가, 비용, 요금, 잔액, 총액	허용범위 점검
명칭	명, 주소, ID, 장소, 고객명	패턴 점검
수량	건수, 매수, 회차, 개수, 거리, 규모, 길이, 무게, 속도, 면적, 온도	허용범위 점검
분류	여부, 유무, Indicator	표준코드 점검
날짜	년월, 년, 연월일, 시분초, 일, 반기, 분기	날짜 점검
율	금리, 이율, 비율, 환율, 백분율	허용범위 점검
내용	내용, 비고, 설명, 정보, 요약	일부 패턴 점검
코드	개별코드, 통합코드, ~구분	표준코드 점검
Key	PK, FK(사번, 조직)	참조무결성 점검



럼에 대하여 정형화된 업무규칙을 적용하여 오류 내역에 대한 통계 데이터를 제공한다.

〈표 12〉 자동화도구의 기능정의

업무기능	기능설명	정보구성
스키마 분석	테이블 목록수준에서 설계문서와 DBMS 비교	테이블 정의서 DBMS 스키마
표준분석	설계내역과 표준정의의 비교	표준메타정보 테이블 정의서
데이터 정합성 분석	ERD 관계정보를 이용하여 쿼리로 점검	테이블 정의서 관계정보 관계컬럼
기능/데이터 상관분석	CRUD 매트릭스를 쿼리로 점검	테이블 정보 프로세스 정보
데이터프로파일링 분석	CTQ 컬럼에 대한 쿼리 분석	컬럼정의서 DBMS 스키마 정보 데이터 값

〈표 13〉 상용 품질진단 도구와의 기능 비교

G업체 품질진단 도구기능	W업체 품질진단 도구기능	본 품질진단 도구기능
테이블스키마 정보 제공	테이블스키마 정보 제공	테이블스키마 정보 제공
컬럼 프로파일링수행	컬럼 프로파일링수행	컬럼 프로파일링수행
관계분석을 통한 참조무결성 오류검증	관계분석을 통한 참조무결성 오류검증	관계분석을 통한 참조무결성 오류검증
코드분석을 통한 오류검증	코드분석을 통한 오류검증	코드분석을 통한 오류검증
품질지표별 업무규칙관리 및 진단수행	품질지표별 업무규칙관리 및 진단수행	품질지표별 업무규칙관리 및 진단수행
-	-	설계 내역과 구현내역 간의 스키마 검증
-	-	데이터 표준 준수여부 검증
-	-	CRUD 상관분석
품질추이 정보제공	품질추이 정보제공	품질추이 정보제공

출처 : DQC, [http://www.dqc.or.kr/guide\\_wise/dq.html](http://www.dqc.or.kr/guide_wise/dq.html).

〈표 12〉에서 제시한 기능과 컨설팅 업체들의 상용 품질진단 도구에서 지원하는 기능을 비교한 내역은 〈표 13〉와 같으며, 본 연구에서 개발할 자동화도구 기능으로 데이터 값뿐 아니라 데이터 구조에 대한 측면을 모두 고려하여 데이터 품질관리 및 진단의 대상을 확장하였음을 확인할 수 있다.

### 3.2 정보분석 및 진단방안

테이블에 저장한 메타데이터 정보와 DBMS 스키마, 데이터 디렉터리에 저장되어 있는 정보를 바탕으로 평가작업을 수행하기 위해서는 우선 데이터베이스 설계내역에서 메타정보를 추출하는 선행 작업과 데이터 분석을 위한 스크립트 작성이 필요하다. 스크립트는 프로그램에 삽입되어 대상정보를 분석하고 오류내역을 추출하는 주요로직에 해당하는 부분이다.

#### 3.2.1 스키마 분석을 위한 진단방안 설명

스키마 생성내역에 대한 진단을 수행하기 위해서는 ERD에서 테이블정보를 추출하는 선행작업이 필요하며, [그림 3]과 같이 ER-WIN의 Report Template Builder를 이용하여 테이블별 컬럼정의, 제약조건, 키정의 내역을 추출해 낼 수 있다.

