

Derivation of Data Quality Attributes and their Priorities Based on Customer Requirements

Kyoung-Ae Jang[†] · Ja-Hee Kim^{**} · Woo Je Kim^{***}

ABSTRACT

There is a wide variety of data quality attributes such as the ones proposed by the ISO/IEC organization and also by many other domestic and international institutions. However, it takes considerable time and costs to apply those criteria and guidelines to real environment. Therefore, it needs to define data quality evaluation attributes which are easily applicable and are not influenced by organizational environment limitations. The purpose of this paper is to derive data quality attributes and order of their priorities based on customer requirements for managing the process systematically and evaluating the data quantitatively. This study identifies the customer cognitive constructs of data quality attributes using the RGT(Repertory Grid Technique) based on a Korean quality standard model (DQC-M). Also the correlation analysis on the identified constructs is conducted, and the evaluation attributes is prioritized and ranked using the AHP. As the results of this paper, the consistent system, the accurate data, the efficient environment, the flexible management, and the continuous improvement are derived at the first level of the data quality evaluation attributes. Also, Control Compliance(13%), Regulatory Compliance(10%), Requirement Completeness(9.6%), Accuracy(8.4%), and Traceability(6.8%) are ranked on the top 5 of the 19 attributes in the second level.

Keywords : Data Quality, Data Quality Model, RGT, AHP, Factor Analysis, Content Analysis

고객의 요구사항에 기반한 데이터품질 평가속성 및 우선순위 도출

장 경 애[†] · 김 자 희^{**} · 김 우 제^{***}

요 약

데이터품질 속성으로는 ISO/IEC 기관 및 국내/외 여러 기관에서 제시한 속성이 존재하지만, 이러한 기준 및 가이드를 현실적으로 조직에 적용하기에는 시간과 비용이 상당히 소요된다. 따라서 조직환경의 제약사항이 존재하여도 적용 가능한 데이터품질 평가속성의 정의가 필요하다. 이 연구의 목적은 고객의 요구사항 기반하에 프로세스를 체계적으로 관리하고, 정량적으로 데이터를 평가하기 위한 데이터품질 평가속성과 우선순위 도출에 관한 연구이다. 본 연구에서는 데이터품질 표준(DQC-M)을 매개체로 RGT 기법을 사용하여 데이터품질 속성의 고객 인지구조(Construct)를 도출하고, 도출된 Construct 간의 상관분석을 수행하여 AHP기법으로 평가속성의 가중치 및 우선순위를 선별하였다. 그 결과 데이터품질 평가속성에서 1레벨에서는 일관된 체계, 정확한 데이터, 효율적 환경, 유연한 관리, 지속적 개선 순위가 결정되었다. 또한 2레벨의 19개 속성 중에서는 통제성(13%), 준거성(10%), 요구완전성(9.6%), 정확성(8.4%), 추적가능성(6.8%)이 상위 5순위로 결정되었다.

키워드 : 데이터품질, 데이터품질 평가모델, RGT, AHP, 요인분석, 내용분석

1. 서 론

1.1 연구배경 및 목적

산업의 성장과 함께 조직의 가치는 고품질의 데이터를 확

보하였는지 여부와 직결된다. 조직이 데이터의 수집과 축적에 그치지 않고 데이터를 지식화하여 효과적으로 고객이 원하는 고부가가치를 도출하기 위해서는 데이터품질 문제 해결이 선행되어야 한다. 따라서 기업 및 공공기관, 연구기관 등 데이터를 처리하는 조직에서 데이터품질 관리는 상당히 중요하다.

오늘날 데이터가 대량화되고 정형데이터와 비정형, 반정형데이터의 여러 유형을 포함하는 빅데이터가 이슈화됨에 따라 데이터의 중요성이 가중되고 있다. 특히 공공기관에서는 공공데이터의 개방과 공유를 위해서 보유하고 있는 데이

※ 이 논문은 서울과학기술대학교 교내 연구비지원에 의하여 연구되었음.
† 준 회 원 : 서울과학기술대학교 IT정책전문대학원 산업정보시스템전공 박사과정
** 비 회 원 : 서울과학기술대학교 IT정책전문대학원 산업정보시스템전공 부교수
*** 비 회 원 : 서울과학기술대학교 글로벌융합산업공학과 교수
Manuscript Received : September 23, 2015
First Revision : November 19, 2015
Accepted : November 20, 2015
* Corresponding Author : Woo Je Kim(wjkim@seoultech.ac.kr)

터의 품질관리에 대한 관심과 투자가 높아지고 있는 추세이다. 한국데이터베이스진흥원에서는 '2014 데이터베이스 백서'를 발간하면서 데이터품질 시장은 빅데이터 시장을 포함하여 연평균 7.5% 성장하여 2016년에는 800억의 시장을 형성할 것으로 전망하였다[1].

데이터를 포함한 정보시스템의 품질관리는 소프트웨어 품질에 대한 관심에서 출발하였다. 소프트웨어 품질향상은 위한 다양한 각도에서 연구가 진행되었고 국제표준과 가이드라인을 제시하는 등 품질향상의 기반을 마련하였다. 품질경영의 전반을 다루는 ISO 9000과 소프트웨어 프로세스에 기반한 생명주기의 ISO/IEC 12207과 SPICE가 있다. 또한 프로세스 품질의 특성기준 ISO/IEC 9126, 제품에 대한 평가 ISO/IEC 14598 등이 있다. 데이터의 품질을 광의의 소프트웨어 품질로 보는 견해도 있으나 소프트웨어 품질요소로 데이터 품질을 관리하기에는 상세 적용의 어려움이 있다.

따라서 데이터품질은 소프트웨어 품질과 별개로 분리하여 데이터 요구사항을 식별하고 데이터품질 지표와 표준을 제정해야 하며, 이를 통해서 데이터품질 측정과 평가가 이루어져야 한다. 데이터품질의 확보를 위한 국제표준으로는 ISO/TS 8000-150과 ISO/IEC 25012, DMBOK 가이드 등이 있으며, 국내 표준으로는 한국데이터베이스진흥원에서 제시하는 데이터베이스 품질과 인증가이드(DQM3, DQC) 등이 존재한다.

그러나 데이터품질 표준은 소프트웨어 품질에 비해서는 부족한 실정이며 실제 현장의 자원에 유연히 적용할 수 있는 데이터품질 지표 마련은 중요한 과업이다. 품질관리는 제품이나 서비스에 대한 고객의 근본 의도와 요구사항을 만족시키는 데 필요한 전반적 요소를 도출하여 관리하는 것이다. 데이터품질은 데이터 사용자의 요구사항을 충족하는 정도를 의미하며 사용자가 작업을 수행할 때 지속적으로 요구되고 기대하는 데이터를 제공하는 것이라고 하였다[2, 3]. 고객이 요구하고 기대되는 데이터의 품질을 확보하기 위해서는 고객이 필요로 하는 데이터품질 요구사항과 품질특성을 정확히 이해해야하고 이를 통해 품질측정과 평가가 이루어져야 한다. 따라서 데이터품질에 대한 연구는 고객의 데이터 요구사항 분석에서부터 출발하여야 하며, 이를 실제 현장에 적용할 수 있도록 실증적인 연구로 귀결되어야 할 필요가 있다. 또한 데이터품질 속성을 정의하고 품질을 객관적이고 정량화할 수 있는 평가기준이 필요하다.

