

CTQ 데이터 선정 모델에 관한 연구

김승희*, 김우제**

Study on the Selection Model CTQ data

Seung-Hee Kim *, Woo-Je Kim **

요약

데이터의 품질은 효율적인 데이터 활용을 위한 가장 기본적인 전제이다. 수많은 연구와 사례를 통해 오류 데이터로 인한 손실과 그로 인한 문제점들이 대두되고 있고, 국가적으로는 데이터 품질 인증제도가 시행되고 있으나 데이터를 생성하고 관리해야 하는 조직 관점의 CTQ 데이터 선정 방법에 대한 연구는 극히 미흡한 상황이다. 본 모델은 조직에서 품질관리 대상이 되는 주요 CTQ 데이터를 선정하여 체계적으로 관리할 수 있도록 업무 및 IT측면의 CTQ 데이터의 기준을 수립하고 그에 따라 데이터를 선별하여 계량화 할 수 있게 있는 전사적 규모의 CTQ 데이터 관리 방법을 구체적으로 제시하였다. 이를 위해 SPSS를 활용하여 요인분석을 수행하고, 계량화를 위해 AHP 방법론을 사용하였다. 특히, DB 품질인증제도의 본격 시행에 따라 실무 적용에 용이하도록 CTQ-DSMM 모델을 활용한 조직 내 데이터 성숙도 관리 방안의 틀도 함께 제시하였다.

▶ Keywords : CTQ, CTQ-DSMM, AHP, 데이터 품질, 품질 평가

Abstract

The quality of the data is the most basic prerequisite for effective use of data. Problems and the resulting loss due to error data has emerged using case studies and a number of, to a national, quality certification system of the data has been enforced, you must manage to generate data study on the method for selecting the point of view of an organization's data CTQ is a very unsatisfactory state of affairs. Selected CTQ main data is subject to quality control in the organization, to develop criteria for CTQ data side of the business and IT so that it can be managed in a systematic manner, the proposed model, to filter the data accordingly presented in

•제1저자 : 김승희 •교신저자 : 김우제

•투고일 : 2013. 2. 12, 심사일 : 2013. 3. 16, 게재확정일 : 2013. 4. 4.

* 서울과학기술대학교 IT정책전문대학원 산업정보시스템전공 박사과정, 한국토지주택공사(Graduate School of Public Policy and Information Technology, Seoul National University of Science and Technology)

** 서울과학기술대학교 기술경영융합대학 글로벌융합산업공학과(School of Global Convergence of Industrial Engineering, Seoul National University of Science and Technology)

detail how to manage enterprise-wide CTQ data that can be quantified Te. By utilizing SPSS, factor analyzes, for which I used the AHP method for quantification. In particular, we present a framework of management measures along the maturity of the data in the organization due to the enforcement of authentic quality certification system of DB, utilizing the CTQ-DSMM model readily applicable to practice.

▶ Keywords : CTQ, CTQ-DSMM, AHP, data quality, quality assessment

I. 서 론

정보화를 통한 업무 수행이 증가하고 DB내에 다양한 데이터가 축적되면서 대내외적으로 방대한 분량의 데이터가 수집되고 있고 데이터의 효과적인 활용이 조직의 경쟁력에 필수적인 요소로 인식되고 있다. 특히, 사업 전략의 수립, 경영의 개량화, 신속한 의사 결정 시 데이터는 가장 중요한 판단의 근거를 제공한다. 그러므로 데이터는 그 품질의 확보가 무엇보다 중요하며, 이는 조직의 경쟁력과 직결되는 요소라 할 수 있다. 그러나 국민의 주민번호가 호적상의 기록과 다른 것이 11만 명이상 발견된 바 있고, 금융기관 데이터 중 주민번호 오류가 400만 건 이상 발견된 바 있으며, 한국데이터베이스진흥원의 '데이터 품질관리의 경제적 효과 분석 연구' 결과에 의하면 한해 평균 데이터 오류는 6,500만 건이며 민간 기업 평균은 430억 건에 달했다. 2009년 공공기관의 공공 정보화 예산 3조2천억 원 중 데이터 품질관리에 들어간 비용은 14.48%인 4,549억 원 가량이었다. 또한 3,500개 민간 기업의 2009년 평균 정보화 비용 중 12.29%인 46조 5,309억 원이 저질 데이터로 인한 피해를 복구하는데 쓰였다고 보도된 바 있다.[16]

그렇다면 조직들이 데이터 품질 확보를 위해 무엇을 관리해야 할 것인가? 모든 데이터를 관리한다는 것은 불가능할 뿐만 아니라 경제적 효율성에도 문제가 있다. 따라서 기업 경영 측면에서 중요하다고 판단되는 핵심 데이터를 선별하여 관리항목과 관리 주체를 정의하여 지속적으로 관리될 수 있도록 하는 방안이 요구된다. 데이터 품질관점에서 핵심(CTQ: Critical to Quality) 데이터란 6sigma에서 유래한 용어로서, 한국데이터베이스진흥원에서는 회사의 고객, 프로세스, 시장 환경, 재무정보 등에 직접적으로 영향을 미치는 중요성이 높은 데이터를 핵심 데이터로 정의하고 있다. CTQ 데이터는 정보시스템상에서 테이블의 컬럼 내 데이터 형태로 존재하게 되지만, 사업측면의 정

보 항목으로써 각 업무별로 선정하되 특정 업무에만 해당되기도 하는 전사적 측면에서 고려되어야 한다. CTQ 데이터 선정 과정은 CTQ 후보 도출 과정과 도출된 CTQ 후보의 우선순위 선정으로 구분할 수 있는데, CTQ 데이터 선정에 대한 연구는 미흡한 상황으로 그 중 CTQ 후보 도출에 대한 연구가 상대적으로 더욱 취약하다. 실제로 6sigma 활동과 관련한 연구보고서에 따르면 VOC(Voice Of Customer)로 부터 CTQ 후보 도출 과정이 6sigma 활동에서 가장 많은 오류를 포함하는 것으로 인식되고 있다[3]. 이러한 배경으로, 정부에서는 2010년 DB 품질 인증 제도(DQC, Database Quality Certification)²⁾를 시행하고 있으며, 한국데이터베이스진흥원을 DB품질인증 기관으로 지정하였다. 한국데이터베이스진흥원의 데이터 관리 인증 (DQC-M : Database Quality Certification-Management)은 성숙수준 5단계로 그 수준을 평가하고 있으나 CTQ 데이터를 생산하고 관리하는 조직의 실정은 CTQ 데이터가 무엇이며, 얼마만큼, 왜 중요한지에 대한 정확한 정의나 관리에 대한 이해조차 미흡한 실정이다. 2012년 DB 그랜드 컨퍼런스에서 한국데이터베이스진흥원 서강수 원장은 "DB 품질이 확보되지 않는다면 아무리 모바일 애플리케이션이 개발한다고 해도 스마트한 비즈니스를 기대할 수 없다"고 언급하며 국내 DB 성숙도가 아직 도입단계에 머물러 있음을 언급하면서, 데이터 품질관리 인증 제도를 본격화 하겠다고 선언한 바 있다

1) DQC(Database Quality Certification: 데이터베이스 품질 인증 제도) : 공공·민간에서 개발하여 활용 중인 정보시스템의 데이터 품질을 확보하기 위한 인증제도임. 데이터베이스를 대상으로 도메인, 업무규칙을 기준으로 데이터 자체에 대한 품질 영향요소 전반을 심의하여 인증하는 DQC-V(Database Quality Certification-Value: 데이터 인증), 행정 및 업무지원, 의사결정 및 정책지원, 지식의 활용 및 제공 등을 목적으로 운영되고 있는 정보시스템에 대한 데이터관리 체계를 심사하여 인증하는 DQC-M(Database Quality Certification-Management: 데이터 관리 인증), 접근 제어, 암호화, 취약점분석, 작업결재 등 데이터베이스 보안에 대한 기술요소 전반을 심사하여 인증하는 DQC-S(Database Quality Certification-Security : 데이터 보안 인증)로 구분된다.[21]

[19].

따라서 DB 품질관리의 기초를 마련하기 위한 각 조직의 노력이 긴급하고 절실한 상황이며, 체계적인 CTQ 데이터 선정 및 관리체계의 도입은 조직 내 수많은 데이터로부터 관리 대상을 정확히 도출하여 단기간에 데이터 품질을 확보할 수 있게 하는데 큰 도움이 될 것이며, 장기적으로 전사적 품질비용을 감소시킬 수 있는 효과적 방안이 될 것이다.

본 연구에서는, 조직의 성공적인 운영을 위해 기업에서 추구하는 최종적인 가치에 대한 공헌도가 높은 데이터를 CTQ 데이터로 정의하고자 하며, 조직관점에서 DB 품질인증 시행에 따라 이들 CTQ 데이터를 어떤 방법으로 선정하고 관리할 것인가에 집중하고자 한다.

본 연구에서 제시하는 CTQ 데이터 선정 모델은 DB 진흥원에서 수행하는 DQC인증에 대비하고, 전사적 측면에서 관리해야 할 데이터를 각 기관에서 정확히 선별하고 객관적, 정량적으로 관리할 수 있도록 하는 모델을 제공하게 될 것이다. 또한 실제 적용 사례를 제시함으로써 모델에 대한 문제점을 검증하고자 하였다. 따라서 제안된 CTQ 선정 모델을 기관 별로 존재하는 내부 통제 프로세스와 결합한다면, 기관 스스로 CTQ 데이터의 품질 수준 관리 및 통제에 있어 능동적이고 합리적인 품질 관리 방안이 될 것으로 사료된다.