	B	C	D	E	F	G
1	Column Name	Attribute Name	Column Datatype	Column is PK	Column is FK	Column Null Option
2	CLASS_CD	분류코드	VARCHAR2(30)	Yes	Yes	NOT NULL
3	CODE_NO	코드번호	VARCHAR2(10)	Yes	No	NOT NULL
4	CODE_NM	코드명	VARCHAR2(500)	No	No	NULL
5	ETC_CD1	기타코드1	VARCHAR2(100)	No	No	NULL
6	ETC_CD2	기타코드2	VARCHAR2(100)	No	No	NULL
7	ETC_CD3	기타코드3	VARCHAR2(100)	No	No	NULL
8	SORT_ORDER	정렬순서	VARCHAR2(10)	No	No	NULL
9	GROUP_CODE	그룹코드	VARCHAR2(20)	No	No	NULL
10	LNK_CLASS_CD	링크_분류코드	VARCHAR2(30)	No	Yes	NULL
11	LNK_CODE_NO	링크_코드번호	VARCHAR2(10)	No	Yes	NULL
12	MEMO	비고	VARCHAR2(4000)	No	No	NULL

[그림 3] ERD에서 추출한 테이블 스키마 정보

추출한 테이블 스키마 정보는 데이터 모델링을 통해 생성한 ‘컬럼설계’ 테이블에 저장한다. DB에 자료를 쉽게 저장할 수 있는 방법으로는 추출한 정보를 엑셀파일로 저장한 다음 TOAD의 Database → Import → Import Table Data 메뉴를 이용하여 작

업을 수행할 수도 있다.

	B	C	D	E	F	G
1	컬럼명	코멘트	타입	Primary key	Foreign key	널허용
2	CLASS_CD	분류코드	VARCHAR2(30)	Yes	No	NOT NULL
3	CODE_NO	코드번호	VARCHAR2(10)	Yes	No	NOT NULL
4	CODE_NM	코드명	VARCHAR2(500)	No	No	NULL
5	ETC_CD1	기타코드1	VARCHAR2(100)	No	No	NULL
6	ETC_CD2	기타코드2	VARCHAR2(100)	No	No	NULL
7	ETC_CD3	기타코드3	VARCHAR2(100)	No	No	NULL
8	SORT_ORDER	정렬순서	VARCHAR2(10)	No	No	NULL
9	GROUP_CODE	그룹코드	VARCHAR2(20)	No	No	NULL
10	MEMO	비고	VARCHAR2(4000)	No	No	NULL

[그림 4] DBMS 디렉터리의 스키마 정보

### 3.2.2 데이터정합성 분석을 위한 진단방안

데이터정합성 분석을 위해서는 ERD에서 테이블 간의 관계정보를 추출하는 작업이 선행되어야 하며, 추출방법으로는 ER-WIN의 Report Template Builder를 이용하며 부모 테이블명명과 자식테이블명명, 관계컬럼 내역을 추출한다. DBMS 디렉터리 테이블(USER\_CONSTRAINT S)에서도 관계정보 추출이 가능하나 임의로 참조무결성 제약조건을 구현단계에서 누락하였다면 체크가 불가능하다.

	A	B	C
1	Parent Table Name	Primary Key Column Name	Child Table Name
2	TB_COM_CODE_DETAIL	CLASS_CD	TB_COM_CODE_DETAIL
3	TB_COM_CODE_DETAIL	CODE_NO	TB_COM_CODE_DETAIL
4	TB_COM_CODE_MASTER	CLASS_CD	TB_COM_CODE_DETAIL
5	TB_COM_MENU	MAINMENU_ID	TB_COM_MENU_AUTH
6	TB_COM_MENU	MENU_ID	TB_COM_MENU_AUTH

[그림 5] ERD에서 테이블 간 관계정보 추출

[그림 6]은 ERD에서 추출한 정보를 매크로 수식 이 지정된 엑셀파일에 복사하여 검증쿼리를 작성한 내역이다. 매크로가 적용된 엑셀시트에 부모테이블명과 관계컬럼, 자식테이블명, 관계컬럼을 입력하면 참조무결성을 검증할 수 있는 쿼리문이 자동으로 생성된다. 이와 동일한 기법을 이용하여 자동화도구에서는 데이터베이스에 생성한 ‘참조무결성점검’ 테이블에 ERD에서 추출한 테이블 정보를 저장한 후 점검을 수행한다.