그러나 기존의 선행연구는 기 정의된 품질표준을 조합하거나 재정의하는 방향으로 진행되었으며[4-7], 고객의 요구사항에 기반한 데이터품질의 특성이나 평가 방안에 관한 연구는 찾아보기 힘들었다. 본 연구에서는 고객의 데이터요구사항을 분석하고 이를 정량화하여 데이터품질 평가를 위한 지표를 개발하였다. 고객의 요구사항을 도출하기 위해서 심리학에서 사용되는 RGT 기법을 적용하였으며, 이를 통하여 고객의 심리적인 인지구조를 분석하여 데이터품질 속성을 도출하였다. 이를 전문가의 양극성 평가(Bipolar Rating)로 정량화하고 내용분석(Content Analysis)으로 정제한 데이터

를 요인분석(Factor Analysis)과 AHP 기법을 통해서 계층화된 평가지표를 개발하였고 기존 품질평가 지표와 비교검증을 수행하였다.

본 연구는 고객의 요구사항분석에 기반하여 객관적이고 신뢰성을 높이는 분석기법 및 의사결정기법을 적용하여 데이터품질 평가 지표를 개발한 것에 그 의의가 있다. 본 연구에서 개발된 데이터품질 평가 지표는 프로젝트가 보유한 자원에 따라 선별적으로 데이터품질 평가를 수행할 수 있으며, 한국데이터베이스진흥원에서 시행하는 데이터품질 관리인증(DQC-M)에 대비하기 위한 가이드로 활용될 수 있다. 또한 데이터품질 관리를 위한 전사적인 관리체계 구축을 위해 활용될 것으로 판단된다.

본 연구의 구성은 다음과 같다. 2장에서는 연구의 이론적 배경과 대표 선행연구를 소개하고, 3장에서는 연구의 설계 방법을 설명하고 실험 결과를 분석한다. 마지막 4장에서는 결론과 향후 추가 연구과제에 대해서 논의한다.

2. 선행연구 고찰 및 이론적 배경

2.1 이론적 배경

1) 데이터품질 표준

국제 데이터품질 관련 표준은 ISO/IEC 기관에서 제시를 하고 있으며 국내에서는 한국데이터베이스진흥원에서 가이드를 제시하고 있다. 데이터품질에 국한한 품질로서 대표적으로 활용되고 있는 것은 ISO/IEC 25000 시리즈에서 데이터품질에 해당하는 ISO/IEC 25012이다. ISO/IEC 25012는 개발이나 운영에서 사용되는 모든 데이터의 품질에 대한 데이터품질 평가 모델을 제시한다[8]. ISO/IEC 25012는 Table 1과 같이 16개의 데이터품질 평가항목을 통해서 품질평가를 가이드하고 있다[8].

Table 1. Data Quality Attribute of ISO/IEC 25012

No.	DQ. Attribute	No.	DQ. Attribute
1	일관성(Consistency)	9	이해성(Understandability)
2	현재성(Currentness)	10	효율성(Efficiency)
3	완전성(Completeness)	11	가변성(Changeability)
4	정확성(Precision)	12	이식성(Portability)
5	정밀성(Accuracy)	13	추적성(Traceability)
6	보안성(Security)	14	신뢰성(Creditability)
7	유효성(Availability)	15	접근성(Accessibility)
8	복구성(Recoverability)	16	준수성(Compliance)

첫 번째는 일관성(Consistency)으로 데이터의 전과 후가 통일성 있게 사용되는지 평가하는 것이고, 두 번째 현재성(Currentness)은 데이터 변경이 발생할 경우 즉시 업무에 반영되는지 평가하는 것이다. 세 번째 완전성(Completeness)은 데이터를 사용하는 관점에서 얼마나 완벽한가를 평가하는 것이고, 네 번째 정확성(Precision)은 데이터가 정확하게

기록되고 수집되는가를 평가하는 것이다. 다섯 번째 정밀성(Accuracy)은 데이터가 정확한 속성을 표현하는가를 평가하는 것이고, 여섯 번째 보안성(Security)은 권한받은 사용자에게만 접근을 허용하는지 평가하는 것이며, 일곱 번째 유효성(Availability)은 데이터가 회복되는지 여부에 대해 평가하는 것이다. 여덟 번째 복구성(Recoverability)은 데이터가 성공적으로 복구되는지에 대해 평가하는 것이고, 아홉 번째 이해성(Understandability)은 데이터의 언어, 심볼, 단어를 이해하고 표현 가능한지에 대해 평가하는 것이다. 열 번째 효율성(Efficiency)은 저장 공간에 대한 효율성에 대해 평가하는 것이고, 열한 번째 가변성(Changeability)은 데이터의 형태, 길이, 값이 변화 가능한지에 대해 평가하는 것이다. 열두 번째 이식성(Portability)은 데이터가 다른 플랫폼으로 이식 가능한지에 대해 평가하는 것이고, 열세 번째 추적성(Traceability)은 데이터의 원본과 변경을 자동적으로 감시할 수 있는지에 대해 평가하는 것이다. 또한 열네 번째 신뢰성(Creditability)은 데이터가 사용자에게 믿어지는 정도에 대한 평가이고, 열다섯 번째 접근성(Accessibility)은 데이터 접근 및 활용정도에 대한 평가이다. 마지막으로 열여섯 번째는 준수성(Compliance)으로 데이터 관련 규칙에 적합한지에 대한 평가를 제시하고 있다.

ISO/TS 8000-150은 마스터데이터 관리 프레임워크로서, 데이터의 교환과 추출에 대한 관리원칙과 요구조건을 제시한다. 또한 데이터품질 관리 프로세스의 프레임워크도 포함되어있으며, 별도의 ISO표준을 결합해서 사용할 수 있도록 방안을 제시하고 있다. 이 프레임워크는 데이터활용, 품질 모니터링, 품질 개선의 3단계로 구성되어있고 3개의 프로세스 레벨과 9개의 프로세스로 구성되어있다[9]. 또한 ISO/TS 8000-150은 국내 주도하에 19개국의 전문가들이 표준화에 참여한 것으로 의미가 크고 향후 발전가능성도 높다.

국내의 한국데이터베이스진흥원에서는 데이터품질 관리를 확산하기 위하여 데이터베이스 품질 인증제도(DQC, Database Quality Certification)를 제시하고 있다. DQC는 정보시스템의 데이터품질을 확보하기 위하여 데이터 값(Value) 자체와 데이터 관리(Management), 데이터 보안(Security)으로 분류하여 범국가적인 데이터품질 제고의 목적으로 실시되고 있다. 데이터 심사대상으로는 Table 2와 같이 데이터 인증(DQC-V), 데이터 관리 인증(DQC-M), 데이터 보안 인증(DQC-S)이 존재하며 상세 점검기준은 다음과 같다[10-13].

데이터 인증(DQC-V)은 데이터 값 자체에 대해서 일관성과 무결성을 보장하는지 여부에 대해서 점검하고 인증하며, 데이터 관리 인증(DQC-M)은 전사 조직의 관리체계를 점검하고 인증하는 것이다. 또한 데이터 보안 인증(DQC-S)은 데이터베이스 시스템에 보관되는 데이터를 허가받지 않은 외부의 침입으로부터 보호할 수 있는 체계가 마련되어있고 기술적으로 시스템보안을 수행하고 있는지에 대해 점검하고 인증한다. 본 연구에서는 고객이 데이터품질 확보를 위해 중요하게 생각하는 속성요소를 도출하기 위하여 데이터 관리 인증(DQC-M)의 요소를 활용한다. 또한 계층적 품질속

성을 도출한 최종 결과를 DQC-M과 매핑하여 활용방안을 제시한다.