II. 선행 연구 고찰

1. 데이터 품질에 관한 연구

1.1 데이터 품질에 관한 정의들

한국데이터베이스진흥원의 DB 품질 가이드라인에 의하면 데이터 품질이란 조직이 목적 달성을 위해 관리되는 데이터가, 조직 구성원, 고객 등 데이터 이용자의 만족을 충족시킬 수 있는 수준이라고 정의하고 있으며, British Columbia 정부에서는 특정 영역에 사용이 적합한 데이터를 만드는 완성도, 유효 기간, 일관성, 시의성 및 정확성의 상태를 데이터 품질이라고 정의하고 있다. 품질 보증 분야에서는 데이터 품질을 데이터의 기능과 특성의 총합, 그 주어진 목적을 충족하는 능력에 기반 하여 데이터와 관련된 요인에 대한 우수성 정도의 합계로 정의하고 있다. Martin Eppler는 정보(Information) 품질이란 정보의 사용에 대한 적합성으로, 데이터 품질이란 비즈니스 요구 사항과 수용 지표에 데이터 값의 적합성을 보장과 관련된 프로세스 및 기술로 정의하였다.

ISO/PAS 26183:2006에서는 제품데이터 품질 가이드라

인의 일종으로, 해당 데이터가 그것을 필요로 하는 모든 사람들에게 제공되는 시기와 결합된 제품 데이터의 정확성 및 적합성의 측정과 같은 데이터 품질을 정의하고 있다.

Larry P. English(20)는 전체 지식 노동자의 기대에 부응하는 것으로 데이터 간의 일치성 유지가 중요하고(Consistently), 모든 데이터는 완전무결할 필요는 없지만 사용자 요구에 부합하는 수준이면 되고(Meeting), 지식노동자와 고객은 데이터를 자신의 업무수행 시 얼마나 도움이 되는지를 근거로 데이터 품질이 만족스러운지 아닌지를 결정하는 정보고객을 의미하며(All knowledge-worker), 데이터 품질의 실제 목표는 사용자와 이해관계자의 만족을 증대시키는 것(Expectations)이라고 기술하였다.

1.2. DB 품질 특성 연구

Ballou & Pazer는 특정 조건 하에서 정보시스템 내의 데이터 항목 오류 및 전파와 변경을 추적할 수 있는 모델을 제시하였는데, 이러한 절차는 광범위하고 효과적으로 특정 문제에 적용이 가능하나 중요한 데이터 흐름/데이터 처리 모델에서 사용이 제한된다.

표 1. 데이터 품질 속성
Table 1. Data quality attributes

사전연구	데이터 품질 기준
Ballou & Pazer(14)	정확성, 완전성, 일치성, 시기적절성
Wang & Strong(15)	신뢰성, 부가가치성, 해석가능성, 접근가능성 등
한국데이터베이스진흥원(2010)(3)(5)	정확성, 일관성, 유용성, 접근성, 적시성, 보안성
미국데이터품질법(2002)(9)	공유성, 변형성, 보안성, 정확성, 시간성, 검색성
CIQM(2)	정보의 범위, 일관성, 통합성, 적시성
MIT TDQM(2)	데이터 량, 일관된 표현, 적시성, 적합성, 객관성, 신뢰성, 평판
이정우(2)	완전성, 적량성, 표현 일관성, 최신성, 갱신주기 정확성, 객관성, 신뢰성, 평판
Larry P. English(2)	completeness, non-duplication, validity, equivalence, concurrency, conformance, accuracy, timeliness

그러나 이 논문은 잠재적인 확장 기능을 평가하기 위한 모델로서 활용할 수 있을 것이다. Wang 과 Strong은 데이터 품질의 특성 연구를 통해 데이터 소비자(데이터를 사용하는 사람)에게 중요한 데이터 품질의 측면을 캡처하는 프레임워크로 118개 데이터 품질을 4가지로 분류하였다. 즉, (1)Intrinsic DQ는 정확성, 객관성, 신뢰성, 그리고 명성으로 구성되어 있다. (2)Contextual DQ는 부가 가치, 관련성, 적시성, 완전성, 데이터의 적절한 금액으로 구성되어 있

다. (3) Representational DQ가 상호운영성, 이해의 용이성, 표현 일치성, 정확한 표현, 일관성, 그리고 간결한 표현, (4) Accessibility DQ는 접근성 및 액세스 보안으로 구성되어 있다. Wang 과 Strong의 연구(15)는 데이터 프레임워크 측정 분석 및 데이터 품질 향상을 위한 근거로 활용 가능하고, 설문조사 등 인식데이터 품질 측정이나 데이터 소비자 인식을 통해 데이터의 품질을 향상시키는 방법을 개발하는 데 도움이 될 수 있을 것이며, 데이터 요구사항 분석 시 체크리스트로 유용하게 활용될 수 있을 것이다. 그러나 경험적 연구 프레임워크의 완전성을 확인하기에 충분하지 않으며, 추가 연구로 특정 작업 컨텍스트에서 이 프레임 워크를 적용해 보는 것이 필요하다.

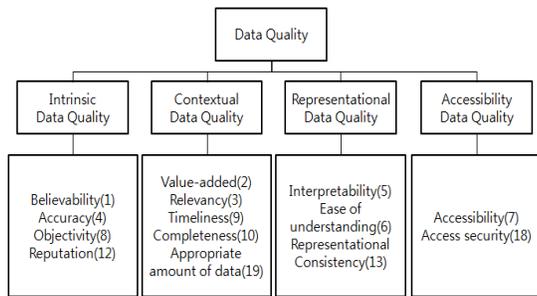


그림 1. 데이터 품질의 개념적 모형(15)
Fig. 1. A conceptual model of data quality (15)

1.3 CTQ 후보 도출 방법에 관한 연구

CTQ 도출방법에 대한 연구는 상대적으로 많이 미흡한 단계인데 김광재의 연구가 유일하다.

Design for Six Sigma(DFSS)는 신제품 개발과 사무 간접 부분의 새로운 프로세스 개발을 위한 6sigma 활동인데 이와 관련하여 초기 고객의 요구사항으로 부터 중요품질 특성을 선정한다. 이와 관련하여 상대적으로 연구가 취약한 CTQ 후보 도출과 관련한 방법론을 연구하였다. 본 논문에서 제안하고 있는 방법은 VOC로 부터 CTQ 후보를 도출하는 과정으로 개발의 대상은 업무 프로세스이지만 VOC에 프로세스, 내/외부 고객, 투입, 산출 정보를 모두 포함한다는 것에 착안하여 사용자(User), 산출물(Output), 입력물(Input), 기능(Function)의 네가지로 시스템 요소를 정의하고 도출 과정을 총 7단계로 제시하였다. CTQ 도출을 위해 주제 유형을 환경, 입력물, 산출물, 제어, 기능, 인터페이스로 구분하였고, 포괄적 속성으로 용이성, 효율성, 정확성, 적절성, 효과성, 역량으로 구분하였다. 그러나 본 논문은 VOC에 기반한 지점표 내에 수많은 항목을 나열식으로 제시하였을 뿐 요인분석이나

상관 분석, 가중치 산정 등 객관화 및 구체적인 계량화 방법이 제시되지 않았다. 또한 주제영역과 속성 영역을 구분하였는데 영역 내 분류 항목의 근거가 구체적이지 않고 객관적인 도출 과정이 제시되지 못하여 실제 지표화 하기위해 응용하기에는 상당히 미흡한 수준이다.

또한 방법론이 신제품 개발이라는 특수한 상황에 고객 요구사항을 설계에 반영하기 위한, 품질관리 항목을 도출하기 위한 방법론으로 본 연구에서 추구하는 CTQ 데이터 항목과 같은 측정 객체를 도출하기위한 방법론과는 적용 목적이 상이하다.

2. 데이터품질 방법론

한국데이터베이스진흥원의 데이터 품질(이하 DQ) 방법론은 데이터품질가이드라인, 데이터 품질관리 가이드라인, 데이터 품질진단 절차 및 기법, 데이터 품질 관리 성숙모형, 품질관리 성과 측정 지침, 데이터 품질 진단을 수행하는 절차와 기법, DB 보안 가이드라인으로 구성된다.

데이터 품질 가이드라인은 데이터 품질진단 방법에 대한 절차와 방법을, 데이터 관리 가이드라인은 요구사항관리, 데이터 표준관리, 데이터 오너쉽 관리, 데이터 구조관리, DB관리, 데이터 흐름관리, 데이터 활용관리, 사용자 뷰관리 절차를, 데이터 품질관리 지침은 데이터 관리, 데이터 구조관리, 데이터 관리 프로세스 등 데이터 품질관리 프레임워크를 정의하고 있다.

데이터 품질진단 절차 및 기법은 품질진단 절차와 정형 및 비정형 데이터 품질 진단 기법을 소개하고 있는데, DB 진흥원에서 데이터품질 인증 심사를 수행하기 위한 데이터 품질진단, 성숙도 수준 측정 방법, 데이터 표준 및 데이터 구조를 관리하거나 DB를 운영하는 것과 관련한 전반적인 개념과 절차로 구성되어 있다. 따라서 DQ 방법론은 각 기관에서 DB 품질인증을 획득하기 위하여 데이터 품질을 어떻게 관리할 것인가에 대한 방법 및 가이드를 제공하기에는 미흡하다. 즉, DQ 방법론은 데이터 품질 진단용 기준 정보 파악을 위해 데이터 모델 정보, 데이터 표준 정보, DB를 분석을 수행하여 도메인 기반 또는 업무규칙 기반의 데이터 품질 진단 시 정보 시스템을 통해 업무규칙이 존재하는 데이터만을 대상으로 하고 있어 업무 규칙과 무관한 조직 관점의 업무 측면의 주요 데이터는 도출에 한계가 존재한다. 진단 DB의 규모가 클 경우 품질 진단 소요 시간 대비 기관의 주요 데이터에 대한 면밀한 검토가 미흡해 질 수 있는 단점이 존재하고, Process 기반 데이터 품질 진단을 수행할 경우 절차를 수행했느냐의 여부만 점검하게 될 우려가 존재하게 되는 것이다. 이를 종합적

으로 판단하면 품질 인증을 위한 절차 수행을 위해 업무 규칙이나 프로세스로부터 CTQ 데이터를 도출하는 것은 조직의 특성, 시간적 한계, 데이터 별 중요도 파악 미흡 등의 문제가 존재하고, 정보시스템으로부터 도출하기 때문에 조직에서 동의하느냐의 문제가 큰 이슈가 된다.