P TABLE	P COLUMN	C TABLE	C COLUMN	R[검증 쿼리문]
ACGOOD_CODE	GOODS_CD, GNGO_K	ACOSSPC_MARK	GOODS_CD, GNGO_K	SELECT 자식컬럼 FROM 자식테이블 MINUS SELECT 부모컬럼 FROM 부모테이블
TB_BIG_INDEX	LOAN_YEAR, FE	TB_BIG_INDEX_HI	LOAN_YEAR, FE	SELECT TB_BIG_INDEX from minus SELECT LOAN_YEAR, RECEIPT_SELECT_CD, SCHOOL_CLASS_CD, BUSI_SELECT_CD, INSP_GROUP_CD, BIG_IDW_CD from TB_BIG_INDEX
TB_BIG_INDEX	LOAN_YEAR, FE	TB_MTD_INDEX	LOAN_YEAR, FE	SELECT TB_BIG_INDEX from minus SELECT LOAN_YEAR, RECEIPT_SELECT_CD, SCHOOL_CLASS_CD, BUSI_SELECT_CD, INSP_GROUP_CD, BIG_IDW_CD from TB_BIG_INDEX

[그림 6] 검증 쿼리문 생성

### 3.2.3 표준 분석을 위한 쿼리문 설명

표준분석 작업은 DBMS에 생성된 스키마에서 컬럼정의 내역 간에 일관성 여부를 점검하고 스키마 정보에 표준용어 또는 표준단어로 정의되어 있지 않은 용어나 단어를 사용하여 정의한 컬럼을 점검한다. 표준 분석을 위해 구현한 SQL은 DBMS 디렉터리의 테이블 스키마 정보를 이용하여 구현하였으며, 동일 컬럼에 대해 일관성 있는 설계가 수행되었는지 점검할 수 있다.

### 3.2.4 데이터 프로파일링 분석용 SQL

컬럼 프로파일링 분석은 <표 14>의 SQL 문을 적용하여 기본점검이 가능하다. 컬럼 속성분석은 대상 컬럼의 총건수·유일값 수·NULL값 수·공백값 수·최대값·최소값 등을 산출하여 그 값이 유효범위 내에 존재하는가 여부를 판단한다. 특히, 컬럼에 대한 패턴분석의 경우 WHERE 절에 점검 대상 컬럼을 특정 문자열로 치환하는 함수인 TR ANSLATE를 이용하여 0~9는 Z, A~Z는 C, a~z는 c, 공백은 B로 치환하여 우편번호나 주민번호 컬럼에 대한 정합성 체크를 수행할 수 있다. 날짜 유형은 사용자 정의함수를 작성하여 날짜유형에 대한 포맷검증을 수행한다.

### 3.2.5 진단결과에 대한 평가기준 정의

데이터 품질 평가를 체계적이고 객관적으로 측정하기 위해서는 평가항목 별 산정기준의 정의가 필요하다. 품질진단 항목은 크게 테이블 구조에 대

<표 14> 컬럼에 대한 프로파일링 분석 SQL

분석구분	실행쿼리
컬럼의 패턴	SELECT [컬럼] FROM [테이블] WHERE TRANSLATE([컬럼], '1234567890ABCDEFHIJKLMNOPQRSTU VWXYZabcdefghijklmnopqrstuvwxyz', 'ZZZZZZZZZZCCCCCCCCCCCCCCCCCCC CCCCCcccccccccccccccccccccccB') <> [분석패턴];
전체레코드수, 최대값, 최소값, 최대길이, 최소길이	SELECT COUNT([컬럼]), MAX([컬럼]), Min([컬럼]), MAX(LENGTH([컬럼]), MIN(LENGTH([컬럼])) FROM [테이블];
유일한 값 개수	SELECT DISTINCT COUNT([컬럼]) FROM [테이블];
NULL 값 개수	SELECT COUNT([컬럼]) FROM [테이블] WHERE [컬럼] IS NULL;
허용범위	SELECT [컬럼], [키컬럼] FROM [테이블] WHERE ([컬럼] < [시작범위]) OR ([컬럼] > [끝범위]);
날자유형	CREATE FUNCTION CHK_DATE (P1 VARCHAR2, P2 VARCHAR2) RETURN VARCHAR2 IS V_TEST _DATE DATE; BEGIN IF P1 IS NULL THEN RETURN NULL; ELSE V_TEST_DATE := TO_DATE(P1, P2, 'NLS_DATE_ LANGUAGE = KOREAN'); RETURN 'T'; END IF; EXCEPTION WHEN OTHERS THEN RETURN 'F'; END;  SELECT [컬럼], [키컬럼] FROM [테이블] WHERE CHK_DATE([컬럼], 'YYYY-MM-DD HH24 : MI : SS') <> 'T'