Table 2. Database Quality Certification of Korea Database Agency

Audit	Audit Define	Audit Contents
Data Value Certification (DQC-V)	Audit the data quality of the database	1 Level: Domain, business rules 2 Level: Number, amount of money, name, quantity, category, date, rate, content, code, key, Manager/ products/ activities/ transactions/ resources/ support/ production domain, business rules
Data Management Certification (DQC-M)	Audit the management level of the organization of an information system	1 Level: Validity, Practicality 2 Level: Accuracy, Consistency, Utilization, Accessibility, Currentness, Security
Data Security Certification (DQC-S)	Audit the level of security on the database	1 Level: DB security policies, access control rules defined 2 Level: DB documentation of security policies, Organizational matters, such as policies, Checking the password and system requirements

2) RGT(Repertory Grid Technique)

RGT는 심리학 분야에서 많이 활용하는 인간의 심리에 기반한 활동요소를 도출하는 연구방법으로, 임상심리학자 George Kelly가 처음 제안한 개인 인지구조 이론(Personal Construct Theory)에 기반한 방법론이다[14]. 인간은 수년간 경험에 의해서 축적된 인지구조를 갖고 있으며, 이 인지구조를 다른 요소와 구별하기 위해서는 해당 요소의 특성에 대해 유사성과 차별성을 통해서 이해한다고 본다. 개인의 경험을 기반으로 축적된 인지구조(Construct)는 양극(Bipolar)으로 존재하며 이렇게 축적된 지식을 얻고자 할 경우는, 직접 물어보는 것이 최선이라고 하였으며[14] 이를 위한 심층적 인터뷰 기법이 RGT이다. 심층적 인터뷰를 위해서는 요소(Element)라는 관심의 대상이 되는 자극제가 필요하고, 요소의 자극에 의해서 개인이 갖고 있는 내면에서 요소에 대한 판별과 해석을 인지구조(Construct)의 형태로 도출하게 된다. 그리고 요소와 인지구조 간의 관계의 정도를 정량적으로 수치화하여 연결(Link)한다. 따라서 RGT의 구성요소는 요소(Element), 인지구조(Construct), 연결(Link)로 구성된다[15].

RGT는 해당 분야의 전문가를 대상으로 심층 인터뷰를 실시하는 것이 효과적이며 자극제인 요소(Element)를 도구로 보여주면서 전문가의 경험적 지식을 도출하고 정량화한다. 해당 분야의 전문가들이 관심이 있는 상황이나, 사례, 도구, 인물 등을 제시하여 “더 좋은 것을 선택해보세요. 왜 그

것이 다른 것보다 더 좋은가요?” 혹은 “세 개 중 두 개와 하나로 분류해보세요. 그 이유는 무엇인가요?”를 반복 질문하여 인지구조를 도출한다. RGT기법을 활용하여 인지구조를 도출하는 방법에는 모나딕(Monadic), 다이아딕(Dyadic), 트리아딕(Triadic) 방식이 있다. 모나딕(Monadic)은 하나의 요소(Element)를 통해 간략한 문구를 설명하도록 하는 방식이고, 다이아딕(Dyadic)은 둘 중에 다른 차이점을 통해서 인지구조를 도출하는 방법이다. 트리아딕(Triadic)은 가장 일반적으로 많이 쓰이는 방식으로 세 개의 요소 중에서 둘과 다른 하나로 분류하게 하고 그것의 이유를 통해서 인지구조를 도출하는 방법이다. 인터뷰는 응답자가 새로운 인지구조를 제시하지 못할 때까지 반복하여 수행해야 하므로, 1시간가량은 인지구조를 찾는 질문에 소모되고 40분 정도는 양극화와 가치의 정량화 등에 소모된다[16] 도출된 인지구조는 ‘크다-작다’, ‘좋다-나쁘다’, ‘딱딱하다-부드럽다’ 등의 양극(Bipolar)개념이 존재하는데 이를 데이터로 수집하여 척도로 정량화(Rating)하면 분석데이터로 활용할 수 있다.

본 연구에서는 한국데이터베이스진흥원의 데이터 관리 인증(DQC-M)방법을 요소(Element)로 사용하여 다이아딕(Dyadic)방식을 적용하여 인지구조를 도출하였다. 도출된 인지구조를 양극화하여 연결(Link)로 정량화된 수치를 데이터 품질 평가요소의 계층화를 위한 기초 데이터로 활용하였다.

3) 내용분석(Content Analysis)

내용분석(Content Analysis)은 유사한 목적으로 도출된 문맥의 내용을 통합하고 분류하는 기술이다[16]. 인터뷰에서는 다양한 인터뷰 대상자가 동일한 의미지만 다른 그들의 용어로 응답하게 된다. 이때 동일한 데이터로 통합하기 위한 데이터표준화의 기술이 필요하다. 내용분석(Content Analysis)은 질문자의 분류(Categorization) 단계, 협력자의 분류(Categorization) 단계, 신뢰도 테이블(Reliability Table)을 통한 상호 검증단계를 통해서 문맥의 내용을 분석하고 표준화하는 기술이다[16]. 분류(Categorization) 단계는 Fig. 1과 같이 7단계를 거쳐서 분류된다. 질문자와 협력자 각자가 분류한 결과를 가로축에는 협력자의 결과, 세로축에는 질문자의 결과를 위치시키고 매트릭스 형태로 결과값을 매핑하여 서로 통합된 결과를 도출하는 방식으로 진행한다. 본 연구에서는 내용분석(Content Analysis)을 통해서 RGT에서 도출된 Construct를 합치고, 분류 및 검증한다.

<p>Step 1) Generating the first item category Step 2) Create a separate category if there are differences in the following items Step 3) Allocated to the appropriate category if the item is similar to the following Step 4) If necessary, create categories (redefine existing categories considered) Step 5) Repeat the process until all the items in this category Step 6) If the categorization is difficult to classify, create a 'Other' category Step 7) Proceeds until no more than 5% of the total other</p>
--

Fig. 1. Process of Content Analysis

4) 요인분석(Factor Analysis)

요인분석은 다중변수의 관계에서 공통의 요인을 분석하는 방법으로, 종속변수와 독립변수를 분리하지 않고 변수들끼리 유사한 분산구조를 기반으로 데이터를 분류하는 방법으로 데이터의 계층화를 용이하게 한다[17].

데이터 분석방법에는 요인분석(Factor Analysis) 이외에도 군집분석(Cluster), 판별분석(Classification)이 있다. 군집분석(Cluster)은 관찰대상의 유사성에 근거하여 유사한 동질의 집단으로 분류하는 방법으로 조사 대상들을 그룹핑할 때 사용되며, 판별분석(Classification)은 사전에 집단이 이미 나뉜 상태에서 집단의 특성과 유사한 데이터를 해당 그룹으로 판별하는 방법이다. 요인분석을 본 연구의 데이터 분석기법으로 선정한 이유는 군집분석(Cluster), 판별분석(Classification)과 달리 요인분석(Factor Analysis)이 사전에 나뉜 집단이 없는 다수의 변수들에서 유사성을 도출하여 변수를 그룹핑하는 방법이기 때문이다.