본 연구에서 다루고자 하는 CTQ데이터 선정과 관련한 문제는 데이터 인증(DQC-V)이나 데이터관리 인증(DQC-M)과 관련되므로 DQC-M의 심사 기준인 데이터 관리 성숙수준 평가 항목과 관련하여 좀 더 살펴본다.

한국데이터베이스진흥원의 데이터관리 인증(DQC-M : Database Quality Certification-Management)은 행정 및 업무지원, 의사결정, 정책지원, 지식의 활용 및 제공 등을 목적으로 운영되고 있는 정보시스템에 대한 데이터 관리 프로세스를 심사하여 인증하는 업무를 수행하고 있다. 인증 등급을 평가하기 위한 척도로 성숙도 수준평가제를 도입하고 있으며, 심사 항목은 정확성, 일관성, 유용성, 접근성, 적시성, 보안성으로 나뉜다. 본 연구와 관련되는 심사항목은 정확성이다. 정확성과 일관성에 대한 성숙 수준 별 심사 내용만을 추출하여 정리하면 표 2와 같다. 표 2에서 보는 바와 같이 성숙수준 평가 시 피 인증기관은 데이터의 정확성을 관리하기 위한 관리 대상, 정확성 확인 업무 절차, 데이터 표준, 관리 대상 정의 여부, 정확성 확인을 위한 계획 및 업무 수행 여부, 관리와 관련한 역할과 책임 등을 제시해야 한다.

표 2. DQC-M 정확성 심사 기준
Table 2. DQC-M accuracy of the screening criteria

기준	심사내용
성숙수준1	관리 대상 정의 여부 심사 정확성 확인 업무 수행 여부 심사
성숙수준2	정확성을 위한 표준 정의 여부 심사 관리대상의 정의 여부 심사 정확성 확인을 위한 계획 및 업무 수행 여부 심사
성숙수준3	역할 및 책임 정의 여부 심사 오류 추적 체계 수립 여부 심사
성숙수준4	관리 지표 존재 여부 심사 활동에 대한 성과관리 업무 수행 여부 심사 오류 해결에 대한 지식 축적 여부 심사
성숙수준5	오류의 근본 원인 제거 업무 수행 여부 심사

그러나 대부분의 조직에서는 이러한 업무는 간과되고 있으며 체계화가 부족한 실정으로, 데이터 품질 인증을 대비하여 데이터 품질관리시스템을 도입한 곳에서조차도 조직측면의 CTQ 데이터 관리 방법에 대한 BPR(Business Process Re-engineering) 없이 인증 심사 업체가 구현한 품질관리 심사 항목만을 지표화하여 관리하는 경우가 대부분이다. 즉,

선정 자체가 체계화되지 못하고, 어떤 목적에 의해 수동적으로 폐과지토리로 관리되며, 전사 측면의 객관적 선정 절차는 갖추지 못한 채 지표만 관리하고 있는 상황인 것이다.

이러한 정확성 평가지표와 관련한 문제를 해소하기 위해서는 데이터 관리 방법을 조직의 상황에 맞게 체계화하고 DB 품질인증제에 대비할 수 있도록 하는 구체적 수행 모델이 제시되어 절차가 시스템화 되어야 할 것이다.

본 연구에서는 데이터 정확성 측면에서 조직의 CTQ 대상 데이터를 선정할 수 있는 체계를 제시하고자 한다.

3. 요인분석

요인분석(Factor Analysis)은 많은 변수들의 상호 관련성을 소수의 요인(Factor)으로 추출하여 전체 변수들의 공통 요인을 찾아내 각 변수가 받는 영향의 정도와 그 집단의 특성을 규명하는 통계분석 방법으로 실제 결과를 초래하게 되는 요인을 찾아냄으로써 목표로 하는 명제를 설명하는 다변량 통계분석방법이다(6). 요인 분석은 다수의 변수를 하나의 그룹으로 묶음으로써 변수를 축소하기 위해 사용하거나, 중요도가 낮은 변수 즉, 불필요한 변수들을 제거하기 위해 사용한다. 또한 그룹으로 하나된 변수는 타 요인들과 상호 독립적인 특성을 가지게 되어 결과적으로 측정 항목의 타당성을 평가할 수 있게 된다. 이러한 요인 분석을 통해 얻어지는 요인 점수를 이용하여 회귀분석, 판별분석 및 군집 분석 등에 적용할 수 있다. 요인 분석에서 주요 요인 추출을 위해 주로 주성분 분석을 사용하게 되는데 주성분 개수는 상관 행렬의 고유값이 1 이상이 되는 고유값의 개수, 고유값에 의해 누적설명비율이 일정한 값 이상에 도달할 때의 고유값 개수, 또는 고유값의 상대적 크기가 급격히 작아지는 고유값 변호를 기준으로 결정한다.

본 연구에서는 CTQ 대상 데이터의 의사결정 조건 항목을 도출하기 위한 방법으로 요인 분석을 실시한다. 즉, CTQ 대상 데이터 의사 결정 조건 항목은 두 가지 방법으로 도출하게 되는데, 첫번째는 한국데이터베이스진흥원의 데이터 성숙도 평가 시의 평가항목이며, 또 다른 하나는 실제 업무 프로세스나 현업 부서 활용으로 부터 파생된 평가 항목들이다. 이는 한국데이터베이스진흥원의 평가 기준 항목을 세분화하는 경우 이를 만족할 수 있는 의사 결정 기준을 도출할 수 있기 때문이며, 반면에 정보시스템 내 업무규칙으로 데이터가 도출되는 한계를 교정하기 위한 것이다.

이러한 과정을 통해 잠정 CTQ 대상 데이터 선정 후보가 도출된 후 적용 조직의 특수성을 반영할 수 있도록 최종 CTQ 데이터 선정 절차를 수행하게 된다. 이때 요인 분석을 통하여

유사 항목은 하나로 합하고, 불필요한 항목은 제거되게 될 것이다. 이러한 과정을 거쳐 관리해야 할 DQ 차원의 CTQ 데이터 선정을 위한 최종 중요 의사결정기준을 도출하게 된다.

4. AHP(Analytic Hierarchy Process)

Satty(1971)에 의해 개발된 계층분석기법은 의사결정의 요소들을 발굴하고 각각 2개씩 짝지어 서로 비교 평가(쌍대비교)함으로써 최종의사결정을 이끌어낸다는 방법으로 본 연구에서는 CTQ 데이터 선정을 위한 의사결정 기준에 대하여 가중치를 부여하는데 활용된다. 즉, 데이터의 정확성 확보를 위해 CTQ 데이터를 선정하고자 할 경우 데이터 마다 중요하다고 여기는 이유와 정도가 다를 수 있기 때문에 가중치를 두어 데이터 중요도를 계량화할 수 있는 방법이 필요한 것이다.

이는 AHP가 각각의 요소들을 평가함에 있어 해당분야의 전문가들로 평가자를 구성하고 이들의 지식, 경험 및 직관을 포착하여 수치화하고 객관적으로 비교하여 최상의 의사결정을 도출해 내고자 하는 의사결정 방법론이기 때문이다. AHP를 적용하기 위해서는 각 계층 요소들이 상위계층 요소에 대해 종속적이지만 동일 계층 요소들 사이에는 독립성이 확보되도록 계층적 분해를 하는 것이 필수적이다. 본 연구에서는 핵심데이터 평가 항목의 중요도 산정을 위해 요소간의 중요도를 평가하는 수준을 9개로 나누는 9점 척도를 사용하였다. AHP에서 일관성 지수란 얼마나 사용자가 일관성 있게 설문에 참여하였는지를 판단하는 척도이다. 보통 지수가 0.1이하로 나오면 일관성이 있다고 판단하는데 본 설문은 처음부터 일관성을 유지하기 위하여 설문에 응답자의 일관성 지수(CI:Consistency Index)가 0.1이 넘을 경우 재설문 방법을 취함으로써 0.1 이하 수준이 유지되도록 하였고 결과적으로 34명 전체의 의견이 AHP에 반영되었다.

III. 본 론

1. CTQ 관리 대상 데이터 선정 모델 (CTQ-DSMM) 제안

DQC-M 품질 인증에 대비하고 데이터를 생성하고 관리하는 조직이 초기 도입단계에 있는 CTQ 데이터 관리 시 BPR을 수행하지 않고도 이를 실무에 범용적, 전사적으로 적용할 수 있는 CTQ 대상 데이터 선정 및 관리 모델인 CTQ-DSMM (CTQ-Data Selection Model for Management)을 제안한다.