한 진단과 데이터 값에 대한 진단으로 구분할 수 있으며, 산정방식은 전체 건수에서 확인된 오류 데이터 건수를 백분율로 계산하여 오류율을 계산한다.

<표 15> 데이터 품질평가 기준(산출식) 정의

부문	평가 항목	설명
데이터 구조	표준 용어 미준 수율	<ul style="list-style-type: none"> <li>◦ 각 시스템에서 사용하고 있는 용어가 표준용어에 얼마만큼 포함되는지 측정</li> <li>◦ &lt;산정방식&gt;</li> <li>◦ (미 준수 건수/시스템의 Unique한 용어 수)×100</li> </ul>
	모델 관리 미준 수율	<ul style="list-style-type: none"> <li>◦ 모델링 관리 Tool의 정보와 물리 DB에 생성되어 있는 테이블의 일치 여부를 파악</li> <li>◦ &lt;산정방식&gt;</li> <li>◦ (불일치 건수/시스템의 전체 컬럼 수)×100</li> </ul>
데이터 값	코드 값 오류율	<ul style="list-style-type: none"> <li>◦ 코드값을 사용하는 컬럼들이 표준 코드 허용범위 내의 값을 얼마나 준수하는지 테이블 단위로 측정</li> <li>◦ &lt;산정방식&gt;</li> <li>◦ (불일치 건수/각 테이블의 데이터 건수)×100</li> </ul>
	데이터 값 오류율	<ul style="list-style-type: none"> <li>◦ 검증유형에 따른 오류 검증규칙에 해당 되는 오류 존재 건수 대비 검증</li> <li>◦ &lt;산정방식&gt;</li> <li>◦ 오류존재 테이블 수/전체 테이블 수 ×100</li> <li>◦ 오류레코드 건수/전체 데이터 건수(단일테이블)×100</li> </ul>

### 3.3 자동화도구 구현 내역 설명

#### 3.3.1 표준용어 조회

데이터의 품질을 확보하기 위해서는 데이터의 표준화가 선행적으로 수행되어야 하며, 그 중에서도 용어에 대한 표준화는 이해관계자 간의 의사소통 오류가 발생하지 않도록 각 단위 시스템에서 사용하고 있는 용어들을 일관성 있게 정의한 후 관리해야 한다.

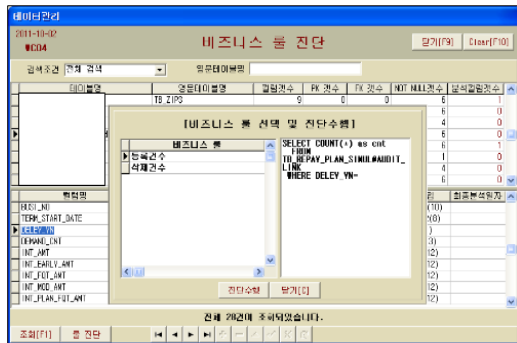
표준용어 조회화면은 프로젝트를 수행하는 사업 기관으로 부터 수집한 표준정보를 DBMS에 저장한 후 데이터베이스 설계시 용어정의에 대한 표준을 준수하였는지 점검할 수 있는 기능을 제공한다.