본 연구에서는 RGT를 기반으로 도출된 인지구조(Construct)의 품질속성 변수가 AHP의 계층구조화를 수행하기에는 변수 축소가 필요하다고 판단되어, AHP의 선행 작업으로 RGT와 병행하여 요인분석을 사용하였다.

5) AHP(Analytic Hierarchy Process)

계층분석과정(AHP)은 Thomas Saaty에 의해 개발된 다기준 의사결정의 문제를 해결하는 방법으로, 문제 해결에 여러 가지 대안이 존재할 때 대안을 계층적으로 구조화하여 분류하고, 대안마다 가중치를 부여하여 최적의 대안을 객관적으로 평가할 수 있도록 하는 방법이다[18]. AHP는 최적의 의사결정을 위해서 대안들을 계층화하고 전문가의 판단에 의해 수집된 대안 데이터를 쌍대비교하여 가중치를 도출하고 이를 통해 우선순위를 산정한다. Saaty는 그의 연구 [18]에서 AHP 분석기법의 적용절차를 계층(Hierarchy)모형 생성, 쌍대비교(Pairwise Comparison), 부분 우선순위(Local Priority) 도출, 일관성 검사(Consistency Check), 전체 우선순위(Global Priority) 도출 및 대안 선택의 5단계로 제시하였고, 이를 조근태와 그의 동료는 4단계로 재분류하였다[19]. 1단계는 의사결정 문제를 상호 관련성 기반으로 계층화하여 의사결정 계층모형을 설정하는 단계이다. 이 단계에서는 복잡한 다기준 의사결정 기준을 갖고 대안의 계층화모형을 만든다. 2단계는 의사결정 대안들 간의 쌍대비교를 통해서 전문가의 판단정보를 수집하는 단계이다. 최하위 계층의 대안이 n 개이면 $n(n-1)/2$ 회의 비교를 수행한다. 3단계는 쌍대비교에서 고유치 방법을 사용하여 상대적 가중치를 종합한다. 일관성 검증, 가중치 추정을 이 단계에서 산정한다. 4단계는 여러 평가에서 종합순위를 산정하기 위해서 상대적 가중치를 종합하고 결론을 도출한다.

본 연구에서는 고객의 요구사항을 기반으로 데이터품질을 향상시키기 위한, 평가속성의 판단에 논리적인 일관성 검증으로 합리적인 의사결정 모형을 제공하기 위하여 AHP기법을 사용하였다. 또한 AHP의 1단계 의사결정 계층모형 구축

은 RGT와 요인분석으로 수행하였고 이를 9점 척도로 쌍대 비교하여 가중치와 우선순위를 도출하였다.

2.2 선행 연구의 고찰

기존 데이터품질의 평가에 관한 연구를 살펴보면, 정혜정은 [5]의 연구에서 ISO/IEC 25012를 기반으로 6개 품질특성의 평가항목 산식을 제시하고 우편번호구축 DB를 기준으로 실증하였으며, 김선호와 이창수의 [6]의 연구에서는 ISO 15504 SPICE를 기준으로 프로세스 평가 모델을 제시하고 ISO 8000-150의 프로세스와 비교하였다. 데이터품질 특성을 선정하고 AHP기법으로 적용한 연구로는 박미영과 승현우의 [4]의 연구방법에서 확인할 수 있었다. 그러나 이들의 연구들은 표준화 기관에서 정의된 표준을 조합하거나 분리하여 새로운 모델을 만드는 연구들이었다. 고객의 요구사항을 기반으로 접근한 연구를 살펴본 결과 Brombacher과 그의 동료들의 연구[24]가 선행되었으며, 이 연구에서는 고객의 요구사항을 실증적인 사례를 기반으로 연구를 수행하였으나 품질속성으로 집약하지 못한 아쉬움이 있었다.

본 연구에서는 데이터품질 평가속성의 가중치와 우선순위를 정량적 기법으로 도출하기 위하여 AHP기법을 사용하였다. AHP 연구는 다양한 연구도메인과 다양한 계층화모형 개발 방법으로 연구되었다. AHP 연구의 대부분은 박미영과 승현우의 연구 [4], Badri의 연구 [7]과 같이 문헌조사에 의해서 계층모형을 생성하고 쌍대비교를 진행하여 결과를 분석하는 절차를 거친다. 그 이외의 AHP 연구에서 계층모형 방법을 살펴보면, 김순영의 메타평가 지표 개발에 대한 연구 [22]에서는 메타평가의 문헌조사와 사업의 특성을 조사한 요인을 통합하여 요인분석(Factor Analysis)을 수행하였고, 김승희와 김우제의 연구 [20]에서는 비즈니스 측면과 정보화 측면으로 나누어 중요한 데이터 선정에 우선고려 요인을 문헌조사와 설문을 통해서 항목을 완성하고 요인분석(Factor Analysis)을 실시하였다. 또한 단계별 다중회귀분석법(SMR)을 활용하여 계층모형화를 시도한 연구 [23]도 있으며, 최근의 연구로는 RGT를 활용하여 심층인터뷰를 실시하고 군집화하여 계층모형을 생성한 연구 [21]도 시도되었다.

AHP 방법론의 1단계 계층화모형 개발 단계를 재분류하면 대안항목 수집 단계와 계층화 단계로 나눌 수 있다. 선행연구의 대안항목 수집단계에는 대부분이 문헌조사와 설문 및 인터뷰를 실시하였고 새로운 시도로는 RGT가 있었다. 계층화 단계에서는 문헌분석, 요인분석, SMR, 군집분석이 있었다. 본 연구에서는 고객 중심의 데이터품질 요구사항을 도출하기 위하여 대안항목 수집 단계에서는 RGT기법으로 고객의 인지구조를 도출하였고, 계층모형화 단계에서는 고객의 요구사항을 소실 없이 축소하고 계층화하기 위하여 요인분석을 실시하였다. RGT와 요인분석을 통합한 AHP 기법은 선행연구에서 수행한 문헌분석의 상이한 연구결과 수준에 따라 중복을 배제하기 어려운 단점을 보완할 수 있고, 설문조사, SMR, 요인분석을 개별적으로 수행한 방법에 비해서는 심층적인 고객의 요구사항을 객관적으로 도출할 수 있는 장점이 있다.

3. 연구 설계 및 결과

3.1 연구모델 설계 및 방법론

데이터품질 평가모형은 Fig. 2와 같으며, 고객의 요구사항을 기반으로 데이터품질 평가속성을 도출하고 계층화모형 및 속성의 우선순위를 산정하여 실제 프로젝트 현장에서 관리시간과 관리비용을 고려하여 선택적으로 적용할 수 있도록 하였다. 1단계는 선행 데이터표준을 분석하여 RGT를 적용하기 위한 요소를 선정하는 단계이다. 데이터품질과 관련된 표준으로 ISO/IEC 25012, ISO/TS 8000-150, DQC-V, DQC-M, DQC-S 등을 분석하였으며, 이 중에서 DQC-M의 품질관리 방법을 요소로 선정하였다. 그 이유는 DQC-M이 데이터관리 전반을 포함하면서 국내 현장에서 적용하기 쉽고 이해가 용이하기 때문이다. 다음 단계는 DQC-M의 관리기준을 통해서 실제 정보시스템의 데이터를 관리하는 고객을 대상으로 RGT를 수행하는 단계이다. 이 단계는 심리적 인터뷰 및 정량화 단계로, 이전 단계에서 선정된 데이터품질 요소를 도구로 활용하여 심층인터뷰와 도출된 인지구조의 정량화(Rating)를 실시한다. “데이터품질 향상을 위해서 중요하다고 생각하는 항목 2가지를 선택하세요. 왜 그것이