대부분의 조직에서 CTQ 데이터 관리 체계가 없는 상황을 감안하여 그림 2와 같이 CTQ 데이터 관리 체계를 도입하여 관리할 수 있는 모델과 CTQ 대상 데이터 선정 모델을 함께 제시하였다. 본 모델의 특징은 CTQ 데이터 선정이 하나의 비즈니스 프로세스로 자동화될 수 있도록 관리체계의 도입과 함께 제시되었고, CTQ 대상 데이터 선정 이후 DQC 인증과 관련하여 무엇을 관리해야 하는지에 대하여 피 인증기관 측면의 관리 항목을 제시함으로써 데이터 품질 관리에 즉시 적용할 수 있다는 장점이 있다.

CTQ 데이터 기준을 도출하기 위한 자료 조사 단계에서는 비즈니스 측면과 정보화 측면에서 핵심적이고 중요한 데이터를 선정하여 관리하고자 할 때 가장 우선적으로 고려해야 하는 것이 무엇이라고 생각하는 지에 대하여 정보화 또는 업무 전문가로 부터 의견을 듣거나 CTQ 도출 시 문헌 조사 자료 등을 통해 CTQ 데이터의 기준을 마련한다. CTQ 데이터 기준 항목 척도 조사는 조사된 CTQ 데이터 선별 기준 항목에 대해 어떤 것이 더 중요한지에 대하여 기준 항목 별 점수를 부여하는 설문을 수행하는 단계이다. 이 때 유의할 사항은 제시된 데이터 기준 항목 이외에 설문자가 의견을 추가하여 점수를 부여할 수 있도록 개방형 설문을 추가할 것을 권고한다.

CTQ 데이터 기준 항목 확정 단계에서는 부여된 점수를 활용하여 CTQ 데이터 선정에 영향이 높은 요인과 그렇지 못한 요인을 선별하고, 유사한 성향을 보이는 요인을 그룹핑하여 요인들을 단순화 시킬 수 있도록 요인분석을 수행한다. 이를 통해 CTQ 데이터 기준 항목을 최종 확정한다. CTQ 데이터 기준 항목별 가중치 선정 단계는 AHP 방법을 활용하여 최종 확정된 CTQ 데이터 기준 항목들의 상대적 중요도를 산정한다. 도입 단계 마지막으로 관리해야 할 CTQ 데이터 선정 기준 점수를 어떻게 할 것인가를 정의한다. 점수의 값을 정의하거나 산출 방법 및 조건으로 정의할 수 있다.

이러한 계량화를 위한 기본 틀이 완성되면 다음 과정부터는 메타데이터를 통해 CTQ 데이터 후보를 추출하여 설문을 통해 CTQ 데이터를 최종 선정 확정하는 단계이다. CTQ 데이터 평가는 메타데이터시스템으로 부터 또는 업무 규정집 등을 통해 도출된 CTQ 데이터 matrix를 정보화부서 및 현업에 제공하여 점수를 부여하도록 하는 절차이다.

원천데이터 여부 및 CTQ 레퍼지토리 존재 유무 단계에서는 CTQ 데이터로 도출된 데이터의 DQI(Data Quality Index:데이터 품질 지표) 적용 및 관리 주체가 될 시스템을 식별하는 과정이며, 이렇게 식별된 데이터가 CTQ 레퍼지토리를 통해 이미 관리되고 있는지를 확인하여 추가 작업 여부를 판단한다. 경험적으로 CTQ 레퍼지토리에는 업무명, CTQ

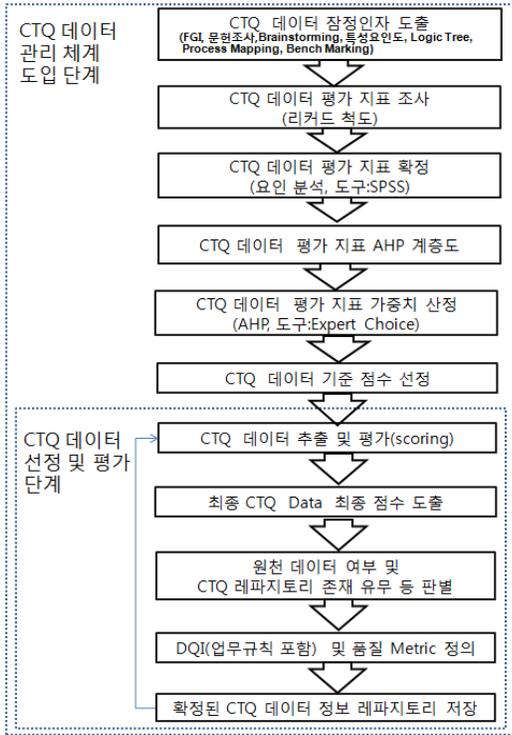


그림 2. CTQ 데이터 선정 모델 제안 개요
Fig. 2. CTQ data selection model proposed outline

데이터명, 도메인 그룹 및 명칭, 데이터타입 및 길이, 업무 규칙, 원천시스템, 활용시스템, 가중치, 품질관리부서, 데이터 품질기준(완전성, 유효성, 유일성, 일관성, 정확성 등)과 세부 품질기준, DQI, DQ 목표 점수 등을 관리한다. DQI 및 품질 Metric 정의 단계는 CTQ 데이터 신규 등록이 필요한 경우나 DQ 품질 기준이 변경된 경우 이를 정량적으로 측정하는 방법과 품질 기준을 정의하는 단계로, 반드시 수치화, 계량화 되도록 해야 한다. 확정된 CTQ 데이터 레파지토리 저장 단계에서는 최종 CTQ 대상 데이터를 확정하고 확정된 정보를 전사적으로 관리할 수 있도록 관련 시스템이나 화일 등에 저장하여 영구화 시키는 것이다.

본 모델을 활용할 경우 각 조직이나 기관 스스로 CTQ 대상 데이터를 선별하여 관리할 수 있을 뿐만 아니라, CTQ 데이터 간에도 중요도의 정도를 판별할 수 있어 효율적 CTQ 데이터 관리에 유용할 것이다.

2. CTQ 데이터 선정 모델 적용

본 모델을 적용하기 위하여 연구자가 속한 회사의 특정 업무를 선정하였고, CTQ 데이터 의사결정 조건을 도출하기 위하여 본 연구자가 소속된 조직의 정보화 부서 및 현업 직원,

유지보수 담당자 등에 인터뷰 및 자료 조사를 실시하였다.

2.1 CTQ 대상 데이터 기준항목 잠정 인자 도출

CTQ 대상 데이터 의사결정 기준을 도출하기 위하여 한국 데이터베이스진흥원에서 피 인증기관의 데이터성숙도형 진단 시 데이터 규칙 및 기준 정보 중요도 항목 후보군으로 제시된 자료와 이 부분이 정보시스템으로 부터 도출된 자료만을 대상으로 하기 때문에 발생하는 적용 조직의 업무 특성을 반영할 수 없다는 한계를 교정하기 위하여 잠정 CTQ 대상 데이터 의사결정 기준 설문조사를 실시하여 이를 통합한 CTQ 대상 데이터 의사결정 요인을 도출하였다. 도출된 결과와 한국데이터베이스진흥원의 데이터 성숙도 측정 시 활용하는 측정 중요도 선정 지표와 매핑한 결과는 그림 3과 같다.

데이터성숙도 측정 중요도 선정 지표		기업 내 CTQ 대상 데이터 의사결정 기준 도출결과	
시스템 중요도	① 데이터 정보시스템에서 차지하는 데이터 중요도	① 기업의 사업영역내 출생 축면에서 주요 핵심 프로세스 상에 놓여있는 업무 인자의 여부	현업/정보화/업무 담당자 대상 CTQ 후보 선정 의견
비즈니스 중요도	② 기업 시스템에서 차지하는 활용 중요도 ③ 시스템 DW, DM 등 시스템간 데이터 불일치에 대한 발생 영향	② 데이터 오류 및 데이터 유출 등 각종 장애 발생 가능성 있는 정도 또는 해결 과정에서 기업 및 재무 측면의 영향 정도 ③ 기업의 경영 관리 측면에서 사업장 분석 및 재무적 여건에 직접적으로 영향을 미치는 정도	정보화/업무 담당자 대상 CTQ 후보 선정 의견
업무 관점	④ 데이터 오류로 인해 장애를 유발할 수 있는 정도 또는 불량을 야기할 수 있는 정도 ⑤ 업무 영향도 ⑥ 2개 이상의 업무 간 영향도	④ 대외적으로 기업의 신뢰도에 영향을 미치는 데이터 인자의 여부 ⑤ 현업에서 실제 활용되는 정도 또는 타 부서, 타 기관에서 활용되는 정도	정보화/업무 담당자 대상 CTQ 후보 선정 의견
IT 관점	⑦ 서비스 서비스에 영향을 미치는 정도 ⑧ 재무 ⑨ 업무 ⑩ 시스템 체계적, 정보계, DM 시스템 간 데이터 불일치 시 발생 영향 정도 ⑪ 데이터 고객(내/외부)의 사용빈도	⑦ 정보시스템 상에서 정보 기술상 해당 시스템 전반에 영향을 미치거나 업무 코드 및 master 성격의 데이터 여부 ⑧ 데이터 조작 시 데이터 수송 조건에 반비례 사용되거나 유일한 데이터를 도출하는 결과 결정적으로 사용되는지 여부 ⑨ 정보시스템 간 내/외부 연계 및 연계된 데이터가 일치하는 지의 여부 ⑩ 데이터 안전성 등 보안 기술이 별도 적용된 데이터 인자의 여부 ⑪ 복잡한 업무 산책이 적용되어 있어 강의 오류 발생 확률이 높은 지의 여부 ⑫ 현재 보관된 데이터가 기업 내 핵심 데이터인가의 여부	정보화/업무 담당자 대상 CTQ 후보 선정 의견

그림 3. CTQ 대상 데이터 후보 의사결정 기준 요인 매핑
Fig. 3. CTQ mapping target data candidate factors the decision-making criteria

그림 3의 매핑과정을 통해 도출된 최종 CTQ 대상 데이터 의사결정 항목을 도출하고 항목별 의미를 정의하였다. 데이터 성숙도 모형에서는 선정 항목이 네 가지로 구분되었으나 사내 잠정인자 도출 시 기업 내 인터뷰 그룹 구분(현업, 정보화 부서)과의 매핑을 위해 시스템 중요도와 IT 관점을 합하여 IT 관점으로, 비즈니스 중요도와 업무 관점을 합하여 업무 관점으로 구분하였다. 예외적으로 현업 활용도 매핑 시 데이터 성숙도 측정 중요도 선정 지표 중 하나인 IT관점의 고객(내/외부)의 데이터 사용 빈도 지표는 업무적 사용 빈도로도 이해할 수 있으므로 업무 관점의 지표인 업무중요도와 외부영향도는 현업활용도로 그룹핑 하였다.