테이블명	영문컬럼명	한글컬럼명	도메인타입	표준준수여부
TEB_COM_HELP	USE_YN	사용여부	CHAR(1)	Yes
TEB_COM_MENU	ETC	info	CLOB	Yes
TEB_COM_MENU	GROUP_MH	그룹명	VARCHAR2(50)	No
TEB_COM_MENU	MAINMENU_ID	메인메뉴ID	NUMBER(7)	No
TEB_COM_MENU	MENU_ID	메뉴ID	NUMBER(7)	Yes
TEB_COM_MENU	MENU_LEVEL	메뉴리벨	NUMBER(1)	No
TEB_COM_MENU	MENU_NM	메뉴명	VARCHAR2(50)	Yes

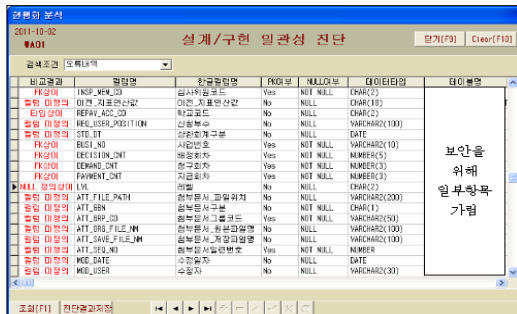
[그림 7] 표준준수여부 진단결과 내역

### 3.3.2 비즈니스 룰 등록관리

비즈니스 룰 등록 및 진단 화면에서는 품질지표의 확장이나 기본 진단내역에 점검항목을 추가하여 진단범위를 확장할 수 있도록 개발하였다. 먼저 '비즈니스 룰 등록관리' 화면에서 추가적인 비즈니스 룰을 등록하고, '비즈니스 룰 진단' 화면에서 추가된 점검항목 선택한 후 사용자가 작성한 질의문 수행을 통해 진단을 수행할 수 있다.



[그림 8] 비즈니스 룰 진단화면



[그림 9] 설계/구현 일관성 진단화면

### 3.3.3 설계/구현의 일관성 진단

설계와 구현내역의 일관성을 진단하는 화면에서

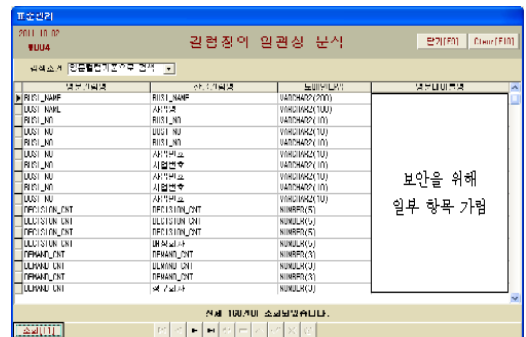
는 설계단계 산출물과 DBMS에 구현된 스키마 정보를 상호 비교하여 산출물 현행화 작업이 미흡한 부분을 찾아내거나 구현 단계에서 누락된 기본키나 외래키 존재여부, 도메인 타입 점검, 컬럼 누락 등의 오류를 사전에 도출하여 운영단계에서 구조적 문제로 인해 발생할 수 있는 장애를 사전에 예방할 수 있다.

<표 16> 설계/구현 일관성 진단항목

진단항목	내용	품질지표
FK상이	◦ FK 컬럼 미매칭 수	완전성
NULL 정의상이	◦ NOT NULL 정의 미매칭 수	
PK상이	◦ PK 컬럼 미매칭 수	
컬럼 미정의	◦ 컬럼 미매칭 수	
타입상이	◦ 컬럼 TYPE 미매칭 수	

### 3.3.4 컬럼정의의 일관성 진단

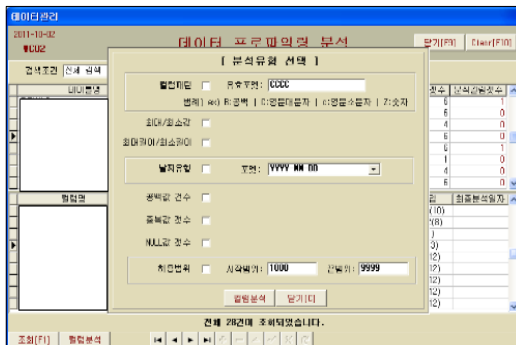
[그림 10]은 동일 영문컬럼에 대하여 정의한 한글컬럼명을 상이하게 정의하거나, 도메인 타입을 잘못 정의한 내역을 점검하기 위한 화면이다. 동일 의미의 컬럼은 데이터 표준을 준수하여 일관성 있게 정의되어야 하며, 정의된 컬럼이 다른 용도로 사용되지 않도록 컬럼명의 의미는 명확해야 한다. 특히, 외래키의 경우 잘못된 실행계획이 수립되지 않도록 하기 위해 데이터타입을 일관성 있게 관리해야 한다.