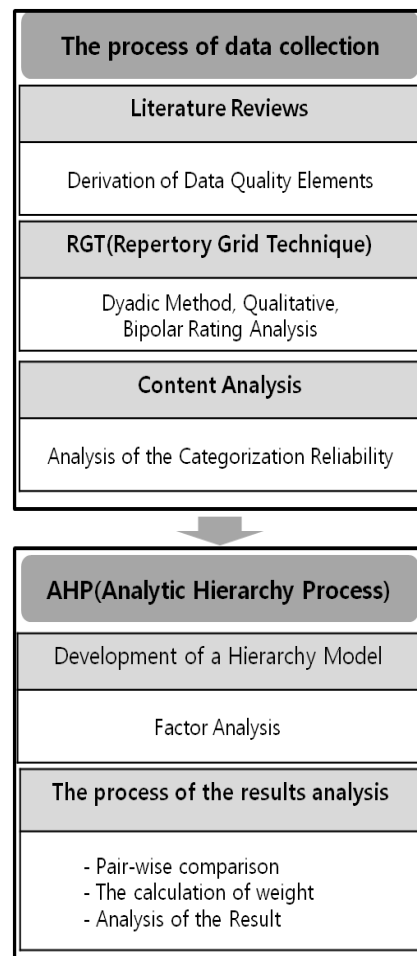


Fig. 2. Research Framework

다른 것에 비해 데이터품질 향상에 중요하다고 생각하나요? 다른 것과 차이점은 무엇인가요?”를 반복하여 질문을 한다. 세 번째 단계는 내용분석(Content Analysis) 단계이다. RGT는 인터뷰로 진행되기 때문에 동일한 용어가 다른 용어로 표출될 수 있으므로 이러한 용어들의 데이터 표준화가 필요하다. 따라서 도출된 데이터품질 속성의 표준화 및 검증을 내용분석 방식으로 수행한다.

이 단계에서는 제3의 전문가를 통해서 RGT 분석결과와 품질속성을 표준화하도록 하고 연구자의 표준화와 비교검증을 통해서 정제작업을 수행한다. 네 번째 단계는 RGT로 표준화된 데이터품질 속성의 정도를 파악하여 속성의 수가 AHP 분석에 적합하도록 조정하는 요인분석(Factor Analysis)을 수행하고 AHP를 준비하는 단계이다. 마지막으로 정제된 데이터품질 속성을 전문가 그룹을 활용하여 AHP분석을 수행하고 각 속성항목의 가중치와 우선순위를 도출한다.

3.2 연구결과 및 토의

1) RGT 및 내용분석

RGT의 데이터 수집 대상은 Table 3과 같이 데이터시스템을 사용한 경험이 있는 사람으로 평균 20년의 경력을 갖춘 기술사 및 특급기술자 19명으로 구성되었다.

Table 3. The Summary of Information of RGT Respondents

Information Item	Summary
Avg. Experience	20 year
Gender	Male(90%), Female(10%)
Professional Composition	Professional Engineer
Data Quality management execution ratio	100%
Data Information System using ratio	100%

고객의 요구사항을 도출하기 위하여 적용하는 RGT방법론에는 요소(Element)라는 자극제가 필요하며, 이 자극제를 기반으로 고객이 더 이상 답변을 하기 어려운 상태까지 질문을 반복하여 인지구조(Construct)를 도출한다. DQC-M에서 정의한 데이터품질 관리 방법 16가지를 선행 FGI를 통해서 7가지 요소로 재정의하여 인터뷰 결과 130개의 Construct를 도출하였으며 이를 중복을 배제하고 표준화하여 Table 4와 같이 21개의 데이터품질 속성으로 통합하였다.

도출된 Construct를 통해서 데이터품질 속성을 표준화하고 이를 전문가 그룹에서 5점 척도로 양극성 평가(Bipolar Rating)로 정량화하였으며, 그 기하평균은 Fig. 3과 같다. a는 통제성, b는 예측성, c는 다양성, d는 일관성, e는 활용도, f는 최선성, g는 보안성, h는 업무수용성, i는 전달성, j는 정확도, k는 상호호환성, l은 시간 효율성, m은 자원 효율성, n은 요구 완전성, o는 가변성, p는 분석성, q는 책임성, r은 준거성, s는 운용성, t는 추적가능성, u는 분류성을 의미하고 1은 영향이 큰 쪽이고 2는 영향이 작은 쪽이다.

Table 4. The Result of Extract of Elements and Constructs

No.	Element Items	Construct Items	Attribute
1	Data structure management	Integrated management of all activities	가변성 (Changeability)
2	Data Integration Management	Continuous data management	보안성 (Security)
3	Standard data management	Identify of the data discarding period	분석성 (Analyticity)
4	Ownership of data management	Management of the integrated user View	상호호환성 (Inter-Compatibility)
5	Data Flow Management	Management of enterprise data needs	시간효율성(Time Efficiency)
6	Management data needs	Control management	업무수용성 (WorkSoluble)
7	Data Monitoring	Consistent data management	예측성 (Predictability)
...	-
21	-	...	요구완전성 (Request Completeness)
...	-	...	-
130	-	Prevention of data abuse	-

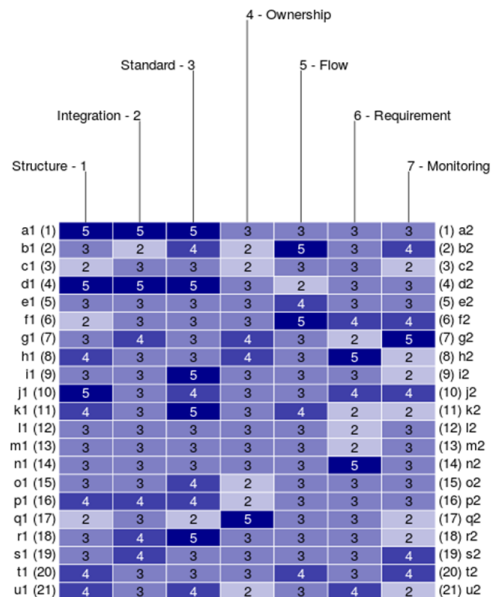


Fig. 3. The Result of RGT Rating

양극성 평가(Bipolar Rating)는 Fig. 3의 왼쪽 요인이 데이터품질에 영향이 큰 경우 최대 5점을 부여하고 오른쪽은 영향이 작은 경우로 최소 1점을 부여하는 방식으로 설문하여 정량화하였다. 세로항목은 통제성, 예측성, 다양성, 일관성, 활용도, 최선성, 보안성, 업무수용성, 전달성, 정확도, 상호호환성, 시간효율성, 자원효율성, 요구완전성, 가변성, 분석성, 책임성, 준거성, 운용성, 추적가능성, 분류성이 순서적으로 정의되어있다. 다음은 정의된 21개의 품질속성을 정제하고 표준화를 검증하기 위하여 내용분석(Content Analysis)을

수행하였다. 연구자의 분류화로 21개의 품질속성이 도출되었으므로, 협력자의 분류화(Categorization) 단계, 신뢰도 테이블(Reliability Table) 분석을 수행하여 최종 데이터품질 속성을 정의하였다. Table 5와 같이 연구자가 분류한 표준과 협력자가 분류한 표준을 행과 열로 배치시키고 협력자와 합의된 표준을 테이블의 사선에 기입하는 방식을 반복 수행한다[16].