표 3. 업무 부문 지표 도출 및 의미 정의
Table 3. Indices business sector and define the meaning

지표		1*	2**	정의
업무 부문	1 핵심업무 민감도	⑨ ⑥	①	사업과 직결된 서비스 및 업무적 중요도가 높은 핵심 업무 프로세스 상의 정보
	2 데이터 장애 심각도	④ ⑦	②	데이터 장애(오류 및 유출, 불법가공)시 장애 정도 및 장애로 인한 재무적 손실에 크게 영향을 미치는 정보
	3 재무 영향도	⑦	② ③	기업의 경영 관리 측면에서 사업성 분석 및 의사결정, 마케팅에 중요한 영향을 미치는 정보
	4 대외 신뢰도	⑧	④	대외적으로 기업의 신뢰도에 영향을 미치는 정보 인지의 여부
	5 현업 활용도	⑤ ⑧ ⑪	⑤	현업에서 실제 활용되는 정도 또는 타 기관, 타 부서, 타 업무에서 활용되는 정도

1*: 데이터 성숙도 측정 중요도 선정 지표
2**: 기업 내 CTQ 대상 의사결정 기준 도출 결과

표 4. IT 부문 지표 도출 및 의미
Table 4. Indices IT sector and define the meaning

지표		1*	2**	설문 문항
IT 부문	1 테이블 영향도	①	⑥	정보시스템 상에서 정보 기술상 해당 시스템 전반에 영향을 미치는 데이터이거나 코드 및 마스터성 데이터
	2 데이터 선별 기여도	② ⑪	⑥ ⑦	데이터 조작 시 데이터 추출 조건에 빈번히 사용되거나 유일한 데이터를 도출하는데 결정적으로 사용되는지 여부
	3 시스템 연관도	③ ⑩	⑧	정보시스템 간 내 외부 연계 및 연계된 데이터가 일치하는 지의 여부
	4 데이터 보안도	-	⑨	규정상 데이터 보안 기술(암호화, Audit)이 적용되어야 하는 데이터
	5 업무규칙 복잡도 (난이도)	-	⑩	업무 규칙 난이도와 복잡도가 높은 데이터로 산출값의 정확성 여부를 주기적으로 검증해야 하는 데이터
	6 데이터 오너쉽	-	⑪	Ownership을 가진 원시데이터로 데이터 입력 오류에 대한 지속적 관리가 필요한 데이터

1*: 데이터 성숙도 측정 중요도 선정 지표
2**: 기업 내 CTQ 대상 의사결정 기준 도출 결과

표 3과 표 4와 같은 매핑 작업 결과 CTQ 대상 데이터의 의사결정을 위한 조건이 모두 도출되었으며, 데이터 성숙도 중요도 선정 지표에는 포함되어 있지 않았던 데이터 보안도, 업무규칙 복잡도(난이도), 데이터오너쉽 등 CTQ 대상 데이터 의사결정 기준 지표가 추가되었다.

한국데이터베이스진흥원에서는 기관에서 무엇을 선정하여 무엇을 측정할 것인가에 대한 사전 측정 지표 관리에 대해 가

이드 되는 바가 없으므로 위와 같은 방법을 활용하며 기업의 특성을 고려한 기초 지표 선정이 가능할 것이다.

2.2 CTQ 대상 데이터 기준 항목 선정을 위한 척도 조사
2.2.1 모집단 선정 및 표집

본 설문은 2011년 11월 본 연구자가 근무하고 있는 회사의 직원과 유지보수 담당 직원을 모집단으로 선정하여 50명을 대상으로 실시하였다. 설문지는 5점 척도로 작성되었고, CTQ 데이터 선별을 위한 기준 항목에 대하여 각각의 중요도를 5점(중요:5점, 중요하지 않음:1점) 척도로 구성하였다.

50명의 설문자중에 34명으로 부터 설문지가 회수되었고 회수된 설문 결과 기초 통계 결과는 표 5와 표 6에 나타내었다.

표 5. 모집단의 일반적 특성
Table 5. The general characteristics of the population

구분	항목	명	비율
성별	여자	10	29.41%
	남자	24	70.59%
연령	20대	2	5.88%
	30대	20	58.82%
	40대	12	35.29%
Total		34	100.0%

표 6. CTQ 대상 데이터 선정 5점 척도 평균과 표준편차
Table 6. CTQ target data selected a 5-point scale the mean and standard deviation

		설문 문항	평균	표준편차
업무 부문	1	핵심 업무 민감도	4.824	0.387
	2	데이터 장애 심각도	4.618	0.697
	3	재무 영향도	1.853	1.54
	4	대외 신뢰도	1.941	1.099
	5	현업 활용도	4.206	0.978
IT 부문	1	테이블 영향도	4.912	0.288
	2	데이터 선별 기여도	4.206	0.808
	3	시스템 연관도	4.471	0.961
	4	데이터 보안도	1.794	1.553
	5	업무 규칙 복잡도	1.824	1.267
	6	데이터 오너쉽	1.559	1.211

설문 결과에서 평균들이 낮게 나온 문항은 담당자들이 CTQ 대상 데이터를 선정 시 의사결정 기준으로서 중요하다고 생각하지 않음을 뜻하며, 업무 부문의 재무 영향도, 대외 신뢰도와 IT부문의 데이터 보안도, 업무 규칙 복잡도, 데이터 오너쉽 등이 여기에 해당 된다.

반면에 평균이 4점 이상으로 높게 나타난 문항은 업무 부문에서 핵심 업무 민감도, 데이터 장애 심각도, 현업 활용도와 IT부문에서 테이블 영향도, 데이터 선별 기여도, 시스템 연관도 등이 해당된다.

2.3 CTQ 대상 데이터 의사결정기준 평가 지표 도출 및 계층도

2.3.1 요인 분석

CTQ 대상 데이터 의사 결정 시 유사한 경향을 보이는 항목 간 그룹핑하고 무의미한 항목을 제거하여 CTQ 대상 데이터 추출을 명확히 할 수 있도록 하기 위하여 그림 4와 같이 SPSS를 활용하여 주성분 요인 분석을 실시하였다.

공통성		설명된 총분산						
추출	성분	후출 제공할 적재값			회전 제공할 적재값			
		개체	% 분산	% 누적	개체	% 분산	% 누적	
핵심업무민감도	.920	1	1,699	33,982	33,982	1,670	33,408	33,408
데이터장애심각도	.587	2	1,155	23,105	57,087	1,040	20,799	54,207
재무영향도	1.000	3	.857	17,141	74,228	1,001	20,021	74,228
대외신뢰도	.485							
현업활용도	.719							

추출 방법: 주성분 분석, 회전 방법: 주성분 분석

공통성		설명된 총분산						
추출	성분	후출 제공할 적재값			회전 제공할 적재값			
		개체	% 분산	% 누적	개체	% 분산	% 누적	
데이터불영향도	.696	1	1,940	32,460	32,460	1,736	28,336	28,336
데이터선별기여도	.819	2	1,291	21,514	53,973	1,325	22,063	51,019
시스템연관도	.634	3	1,085	17,758	71,731	1,243	20,713	71,731
데이터보안도	.865							
업무규칙복잡도	.715							
데이터오너쉽	.585							

추출 방법: 주성분 분석, 회전 방법: Kaiser-Meyer-Olkin이 있는 배리역스

그림 4. 업무부문(상)과 IT부문(하) 요인분석결과
Fig. 4. Business sector (Up) and the IT sector (Down) the results of factor analysis

업무 관점에서는 3개의 요인으로 강제로 분류한 결과 74%의 설명력을, IT 관점에서는 3개의 요인이 71%의 설명력을 보이는 것으로 나타났다.

scree plot 도표를 통해 1보다 큰 값을 가지는 요인이 업무관점으로 2개 요인, IT관점으로 3개 요인이 도출됨을 알 수 있었다. 한편, 명확한 그룹핑을 위하여 업무관점 요인 수를 3으로 고정시킨 후, 부하량이 비슷해지는 것을 막기 위해 varimax 회전을 사용하여 그림 6과 같이 요인 분석을 수행하였다.

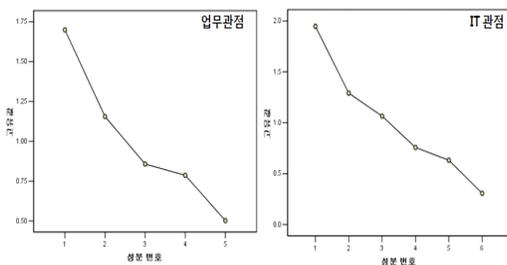


그림 5. scree plot 그래프
Fig. 5. scree plot graph

회전된 성분행렬에 의하여, 업무관점에서는 첫 번째 데이터 장애 심각도 그룹, 두 번째 핵심 업무 민감도 그룹, 세 번째 재무영향도 그룹이 추출되었고, IT관점에서는 첫 번째 데이터 보안도 그룹, 두 번째 데이터 선별 기여도 그룹, 세 번째 테이블영향도 그룹을 도출할 수 있었다.