[그림 10] 컬럼정의 일관성 진단화면

### 3.3.5 데이터 프로파일링 분석

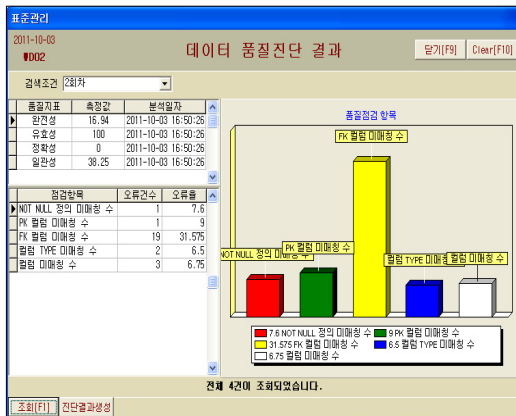
데이터 프로파일링 분석은 컬럼별 데이터 값의 현황을 파악하여 데이터의 품질 수준을 파악하고 진단을 통해 확인된 오류는 각각의 해당되는 품질 지표 오류율에 반영된다. 다음 화면은 데이터 프로파일링 수행화면과 수행결과 발견된 오류를 조회하는 화면이다.



[그림 11] 데이터 프로파일링 분석화면

### 3.3.6 데이터 품질진단 결과

진단결과 화면에서는 앞에서 소개한 기능화면들에서 진단한 결과를 취합하여 종합적인 진단결과를 제시한다. 아래의 예시 화면에서 완전성과 관련하여 진단을 수행한 항목은 총 5개이며, 진단결과 완전성에 대한 품질수준은 16.94%로 측정되었다.



[그림 12] 데이터 품질진단 결과화면(예시)

## 4. 감리적용 검증사례 연구

### 4.1 자동화도구의 실용성 검증

정보시스템 감리를 통해 지적된 내용 중 데이터 품질과 관련된 내용과 보고서에 기술된 증적내역을 제시한 것으로 보고서상의 증적자료는 수감기관의 내부정보로 보안이 필요하기 때문에 실제 기술된 내용에 비해 간략하게만 기술하였다. 주요 감리적 사항으로는 데이터베이스 설계서 기술이 미흡하거나 설계내역과 DB 오브젝트 생성 내역간의 비일관성, 데이터 무결성 오류, 존재하지 않는 코드값 참조 등의 문제점이 지적되었다.

## 5. 결 론

본 연구에서는 공공기관의 정보시스템 구축시법에서 의무적으로 규정한 정보시스템 감리를 통해 정보시스템 구축 단계에서 데이터 품질진단을 원활하게 수행할 수 있도록 지원하기 위한 목적에서 연구를 수행하였으며, 연구방향 설정을 위해 국/내외에서 데이터 품질표준과 관련하여 진행되고 있는 연구 내용을 참고하여 진단방안과 점검항목을 도출 후 데이터 품질진단도구 개발을 위한 설계 및 구현단계를 진행하였다. 개발된 자동화도구는 데이터 프로파일링 기법을 적용하여 데이터 품질 수준을 진단하고 추가적으로 비즈니스 룰을 도출하여 품질핵심정보(CTQ)에 대한 진단이 가능하도록 기능을 구현하였으며, 자동화도구의 진단기능과 정보시스템 감리에서의 지적사항을 비교 검토하여 진단결과에 대한 효과성을 검증하였다. 진단기능 개발과 관련하여 현장에서 활동 중인 감리원들의 의견을 참고하여 산출물 현행화 여부 진단 및 표준용어의 사용의 일관성 점검, 참조무결성 데이터 점검 등 감리현장에서 필요로 하는 진단기능을 추가적으로 반영하였다.