최종으로 협력자와 합의에 의해 표준화된 품질속성을 19

개 도출할 수 있었다. 도출된 품질속성의 신뢰도 검증을 위해서 Table 6과 같이 신뢰도 테이블을 생성하였다. 신뢰도의 측정은 분류화 지수(Categorization Index)로 판단하는데, 일반적으로 분류화 지수(Categorization Index)가 80% 이상이면 적정하고 90% 이상이면 우수한 평가를 받는다[16]. 본 연구에서 데이터품질 속성의 표준화에 대한 분류화 지수는 89%로 우수한 수준의 신뢰도를 확인할 수 있었다.

Table 5. The Result of Reliability Table

Collaborator	가변성	보안성	분석성	상호호환성	시간 효율성	업무수용성	예측성	요구 완전성	운용성	일관성	자원 효율성	최신성	전달성	정확도	준거성	책임성	추적가능성	통제성	활용도
Interviewer																			
가변성	33																		
보안성		11, 63, 64																	
분석성			56, 62																
상호호환성				18, 31, 46															
시간 효율성					19, 24, 26, 30														
업무수용성						12, 57, 4, 49													
예측성			42, 61				27, 38, 60												
요구 완전성								22, 35, 5											
운용성									53										
일관성										14, 7, 21									
자원 효율성											25, 32, 36, 52								
최신성												10, 55, 59							
전달성													15, 47, 48						
정확도														29, 17, 43, 44					
준거성															41, 37				
책임성																39, 50, 20			
추적가능성																	1, 54		
통제성																		2, 6, 9, 13	
활용도			8																16, 23
다양성																			

Table 6. The Result of Computer Categorization Index

Collaborator	가변성	보안성	분석성	상호호환성	시간 효율성	업무수용성	예측성	요구 완전성	운용성	일관성	자원 효율성	최신(최신)성	전달성	정확도	준거성	책임성	추적가능성	통제성	활용도	Agreed Total	Total			
Interviewer																								
가변성	33																				1	1		
보안성		11, 63, 64																				3	3	
분석성			56, 62																			2	2	
상호호환성				18, 31, 46																		3	3	
시간 효율성					19, 24, 26, 30																	4	4	
업무수용성						12, 57, 4, 49																5	5	
예측성			42, 61				27, 38, 60															3	6	
요구 완전성								22, 35, 5														3	3	
운용성									53													1	1	
일관성										14, 7, 21												2	3	
자원 효율성											25, 32, 34, 52											6	6	
최신(최신)성												10, 55, 59										3	4	
전달성													15, 47, 48									3	3	
정확도														29, 17, 43, 44								6	6	
준거성															41, 37							2	2	
책임성																39, 50, 20						3	3	
추적가능성																	1, 54					2	2	
통제성																		2, 6, 9, 13				4	5	
활용도			8																			16, 23	2	3
Agreed Total																								
Total	1	3	2	3	4	5	3	3	1	2	6	3	3	6	2	3	2	4	2			58		
	1	4	5	3	4	5	3	3	2	4	6	3	3	6	2	3	2	4	2				65	

Categorization Index = 58/65 *100 = 89%

2) 요인분석 및 AHP 모형개발

AHP기법을 적용하여 데이터품질 속성 평가를 위한 신뢰성 있는 의사결정 모형을 개발하기 위해 계층모형을 설정하고 의사결정 대안들 간의 쌍대비교 과정을 거쳤다. 계층에 포함된 비교대상은 최대 7±2가지로 제안할 필요가 있다고 Saaty는 제안하고 있다[25]. 따라서 도출된 데이터품질 속성을 의사결정 계층모형으로 구조화하기 위해 항목을 축소하는 요인분석을 수행하였다. RGT결과로 수집된 고객 요구사항 데이터를 분석데이터로 사용하였으며, 요인 추출방법은 주성분분석을 사용하고 회전방법으로는 직각회전의 베리릭스(Varimax)방식을 사용하여 분석하였다. 또한 요인 추출과정은 고유값 기준으로 1보다 큰 요인과 요인적재치 0.5 이상인 변수를 선정하는 방식을 선택하여 SPSS Version 20을 사용하여 분석하였다. 요인분석의 공통성은 요인에 의해 설명되는 비율을 나타내며 0.4 미만이면 상관관계의 설명력이 부족하여 추출 변수에서 제외한다[17]. 그러나 본 연구에서는 Table 7과 같이 실험데이터가 모두 0.4 이상으로 설명력이 우수한 항목들로 나타났다.

또한 설명된 총 분산은 5개의 요인으로 설명 가능하며 설명력은 Table 8과 같이 60.92%로 60% 이상의 설명력을 보여 유효한[17] 것으로 나타났다. 요인분석에서 도출된 결과에 회전성분 행렬을 반영하여 정확도를 높인 결과가 Table 9와 같이 나타났다. Table 9의 요인분석에 의해 도출된 5개 요인성분을 0.5 이상의 성분의 요인으로 Table 9와 같이 그룹핑하여 의미를 유지하면서 요소를 5개로 축소하였다.

Table 7. The Result of Factor Analysis(Commuality)

	Initial	Extraction
Currentness	1.000	.625
TimeEfficiency	1.000	.706
Security	1.000	.598
WorkSoluble	1.000	.718
InterCompatibility	1.000	.717
Changeability	1.000	.413
Interoperability	1.000	.633
Predictability	1.000	.601
Consistency	1.000	.651
Utilization	1.000	.623
Communication	1.000	.474
Accuracy	1.000	.554
ResourceEfficiency	1.000	.711
RequestCompleteness	1.000	.554
Traceability	1.000	.603
Analyticity	1.000	.579
Accountability	1.000	.558
RegulatoryCompliance	1.000	.692
Control	1.000	.565

Table 8. The Result of Factor Analysis(Total Variance Explained)

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Extraction Sum of Squared Rotated		
	Total	% Variance	% Comulative	Total	% Variance	% Comulative	Total	% Variance	% Comulative
1	4.717	24.825	24.825	4.717	24.825	24.825	2.570	13.524	13.524
2	2.344	12.339	37.164	2.344	12.339	37.164	2.523	13.279	26.804
3	1.777	9.350	46.514	1.777	9.350	46.514	2.489	13.102	39.906
4	1.532	8.061	54.575	1.532	8.061	54.575	2.159	11.365	51.272
5	1.207	6.353	60.928	1.207	6.353	60.928	1.835	9.657	60.928
6	.997	5.248	66.176						
7	.934	4.913	71.089						
8	.833	4.382	75.471						
9	.735	3.870	79.341						
10	.661	3.479	82.820						
11	.591	3.113	85.933						
12	.519	2.729	88.662						
13	.441	2.321	90.983						
14	.406	2.136	93.120						
15	.356	1.872	94.992						
16	.306	1.612	96.603						
17	.279	1.468	98.072						
18	.198	1.043	99.114						
19	.168	.886	100.000						

Table 9. The Result of Factor Analysis(Rotated Factor Matrix)