또한 요인 분석 결과, 데이터 보안도 그룹, 테이블영향도 및 시스템 연관도 그룹으로 묶일 수 있으므로 데이터보안 및 업무 규칙 복잡도, 데이터 선별 기여도 및 연관도, 테이블 영향도로 그룹핑하고, 데이터 오너쉽은 제외한다. 이러한 측정 항목의 신뢰성을 다시 한번 확인하고자 신뢰도 분석을 실시하였다.

회전된 성분행렬 ^a	성분		
	1	2	3
핵심업무민감도	.014	.956	-.075
데이터장애심각도	.766	.027	.015
재무영향도	.050	-.076	.996
대외신뢰도	.620	-.314	.048
현업활용도	-.835	-.145	-.036

요인추출 방법: 주성분 분석, 회전 방법: Kaiser-Meyer-Olkin이 있는 배리역스, a. 4 번째 계산에서 요인회전이 수렴되었습니다.

회전된 성분행렬 ^a	성분		
	1	2	3
데이터불영향도	.211	-.065	.865
데이터선별기여도	-.017	.898	-.111
시스템연관도	-.144	.596	.508
데이터보안도	.912	.127	.079
업무규칙복잡도	.768	-.374	.028
데이터오너쉽	.514	-.056	-.564

요인추출 방법: 주성분 분석, 회전 방법: Kaiser-Meyer-Olkin이 있는 배리역스, a. 5 번째 계산에서 요인회전이 수렴되었습니다.

그림 6. Varimax 회전에 의해 생성된 성분행렬
Fig. 6. Composition matrix generated by the Varimax rotation

그 결과 신뢰도 측정결과 업무 부문은 0.626, IT부문은 0.601로 나타났다. Cronbach's Alpha값은 1에 가까울수록 신뢰도가 높다고 판단하며, 일반적으로는 0.6보다 크면 해당되는 몇 개의 항목들이 신뢰도가 높다고 판단한다.

따라서, CTQ 데이터 선정 기준 요인의 신뢰도는 유의미한 수준으로 판단 가능하다. 그림 7은 이러한 과정에 의해 도출된 최종 CTQ 대상 데이터 의사결정 지표이다.

	문항	Corrected Item-Total Correlation	Cronbach's α
업무관점	업무규칙복잡도	.691	0.626
	데이터보안도	.266	
	재무영향도	.292	
	대외신뢰도	.691	
	데이터오너쉽	.162	
IT 관점	데이터장애심각도	.627	0.601
	시스템연관도	.515	
	현업활용도	.353	
	테이블영향도	.245	
	데이터선별기여도	.229	
	핵심업무민감도	.127	

그림 7. CTQ 대상 데이터 선정 기준 요인 신뢰도 분석
Fig. 7. CTQ target data selection based on factor reliability analysis

2.3.2 CTQ 대상 데이터 선정 평가 지표 세분화

CTQ 대상 데이터 선정 지표를 정의에 맞게 구체적으로 명확하게 하기 위하여 그림 8과 같이 지표를 세분화한다.

CTQ 대상 데이터 지표를 정의에 맞게 구체적으로 명확하게 데이터를 구분할 수 있도록 한국데이터베이스진흥원에서 성숙도 평가 시 활용하는 중요도 점수 부여 기준을 참조하여 평가 지표를 세분화 하였다.

평가지표	표식	세부 분류
업무관점	데이터 장애 심각도	기업의 경영관리 측면에서 사업성 분석 및 의사결정, 마케팅에 중요한 영향을 미치는 정도
	핵심업무 민감도	사업과 직결된 서비스 및 업무적 중요도가 높은 핵심 비즈니스 프로세스 정도
	재무 영향도	데이터 장애(오류 및 유출, 불법가공) 시 장애 정도 및 장애로 인한 재무적 손실에 크게 영향을 미치는 정도
IT 관점	데이터 영향도	정보시스템 상에서 정보 기술상 해당 시스템 전반에 영향을 미치거나 업무 코드 및 master성 데이터
	데이터 선별 기여도	데이터 조작 시 데이터 추출 조건에 빈번히 사용되거나 유일한 데이터를 도출하는데 결정적으로 사용되는지 여부
	데이터 보안도	규정상 데이터 보안 기술(암호화, Audit)이 적용되어야 하는 데이터

그림 8. CTQ 대상 데이터에 대한 선정 평가 지표 도출 결과
Fig. 8. Evaluation Indices for selected target data CTQ results

표 7. CTQ 대상 데이터 선정 지표
Table 7. CTQ target data selected indicators

평가지표	표식	세부 분류
데이터 장애 심각도	지표1-A	데이터 오류 및 유출 시 사업 전체에 미치는 장애로 인한 영향의 심각성
	지표1-B	데이터 오류 및 유출 시 유출된 데이터로 인해 고객 불만 제기의 심각성
	지표1-N	특이 사항 없음
핵심 업무 민감도	지표2-A	사업과 직결된 서비스 업무 핵심 Biz. 프로세스 상의 정보
	지표2-B	사업과 직결된 서비스를 지원하는 핵심 Biz. 프로세스 지원 정보
	지표2-N	특이 사항 없음
재무 영향도	지표3-A	데이터 오류 및 유출 등 사고발생 시 영업이익 및 매출에 대한 영향도
	지표3-B	데이터 오류 및 유출 시 해결을 위한 재무 측면의 영향도
	지표3-N	특이 사항 없음

평가지표	표식	세부 분류
데이터 영향도	지표4-A	정보시스템 상에서 설계상 시스템 전반에서 영향을 미치는 데이터 인지의 여부
	지표4-B	단순 마스터성 데이터나 코드성 데이터
	지표4-N	특이 사항 없음
데이터 선별 기여도	지표5-A	유일한 값을 구별하는데 사용되는 데이터에 대한 중요도 (주키)
	지표5-B	데이터 추출 조건에 빈번히 사용되는 컬럼의 중요도(주키 외 인덱스)
	지표5-N	특이 사항 없음
데이터 보안도	지표6-A	규정상 데이터 암호화 구현이 필요한 데이터(개인정보 등)
	지표6-B	데이터 가공 및 변경에 대한 증거가 요구되는 데이터(증적 및 Audit 대상)
	지표6-N	특이 사항 없음

2.4 CTQ 대상 데이터 선정 지표 중요도 산정

2.4.1. CTQ 대상 데이터 선정 지표 AHP 계층도

최종 CTQ 대상 데이터 선정 평가 지표의 중요도를 평가하기 위하여 표 7에서 도출된 CTQ 대상 데이터 선정 지표를 기준으로 그림 9와 같이 AHP 계층도를 완성한다.



그림 9. CTQ 데이터 선정 평가 지표 계층도
Fig. 9. CTQ data selected the evaluation index hierarchy

CTQ 데이터 평가 기준의 중요도를 결정하기 위하여 업무 및 IT 관점별, 업무 관점 별, IT 관점 별 상대비교를 수행하며, 도출된 가중치를 반영한 가중 중요도를 산출하며, 중요도는 표8과 같이 9점 척도로 구성하였다.

표 8. 데이터 중요도 기본 척도
Table 8. the primary measure of the Data importance

중요도	정의	설명
1	동등하게 중요(equal)	어떤 기준에 대하여 두 인자가 비슷한 영향을 가진다고 판단됨
3	약하게중요 (moderate)	경험과 판단에 의하여 한 인자가 다른 인자보다 약간 더 영향을 미침
5	강하게 중요 (strong)	경험과 판단에 의하여 한 인자가 다른 인자보다 확실하게 영향을 미침
7	매우 강하게 중요(very strong)	경험과 판단에 의하여 한 인자가 다른 인자보다 강하게 영향을 미침
9	절대적으로 중요 (extreme)	경험과 판단에 근거한 인자가 다른 인자보다 극히 강하게 영향을 미침
2,4, 6,8	위 값들의 중간 값	경험과 판단에 의한 비교값이 위 값들의 중간에 해당될 경우 사용함
역수값	a 요소가 b요소 보다 중요한 경우 b의 중요도는 a의 중요도와 비교하여 그 역수 값을 갖는다	

2.4.2 모집단 선정 및 설문

AHP 중요도 산출을 위한 설문은 연구자가 근무하는 회사 내 요인 설문 조사 시의 참가자(34명)를 대상으로 실시하였다. 설문 결과 응답자의 논리적 모순으로 인해 일관성 지수(CI: Consistency Index)가 0.1보다 큰 경우가 발생되었다. 설문 회수율이 높지 않은 관계로 신뢰도 향상을 최대로 높이기 위해 응답자에게 재설문을 의뢰하는 방식을 취하였다. 계층적으로 분석된 설문 결과를 Expert Choice 11.5 S/W를 활용하여 입력 및 분석 작업을 수행하였다. 최초 설문지를 배포하여 회수한

설문지 중 총 7개의 설문에서 비일관성 지수가 0.1이 초과되었다. 아무 설명 없이 재 설문을 요청한 결과 1개만이 일관성 지수를 만족했다. 재설문의 사유를 상세히 설명한 후 재설문한 결과 재 설문대상이 2개로 축소되었고 마지막 하나의 재 설문을 수행한 결과 전체 일관성 지수가 0.1 이하로 만족되었다.

2.4.3. 가중치 산정

중요도 산정 결과 그림 10에서와 같이 IT관점 보다 업무 관점에서 중요도가 높게 나타났고, 업무 관점은 데이터 장에 심각도, 핵심업무 민감도, 재무영향도, IT 관점은 테이블 영향도, 데이터 선별 기여도, 데이터 보안도 순으로 중요도가 높게 산정되었다.