본 연구에서 개발한 자동화도구는 데이터 값에

〈표 17〉 감리보고서 지적사항에 대한 자동화도구 지원가능성 검토내역

번호	구 분	내용
1	지적 사항	<ul style="list-style-type: none"> <li>논리 ERD의 내용 중 엔티티명이 누락되었거나 한글 속성정의가 미흡함.</li> <li>논리 ERD에 한글 속성명이 아닌 영문 약어명을 그대로 기술한 내역이 확인됨.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>자동화도구 메뉴 중 메타표준 준수여부 진단 화면에서 메타사전에 등록된 용어를 사용하여 테이블 컬럼을 명명하였는지 전수 진단이 가능하다. 단, 메타표준을 제시하지 않았거나 엔티티 누락에 대해서는 해당 진단 기능을 사용할 수 없다.</li> </ul>
2	지적 사항	<ul style="list-style-type: none"> <li>ERD의 테이블 정의 내역과 DBMS에 생성되어 있는 컬럼 내역을 비교한 결과 다음 도표와 같이 컬럼 TYPE, PK 설정 여부, 널 옵션 설정여부 등에서 상호 일치하지 않는 부분이 확인됨.</li> <li>DBMS 객체 생성내역을 조사한 결과 기본키(Primary key)가 생성 안 된 테이블이 확인됨.</li> <li>점검결과 테이블정의서에 표기한 기본키(Primary key) 선정항목과 운영DBMS에 생성되어 있는 기본키 선정항목이 상이하였음.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>설계내역과 구현내역(DBMS 오브젝트 생성내역) 간의 비교는 설계/구현 일관성 진단 화면에서 전수 진단이 가능하다. 진단 가능 항목으로는 PK컬럼정의 여부, FK컬럼 정의여부, 컬럼 미정의 여부, NOT NULL 제약조건 정의여부, 데이터 타입의 일치여부에 대한 점검이 가능하며, 설계 내역을 기준으로 진단을 하거나 구현내역을 기준으로 진단을 수행할 수 있다.</li> </ul>
3	지적 사항	<ul style="list-style-type: none"> <li>ERD에 기술된 관계표현을 근거로 테이블간의 정합성 검증을 수행한 결과 참조무결성 오류 데이터가 다수 확인됨.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>ERD에서 추출한 관계정보를 기준으로 참조무결성 분석 화면에서 참조무결성 오류 데이터 분석이 가능하다.</li> </ul>
4	지적 사항	<ul style="list-style-type: none"> <li>테이블정의서와 ERD에 기술된 영문테이블명을 비교 검토한 결과 다음 도표와 같이 ERD의 테이블명이 테이블 명명규칙에 맞지 않게 기술되어 있어 재정의가 필요한 상태임.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>설계내역 간의 일관성 진단은 자동화 도구에서의 진단기능을 제공하지 않고 있으며, 그 이유는 데이터베이스 산출물이 사업기관별로 다른 표준 템플릿을 사용하고 있으며, 데이터베이스 산출물을 자동화도구의 입력데이터로 정형화하는 작업에 많은 공수가 소요될 뿐만 아니라 사전 정형화 작업을 수행하였을 경우 엑셀 상에서 간단하게 비교 검토가 가능하기 때문이다.</li> </ul>
5	지적 사항	<ul style="list-style-type: none"> <li>동일 영문컬럼명에 대해 도메인 정의가 상이한 내역이 확인됨.</li> <li>동일 한글컬럼명에 대해 영문컬럼명 또는 도메인 정의가 상이한 내역이 확인됨.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>컬럼정의 내역의 일관성 진단은 컬럼정의 일관성 분석 화면에서 진단이 가능하다. 다음 그림에서 제시된 것처럼 동일 영문 컬럼에 대해 한글 컬럼명이나 도메인 타입을 상이하게 정의하거나, 동일 한글 컬럼명에 대하여 영문명이나 도메인 타입을 상이하게 정의된 내역에 대해 전수검사를 수행할 수 있다.</li> </ul>
6	지적 사항	<ul style="list-style-type: none"> <li>공통코드 테이블에 존재하지 않는 코드값을 참조하는 레코드의 존재여부 점검결과에서 오류 내역이 확인되었음.</li> </ul>
	도구 지원 내용	<ul style="list-style-type: none"> <li>공통코드 컬럼이나 데이터 값이 유효범위를 갖는 컬럼에 대한 진단은 데이터 프로파일링 분석 화면에서 점검이 가능하다. NULL 값이나 공백값 뿐만 아니라 데이터 값이 시작이나 끝 범위를 갖는 경우 다음 그림과 같이 유효성 검사 항목을 통해 지정된 범위를 벗어난 오류데이터를 점검할 수 있다.</li> </ul>