	Factor Value				
	1	2	3	4	5
Consistency	.781	.044	.155	-.109	-.058
Control	.738	.130	-.019	-.042	-.038
RegulatoryCompliance	.651	.083	-.166	.072	.478
Communication	.488	.174	.201	.406	.002
ResourceEfficiency	.122	.786	.251	.061	.109
Interoperability	.090	.726	.186	.194	.160
Security	.004	.694	-.255	-.034	.224
TimeEfficiency	.096	.666	.397	.298	-.080
WorkSoluble	.036	.091	.826	.051	.151
RequestCompleteness	-.101	.163	.659	.075	.279
Utilization	.105	.412	.632	.189	.082
Accuracy	.461	-.111	.557	.126	.050
Predictability	-.203	.107	.136	.728	.026
InterCompatibility	.489	-.032	.035	.670	.162
Currentness	-.370	.147	.250	.629	.093
Changeability	.253	.145	-.050	.521	-.233
Traceability	.040	.090	.152	-.035	.754
Accountability	-.028	.338	.152	-.122	.637
Analyticity	.022	.049	.344	.352	.578

요인분석 결과 그룹핑된 각각의 데이터품질 속성을 분석하여 Table 10과 같이 ‘일관된 체계’, ‘효율적 환경’, ‘정확한 데이터’, ‘유연한 관리’, ‘지속적 개선’으로 명명하고 AHP 분석을 위한 계층모형을 구조화하였다.

Table 10. Hierarchical Factor Using RGT and Factor Analysis

Factor	Factor Define	Second Factor
FA1 일관된 체계 (Consistent systems)	The management system, organization, implementation processes and activities of enterprise aspects for data quality management	일관성(Consistency)
		통제성(Control)
		준거성(Regulatory Compliance)
		전달성(Communication)
FA2 효율적 환경 (Efficient environments)	Activities to improve the quality of data to capable of corresponding, human and material resources in the external environment of the outside intrusion, disorders	자원효율성(Resource Efficiency)
		운용성(Interoperability)
		보안성(Security)
		시간효율성(Time Efficiency)
FA3 정확한 데이터 (Accurate data)	Activities to accommodate the customer's requirement and can work well with high utilization and ensure accurate data	업무수용성(Work Soluble)
		요구완전성(Request Completeness)
		활용도(Utilization)
		정확도(Accuracy)
FA4 유연한 관리 (Flexible management)	Activities flexible adaptable to changes between system such as flow of data or time, organically linked data	예측성(Predictability)
		상호호환성(Inter-Compatibility)
		최신성(Currentness)
		가변성(Changeability)
FA5 지속적 개선 (Continuous improvement)	Continuous management activities that complement and enhance the risk by tracking and analyzing the required feasibility	추적가능성(Traceability)
		책임성(Responsibility)
		분석성(Analyticity)

3) AHP분석 및 가중치 산정

AHP 분석을 위한 RGT와 요인분석으로 도출된 계층화모형은 Fig. 4와 같이 나타났으며, 이를 기반으로 데이터품질 속성별 가중치를 확인하기 위한 항목 간 쌍대비교를 수행하였다.

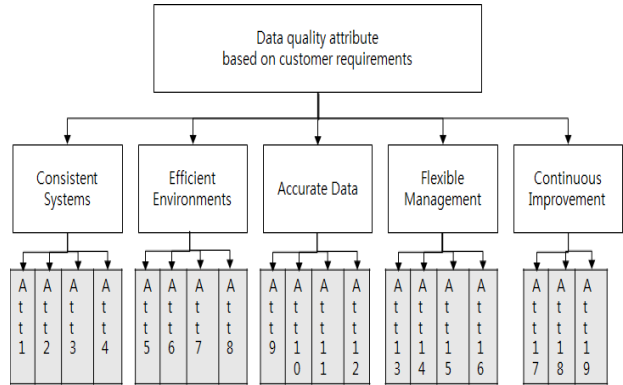


Fig. 4. Hierarchical Structuring for AHP

AHP분석을 위하여 전문가 25명에게 중요도 비교설문을 실시하였으며, 그중 일관성 지수 CI(Consistency Index)가 0.1보다 크게 나온 경우와 미응답을 제외하고 20명의 응답 결과를 연구 자료로 채택하였다. 설문은 평균 17년 이상의 데이터시스템을 활용한 경험이 있는 전문가를 대상으로 실시되었으며, 양쪽속성 동등 1점, 약간 중요 3점, 분명히 중요 5점, 크게 중요 7점, 절대적 중요 9점으로 척도되었다.

AHP설문을 분석한 결과 Table 11과 같이 각 데이터품질 속성의 중요도와 우선순위를 도출하였다.

Table 11. Results of AHP

First DQ Item	Second DQ Item	Rate	Rank
일관된 체계: Consistent systems (0.337, 1)	일관성(Consistency)	0.067	6
	통제성(Control)	0.134	1
	준거성(Regulatory Compliance)	0.100	2
	전달성(Communication)	0.036	11
효율적 환경: Efficient environments (0.146, 3)	자원효율성(Resource Efficiency)	0.046	8
	운용성(Interoperability)	0.030	16
	보안성(Security)	0.038	9
	시간효율성(Time Efficiency)	0.034	14
정확한 데이터: Accurate data (0.282, 2)	업무수용성(Work Soluble)	0.066	7
	요구완전성(Request Completeness)	0.096	3
	활용도(Utilization)	0.036	12
유연한 관리: Flexible management (0.101, 5)	정확도(Accuracy)	0.084	4
	예측성(Predictability)	0.015	19
	상호호환성(Inter-Compatibility)	0.027	17
	최신성(Currentness)	0.037	10
지속적 개선: Continuous improvement (0.134, 4)	가변성(Changeability)	0.021	18
	추적가능성(Traceability)	0.068	5
	책임성(Responsibility)	0.035	13
	분석성(Analyticity)	0.030	15

고객의 요구사항에 기반한 데이터품질 5개 대분류 속성에서는 일관된 체계가 가장 우선순위가 높았으며, 정확한 데이터, 효율적 환경, 지속적 개선, 유연한 관리순으로 중요하다는 결과로 나타났다. 또한 두 번째 데이터품질 세부 19개 속성은 통제성, 준거성, 요구완전성, 정확도, 추적가능성, 일관성, 업무수용성, 자원효율성, 보안성, 최신성, 전달성, 활용도, 책임성, 시간효율성, 분석성, 운용성, 상호호환성, 가변성, 예측성순으로 중요도가 나타났다. 가장 중요하게 고려하는 속성은 일관된 체계의 하부 속성인 통제성과 준거성이었으며, 나머지는 정확한 데이터의 하부속성인 요구완전성, 정확도, 추적가능성 등이 중요하다는 결과로 나타났다. AHP 결과의 상위 우선순위 기준으로 분석해볼 때, 고객이 데이터에서 요구하는 품질은 일관된 체계하에 기준과 고객의 요구사항을 준수하여 업무를 수용할 수 있는 데이터를 중요한 요소로 판단하는 것으로 파악되었다. 그러나 데이터의 지속적이고 유연한 관리를 통해서 시스템 간의 연계 혹은 분석, 외부변화에 대응을 위한 부분은 시급성이 낮다고 판단하여 덜 중요하게 고려하는 것으로 나타났다.