2.5 CTQ 대상 데이터 선정 기준 점수 산정

CTQ 데이터 Scoring 전에 CTQ 데이터 선정 기준 점수를 선정한다. 즉, CTQ 데이터 후보를 설문을 통해 점수화 한 후 어느 데이터를 CTQ 대상 데이터로 확정할 것인가에 대한 점수의 하한을 정하는 것이다.



그림 10. CTQ 대상 데이터 선정 평가 기준 별 가중치
Fig. 10. CTQ target data selection weighted ratings

기준 점수는 전사 표준 점수로 할 수도 있고, 해당 업무마다 고유성을 인정하여 업무별로 기준점수를 선정할 수도 있다. 이 경우 기준 점수 산정 방법이 명시되어야 한다. 정보 시스템 수가 많고 전사적 데이터 품질관리를 수행하는 조직인 경우 업무별 CTQ 대상 데이터 기준 점수를 적용하여 객관성을 확보하기 위하여 동일한 기준 점수를 적용하는 것이 바람직하다.

2.6. CTQ 대상 후보 데이터 추출 및 평가

2.6.1 업무 별 CTQ 대상 데이터 후보 도출

현업 인터뷰와 설문, 규정집, DB, 메타데이터, 각종 업무용 로그 화일로 부터 다양한 방법을 통해 CTQ 데이터 후보를 도출한다. 예를 들면, 주요 마스터 테이블과 코드성 테이블은 표준화 시스템인 메타데이터 시스템을 통해 도출할 수 있다. 만일 이러한 정보시스템이 없는 경우 DB와 응용프로그램 등으로부터 실제 운영되고 있는 소스를 추출하여 피벗 테이블을 통해 그림 11과 같이 도출할 수 있다. 데이터 표준화가 잘 준수되고 있는 경우 연계 테이블 명명 규칙에 근거하여 연계 데이터 정보를 쉽게 알 수 있다.

그러나 ESB(Enterprise Service Bus)에 기본을 둔 EAI(Enterprise Application Integration)와 같은 Web 서비스의 별도 연계 솔루션을 통해 구현된 경우 기본 정보시스템 외에 EAI DB로 부터 추가적인 데이터 도출이 필요할 수도 있다.

DB Link 연계와 같이 단순한 데이터 연계만 있는 경우나 별도의 DB 연계용 S/W를 사용하는 경우 테이블 별로 부여

데이터			업무 부문 중요도						업무 부문 CTQ Score	
업무구분	업무단위 (레이블명)	주요데이터 (컬럼명)	업무관점 0.521							
			데이터장에 실각도		핵심 업무 민감도		재무 영향도			
			0.444		0.34		0.216			
			지표1-A	지표1-B	지표2-A	지표2-B	지표3-A	지표3-B		
상	사	년	ID	5	3	4	3	4	5	2572
상	사	년	번호	1	2	2	3	4	5	1569
상	사	년	번호	2	5	3	4	5	4	2425
상	사	년	번호	4	2	3	5	3	3	2259
상	사	년	번호	3	1	2	5	2	3	1780
상	사	년	번호	1	3	2	2	2	4	1429
상	사	년	번호	5	1	1	2	1	5	1590

그림 13. CTQ 후보 대상 데이터 평가(업무 부문)
Fig. 13. CTQ candidate assessment data (business sector)

데이터			IT 부문 중요도						IT 부문 CTQ Score	
업무구분	업무단위 (레이블명)	주요데이터 (컬럼명)	IT 관점 0.479							
			데이터 영향도		데이터 선택 기여도		데이터 보안도			
			0.412		0.346		0.243			
			지표4-A	지표4-B	지표5-A	지표5-B	지표6-A	지표6-B		
상	사	년	ID	5	5	4	5	4	5	3046
상	사	년	번호	5	5	2	5	3	4	2695
상	사	년	번호	5	5	3	4	5	4	2814
상	사	년	번호	4	2	2	5	2	3	1989
상	사	년	번호	3	2	5	5	2	3	2189
상	사	년	번호	1	3	2	2	2	4	1408
상	사	년	번호	5	1	4	2	1	5	1933

그림 14. CTQ 후보 대상 데이터 평가(IT 부문)
Fig. 14. CTQ candidate assessment data (IT sector)

그림 13과 그림 14와 같이 업무 부문의 CTQ 데이터 평가에서 사업지구 번호, 소재지, 본번 순으로 중요도가 높게 산출되었고, IT 부문의 CTQ 데이터 평가에서는 사업지구 번호, 소재지, 지번관리번호 순으로 중요도가 높게 산출되었다.

그림 13과 그림 14는 내용의 이해를 돕기 위해 실제 평가 사례에서 일부만을 발췌한 것으로 최종 중요도 점수가 업무 부문 CTQ점수와 IT부문 CTQ 점수를 합하여 데이터 선정 기준 점수(사례에서는 3.7 이상)를 만족하는 항목만 선별되며, 본 사례에서는 총 5개의 최종 CTQ 데이터가 선정된다.

2.7 CTQ 데이터 레파지토리 관리

2.7.1 CTQ 데이터와 관련된 부가 정보들

CTQ 데이터 선정이 완료되면 이를 위해 CTQ 데이터와 관련한 추가 정보의 파악이 필요하다.

먼저 도출된 CTQ 데이터가 최초로 입력되는 시스템인 원천 시스템(업무)을 평가하여 운영 중인 레파지토리 내 해당 데이터가 존재하는 지를 확인한다. 만약 존재하는 경우 레파지토리 정보에 대해 갱신 여부를 판단 후 그 결과를 반영한다. 최초로 CTQ 데이터 관리를 시작하는 경우 이러한 레파지토리는 존재하지 않으므로 레파지토리 구현을 수행한다. 경험적으로 CTQ 데이터 레파지토리에는 데이터 일련번호, 데이터 분류(사업, 재무, 보안, 기술, 경영, 표준), CTQ데이터 한글

명, 업무명, 단위업무명, 원천시스템, 도메인 그룹 및 명칭, 데이터 타입 및 길이, 업무 규칙, 중요도점수, 오너쉽 코드, 오너쉽 부서명, 데이터 품질 관리 기준 즉, 정확성, 일관성, 유용성, 적시성, 접근성, 보안성 중 관리 부문 매핑과 세부 품질 기준, DQI 및 측정 방법, 목표 점수, 주기별 측정 점수 등으로 구성할 수 있다.

DQI 및 품질 Metric 정의는 CTQ 데이터 신규 등록이 필요한 경우나 DQ 품질 기준이 변경된 경우 이를 정량적으로 측정하는 방법과 품질 기준을 정의하는 단계로, 반드시 수치화, 계량화 되도록 해야 한다. 만약 측정이 어려운 경우 업무적으로 지침이나 규정의 강화나 절차 등을 통해 품질을 높일 수 있는 방안은 없는지를 모색하여 BPR 요건으로의 접근도 필요하다.

2.7.2 CTQ 대상 데이터 레파지토리 저장

본 절치는 확정된 CTQ 대상 데이터와 관련 정보를 레파지토리에 저장하여 CTQ 데이터가 전사적으로 관리될 수 있도록 하기위한 방법이다. 전사적 차원의 특정 레파지토리를 구성하여 관리함으로써 조직 내 필요로 하는 업무 관련자들이 언제나 확인하고 업무에 활용할 수 있게된다. 저장된 최종 CTQ 데이터는 주기적으로 재 조사를 실시하여 지속적으로 관리되도록 체계화하며, 일부 주요 데이터에 대해서는 실제 품질을 측정하여 목표 점수가 유지되도록 꾸준히 관리해야 한다.

3. CTQ 대상 데이터 관리 활동 제언

3.1 CTQ 대상 데이터 중요도를 활용한 matrix분석

본 연구에서 제시한 CTQ 대상 데이터 선정 방법인 CTQ-DSMM 모델 적용을 간소화하여 CTQ 대상 데이터를 매트릭스 분석(matrix analysis)을 통해 대상 여부를 직관적으로 파악할 수 있는 방안을 소개하고자 한다.

먼저, 통계량을 기초로 CTQ 데이터 matrix를 구성하고, 한글 컬럼명에 차례로 숫자 레이블(예: 사업지구ID에 ①번 표시)을 표시한다.

matrix 분석은 CTQ 핵심 데이터의 기준 점수에 따라 영역이 달라지므로 이 값을 어떻게 선정하느냐는 데이터를 활용하는 조직의 핵심 데이터 관리 목표 수준과 정보시스템의 특성을 고려하여 선정하여야 한다. 본 사례에서는 matrix 기준 점(norm)을 업무관점 중요도와 IT관점 관점 중요도 각각의 평균값으로 설정하였다.

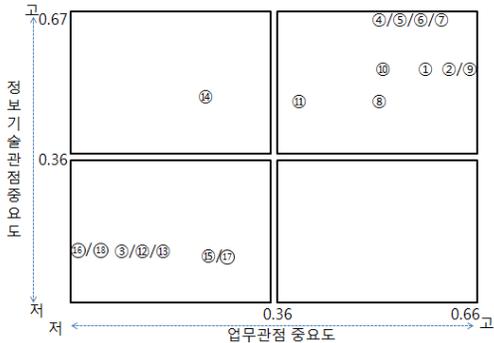


그림 15. CTQ 데이터 matrix 분석
Fig. 15. CTQ data matrix analysis

그림 15의 개별 중요도 점수의 matrix 분석 결과를 보면 업무 관점과 IT관점 중요도 모두에서 평균 이상을 득점한 1번과 2번 데이터가 CTQ 대상 데이터임을 직관적으로 파악할 수 있다. 이 데이터에 대하여 CTQ 레파지토리를 구성하여 전사적으로 관리하고, 주기적으로 점검을 수행함으로써 조직의 데이터의 품질을 효율적으로 높일 수 있을 것이다.