대한 문제 뿐 아니라 데이터 구조적 측면에서의 문제점까지 진단범위를 확대하였고, 설계 단계에서 작성된 산출물 내역을 진단함으로써 결과물에 대한 지적이 아닌 진행단계에서의 개선을 목적으로 한다는 점에서 차별성을 가지고 있다.

자동화도구를 감리 실무에 적용하여 실용성 검

증을 수행한 결과, 진단대상에 대한 전수 조사를 통해 피 감리기관이 개선해야 할 대상을 구체적으로 제시할 수 있었으며, 정량적으로 품질현황을 제시함으로써 운영단계에서 품질관리 지표의 목표 수립 및 지속적인 점검을 통해 품질지표를 유지할 수 있는 기반을 마련할 수 있었다.



## 참 고 문 헌

- [1] 곽송해, 「감리생산성 향상을 위한 자동화도구 활용 연구」, 2009.
- [2] 김문영, 「체계적인 데이터 품질 관리를 위한 방안을 찾아서」, 디지털콘텐츠, 2005.
- [3] 김수경, 「데이터 품질 매트릭스 설정 연구」, 2001.
- [4] 박주석, 「진정한 데이터 품질 관리의 조건」, 2009.
- [5] 이병태, 양해술, 「산업용 소프트웨어의 평가 기준 및 모듈의 구축」, 2008.
- [6] 이형로, 이유진 등, 「차세대 데이터 품질」, 투이컨설팅, 2009.
- [7] 정혜정, “데이터 품질 평가에 관한 연구”, 『한국 인터넷정보학회』, 제8권, 제4호(2007), p.120.
- [8] 한국데이터베이스진흥원, 「데이터아키텍처 전문가 가이드」, 2010.
- [9] 한국데이터베이스진흥원, 「데이터 품질 관리성숙모형 V1.0」, 한국데이터베이스진흥원, 2006.
- [10] 한국데이터베이스진흥원, 「데이터 품질진단 절차 및 기법 V1.0」, 2009.
- [11] 한국전산원, 「감리 자동화 방안 연구」, 2000.
- [12] Lee, Y. W., 「데이터 품질로의 여행」, 서울경제경영, 2008.
- [13] ISO/IEC, 「25012 Software engineering」, 2006.

## ◆ 저 자 소 개 ◆



**고 재 환 (koj7@kisac.co.kr, cream192@naver.com)**

건국대 정보통신 대학원에서 석사학위를 취득하였으며, 현재는 감리법인 (주)키삭에서 정보시스템 감리업무를 수행하는 수석 컨설턴트이다. 주요 관심분야는 데이터베이스, 개인정보보안, 정보시스템감리, 프로젝트관리, PMO 등이다.



**김 동 수 (dskim@kisac.co.kr, dskim54@gmail.com)**

국민대 BIT 대학원에서 MIS 박사학위를 취득하였으며, 현재 건국대 정보통신 대학원 겸임교수로 재직하고 있다. 정보 시스템 및 통신 전문감리법인인 (주)키삭의 대표 컨설턴트이며, 주요 관심분야는 U-city 감리, 정보시스템 감리, 통신 감리, DB Quality, IT Outsourcing, Enterprise Architecture, Security 등이다.



**한 기 준 (kjhan@db.konkuk.ac.kr)**

서울대학교에서 이학사를 취득하였고, KAIST 전산학과에서 공학석사 및 공학박사 학위를 취득하였으며, 현재 건국대학교 컴퓨터공학부 교수로 재직 중이다. Stanford 대학 전산학과 Visiting Scholar, 한국정보과학회 데이터베이스연구회 운영위원장, 한국ITS학회 기획위원장, 한국공간정보시스템학회 회장, 한국정보시스템감리사협회 회장 등을 역임하였으며, 주요 관심분야는 데이터베이스, GIS, LBS, 텔레매틱스, 정보시스템 감리 등이다.