그러나 결과에서 알 수 있듯이 매트릭스 분석은 점수부여 도표에 비해 상대적으로 중요도가 높은 데이터만을 직관적으로 인식하는 데는 도움이 되지만 선별된 CTQ 데이터에 대하여 '얼마나 더'라는 상세 비교에는 적합하지 않다. 따라서 실무에서 어떠한 방식으로 CTQ를 도출하여 관리할 것인가 하는 문제는 데이터의 성격을 고려하고 조직의 CTQ 데이터 관리 수준에 따른 선택의 문제가 될 것이다.

3.2 기관 별 DQC 측정에 대비한 데이터 성숙도 관리 절차

DB 품질 인증제도에 대비하고 조직의 특수성을 감안한 데이터 품질관리가 요구되고 있다. 이를 위해 업무 재설계를 통해 업무 절차, 조직 별 역할과 책임 등을 정의하게 되는데 이것이 불가능한 경우 CTQ-DSMM모형과 한국데이터베이스진흥원에서 피인증기관의 데이터 성숙도 측정을 위한 품질 진단 수행 절차를 그림 16과 같이 점목시켜 응용할 것을 제안한다.

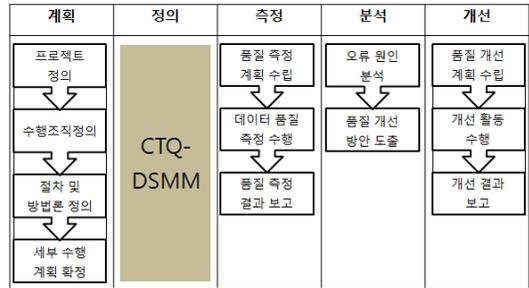


그림 16. CTQ-DSMM을 활용한 데이터 성숙도 관리 모형
Fig. 16. CTQ-DSMM utilizing data management maturity model

IV. 결론

정보의 품질은 수요자의 만족도와 지속적 이용에 영향을 미친다[17]. 데이터 품질관리 관점에서 볼 때, 데이터의 품질은 품질정책, 품질조직, 업무프로세스, 업무규칙 등 여러 요인에 의해 영향을 받는다[18]. 정보화 성숙도가 높아지고 수많은 업무 데이터를 보유하게 되면서 데이터 품질관리의 중요도가 점점 높아지고 있으나 조직들이 어떤 데이터를 선별하여 어떤 품질관리 부문에 어떻게 품질 관리 활동을 수행해야 할 것인가에 대한 연구는 거의 이루어지지 않고 있다.

특히 CTQ 대상 데이터 도출에 대한 연구는 거의 이루어지지 않고 있는 상황인데 데이터 품질 관리 주관기관인 한국데이터베이스진흥원에서는 2010년부터 DQC 제도를 도입하여 정보시스템의 데이터 품질 인증을 추진하고 있다.

본 연구는 이러한 시대적 경향을 반영하여 피인증기관을 위한 CTQ 대상 데이터 선정 모델(CTQ-DSMM)을 개발하고 이를 실제 업무에 적용해 봄으로써 CTQ 대상 데이터를 객관적으로 도출하여 계량화하는 방법을 제시하였다. 또한, CTQ 데이터 matrix 분석을 수행하여 직관적으로 CTQ 데이터를 도출할 수 있는 방법을 제안하였으며, 데이터 성숙도를 높이고자 하는 기관에서 활용할 수 있도록 본 연구에서 제공한 CTQ-DSMM과 한국데이터베이스진흥원에서 피인증기관 평가용으로 제시된 데이터 품질(DQ) 진단 절차를 혼합하여 피인증기관용 데이터 성숙도 관리 프로세스도 함께 제안하였다.

본 연구에서는 이러한 실무적 접근의 첫 시도으로써, 11개의 CTQ 후보 요인을 도출하여 주성분 요인 분석을 통해 그룹핑 및 소거를 통해 최종 도출된 CTQ 대상 데이터 선정 지표를 활용하여 AHP 계층도를 완성하고 이를 평가하고 계량화함으로써 각 회사나 기관에서 간과하거나 추상적인 개념으로 관리해 오던 CTQ데이터를 정량적으로 관리할 수 있는 방안을 실

무적으로 제시했다는데 그 의의가 있다. 따라서 이러한 절차를 조직의 데이터품질관리 시스템 도입 시 반영한다면 기관의 특성을 반영한 품질관리 시스템 구축이 가능할 것이라 판단된다.

그러나 몇 가지 한계도 발견되었다.

첫째, CTQ 데이터 도출에 대한 초기 연구로, CTQ 데이터의 요인 도출이 브레인스토밍을 통하여 설문을 수행하고 요인 분석을 통해 유의미한 요인만을 도출하는 방법을 제안하였으나 사전 연구가 다수 이루어지면 이들 요인을 통합하여 CTQ 데이터 요인을 정의하는 추가 연구가 필요할 것이다. 둘째, DB와 메타데이터시스템을 통한 CTQ 대상 데이터 추출 방법은 DB 규모에 따라 후보 데이터가 지나치게 많아질 우려가 있으므로, 실제 사용 현황을 근거로 객관화된 중요도 평가 방안의 연구가 필요할 것으로 사료된다. 셋째, 전문가로 분류된 담당자가 중요도를 평가하는 방법이 있어서 직접 점수를 부여하는 방법은 주관적이며, 전문가가 없는 경우 사용에 제약이 따르므로 객관적인 CTQ 대상 데이터 도출 및 평가 방법의 연구가 필요하다.

이러한 한계에도 불구하고 CTQ-DSMM 모형은 CTQ 데이터 관리체계 도입에서 품질 레퍼지토리 운영에 이르는 전 과정을 모두 다루고 있으므로 업무적 활용성이 높고, 전사 차원의 CTQ 데이터 체계 수립 시 유용할 것으로 사료된다.

참고문헌

- [1] Jae-Seob Lee et al., "A Feasibility Analysis of Integrated Project Delivery by MAUT-AHP in Public Construction Projects", 2010.
- [2] Korea Database Agency, "Data quality management guidelines", 2011.
- [3] Changhwan Lee, "A pleasant change, depends on the quality challenge", I Love Six Sigma, POSCO, pp. 10-11, 2004.
- [4] MiYoung Park, Hyon-Woo Seung, "A Selection Method of Database System Quality Characteristics Using the Analytic Hierarchy Process", 2009.
- [5] Korea Database Agency, "Guideline for the management of data quality", 2010.
- [6] Korea Database Agency, "Data Quality Management Maturity Model", 2010.
- [7] Korea Database Agency, "Process-based data quality master courses", 2012.
- [8] Chan Soo KIM, Joo Seok Park, "A Study of Data Quality Management Maturity Model", 2003.
- [9] Hcsong, Cwkang, "A Study on the Data Architecture Implementation and the Evaluation Method of Quality Indicator", 2010.
- [10] IEEE-1471, "Recommended Practice for Architectural Description of Software-Intensive Systems", 2000.
- [11] Richard Y. Wang AND Diane M. Strong. "A conceptual framework of data quality (Beyond Accuracy : What Data Quality Means to Data Consumers)", Journal of management information systems, 1996.
- [12] Zachman John, "A Framework for Information System Architecture", IBM System Journal V.26 No.3, pp.276-292, 1987.
- [13] Thomas. C. R., "The Impact of Poor Data Quality on the Typical Enterprise", Communications of the ACM, Vol. 41. No. 2, February, pp.79-82, 1998.
- [14] Ballou. D. P. And Pazer. H., "Modeling Data and Process Quality in Multi-input. Multi-output Information System". Management Science, Vol. 31. No 2. February, pp.150-162, 1985.
- [15] Wang. R. Y. and Strong D. M ., "Beyond Accuracy : what Data Quality Means to Data consumers", Journal of Management Information Systems, Vol 12, No. 4, Spring, pp.623-640, 1996.
- [16] Korea Database Agency, "Data quality management practices", 2011.
- [17] byeongho Jeun, byeonggoo Kang, "Effects of Information Quality on Customer Satisfaction and Continuous Intention to buy in Social Commerce", Journal of The Korea Society of Computer and Information
- [18] Sunngho Shin et al., "A data cleansing strategy for improving data quality of National R&D Information - Case study of NTIS", Journal of The Korea Society of Computer and Information, Vol.16 No.6, pp.119~130, 2011
- [19] DB domestic quality management maturity yet introduced step, <http://www.itdaily.kr/news/>

articleView.html

- [20] The Essentials of Information Quality Management,
<http://www.information-management.com/issues/20020901/5690-1.htm>
- [21] Database Quality Management,
<http://www.dqc.or.kr>

저 자 소 개



김 승 희
2003: 동국대학교
컴퓨터멀티미디어공학 공학사
2005: 연세대학교
산업정보경영 공학석사
2011: 서울과학기술대학교
IT정책전문대학원 박사 과정
현 재: 한국토지주택공사
경영정보처 근무
관심분야: 품질공학, IT서비스 최적화
Email : kim.sh@LH.or.kr



김 우 제
1986: 서울대학교
산업공학과 공학사
1988: 서울대학교
산업공학과 공학석사
1994: 서울대학교
산업공학과 공학박사
현 재: 서울과학기술대학교
글로벌융합산업공학과 교수
관심분야: 소프트웨어공학, IT서비스
Email : wjkim@seoultech.ac.kr