Table 12. Compared with other Data Quality Standards

DQC-M	DQ based on customer requirements		ISO/IEC 25012
	First DQ	Second DQ	
유효성 (Validity)	일관된 체계 (Consistent systems)	일관성(Consistency)	일관성(Consistency)
		통제성(Control)	*
		준거성(Regulatory Compliance)	준수성(Compliance)
	정확한 데이터 (Accurate data)	전달성(Communication)	*
		업무수용성 (Work Soluble)	이해성 (Understandability)
		요구완전성 (Request Completeness)	완전성 (Completeness)
		활용도(Utilization)	접근성(Accessibility)
	정확도(Accuracy)	정밀성(Accuracy), 정확성(Precision)	
활용성 (Practicality)	효율적 환경 (Efficient environments)	자원효율성 (Resource Efficiency)	효율성(Efficiency)
		운용성(Interoperability)	유효성(Availability), 복구성(Recoverability)
		보안성(Security)	보안성(Security)
		시간효율성 (Time Efficiency)	효율성(Efficiency)
	유연한 관리 (Flexible management)	예측성(Predictability)	*
		상호호환성 (Inter-Compatibility)	이식성(Portability)
		최신성(Currentness)	현재성(Currentness)
		가변성(Changeability)	가변성 (Changeability)
지속적 개선 (Continuous improvement)	추적가능성(Traceability)	추적성(Traceability)	
	책임성(Responsibility)	*	
	분석성(Analyticity)	*	
Total DQ			신뢰성(Creditability)

4) 기존 데이터품질 표준과 비교

고객의 인지구조 내에서 중요하게 고려하는 데이터품질 속성이 기존 데이터표준 기관에서 제시하는 데이터품질 속성과 어떠한 차이가 있는지 비교하였다. 먼저, 본 연구는 국내의 데이터베이스진흥원에서 제시하는 DQC-M의 품질관리 가이드를 촉매제로 사용하였으므로 DQC-M과 비교하여 매핑한 결과, Table 12과 같이 유효성(Validity)에는 일관된 체계, 정확한 데이터를 매핑할 수 있었고, 활용성(Practicality)에는 효율적 환경, 유연한 관리, 지속적 개선을 매핑할 수 있었다. 또한 기존의 DQC-M의 세부항목이 정확성, 일관성, 유용성, 접근성, 적시성, 보안성으로만 제시된 기준이었다면, 고객의 요구사항 기반 데이터품질 속성은 좀더 상세한 19개의 품질측정 속성들을 도출할 수 있었다.

이렇게 도출된 19개 속성을 ISO/IEC에서 제시한 데이터 품질 표준 25012와 비교한 결과 대부분의 품질평가 속성과 일치하였다. 다만 Table 12의 * 표시한 통제성, 전달성, 예측성, 책임성, 분석성이 본 연구에서 추가로 도출되었다. 통제성과 전달성, 책임성이 추가로 도출된 이유는 조직에서는 좀더 체계화된 데이터 가버넌스(Data Governance) 측면의 포괄적인 관리가 중요시되고 있기 때문으로 판단된다. 또한 예측성, 분석성이 추가로 도출된 이유는 정형데이터를 넘어서 비정형데이터에서 가치를 이끌어내려는 외부환경과 데이터의 변화에 따른 결과로 판단된다.

4. 결론 및 향후 연구

본 연구에서는 고객의 요구사항에 기반한 데이터품질 평가속성 및 우선순위를 도출하기 위하여, 데이터품질을 20년 이상 처리한 경험이 풍부한 고객을 대상으로 경험치를 이끌어내는 심리학 분석기법인 RGT기법을 적용하였으며, 데이터품질 평가 모델을 구축하기 위하여 RGT기법의 결과와 요인분석 기법을 결합하여 AHP의 계층화모형을 구축하고 AHP분석을 실시하여 가중치와 우선순위를 도출하였다. 또한 도출된 품질평가 속성의 신뢰도를 ISO/IEC 25012와 비교분석하여 전체를 포괄하고 있음을 확인할 수 있었다.

평가속성의 1레벨은 일관된 체계, 효율적 환경, 정확한 데이터, 유연한 관리, 지속적 개선을 도출하였고, 2레벨의 속성으로는 일관성, 통제성, 준거성, 전달성 등 19개 데이터품질 속성을 도출할 수 있었다. 본 연구결과 1레벨에서는 일관된 체계를 가장 중요한 요소로 고려하고 있었고, 정확한 데이터, 효율적 환경, 지속적 개선, 유연한 관리순으로 중요하다는 것을 확인할 수 있다. 2레벨에서는 1레벨의 일관된 체계의 하부 속성인 통제성과 준거성을, 그리고 요구 완전성과 정확도, 추적가능성이 중요한 속성으로 나타났다. 이 결과에서 일관된 체계하에 기준을 준수하고 고객의 요구사항을 수용하여 업무에 적용될 수 있는 데이터가 품질이 높은 데이터로 확인되었다. 구축된 데이터품질 평가속성을 DQC-M 및 ISO/IEC 25012의 품질속성과 매핑하여 비교한 결과, DQC-M

의 속성과의 비교에서는 기존 대비 균일하고 상세한 세부 실천 속성을 얻었음을 확인하였고, ISO/IEC 25012과의 비교에서는 구축된 모델의 신뢰성을 확인함과 동시에 5가지의 추가속성이 도출되었음을 확인하였다. 이는 사용되는 데이터가 정형에서 반정형, 비정형데이터로 변화되면서 축적되는 데이터와 데이터의 가치에 대한 외부환경의 변화에 따른 결과로 판단된다.

따라서 본 연구는 선행연구에서 수행한 표준화기관 속성 간 조합 연구가 아니라, 고객의 심리적 심층인터뷰 기법을 활용하여 요구되는 데이터품질 속성을 도출한 것에 의의가 있다. 또한 이를 통해서 데이터 활용 환경의 변화에 따른 추가 속성을 도출하였고 DQC의 하부속성을 상세화할 수 있었다. 연구방법론 적용 측면에서는 AHP기법의 계층화모형 구성에 RGT와 요인분석을 적용한 연구를 시도하여 고객의 인지구조에 내재된 요구사항을 정량적이고 객관적으로 식별하여 분석하였다는 데 그 의의가 있다.

본 연구에서 구축된 모델은 데이터품질 관리를 위한 전사적인 관리체계 구축 및 가이드 제공에 활용될 수 있으며, 국내의 DQC-M 인증 대비를 위한 상세속성으로 활용될 수 있을 것이다. 또한 품질속성별로 제시된 가중치와 우선순위에 따라서 프로젝트가 보유한 자원을 고려하여 선별적으로 데이터품질 속성을 관리할 수 있을 것이다. 향후 구축된 평가모형을 실제 조직에 적용하고 사례분석을 실시하여 모델을 검증하고 미흡한 부분을 지속적으로 보완하여 완성도를 높일 수 있는 연구가 추가되어야 할 것이다.

References

[1] Korea Database Agency, "Database White Paper," Korea Database Agency, Seoul, 2013.

[2] Richard Y. Wang, Mostapha Ziad, and Yang W. Lee, "Data quality," Kluwer Academic Pub., 2001.

[3] PMI, "Construction Extension to the PMBOK Guide," 3rd ed., Project Management Institute, 2008.

[4] Mi-Young Park and Hyon-Woo Seung, "A Selection Method of Database System Quality Characteristics Using the Analytic Hierarchy Process," *Journal Of The Korean Bibliia Society For Library and Information Science*, Vol.20, No.4, pp.191-204, 2009.

[5] Hye-Jung Jung, "A Study of the Data Quality Evaluation," *Journal of Internet Computing and Services*, Vol.8, No.4, pp.119-128, 2007.

[6] Sunho Kim and Changsoo Lee, "The Process Reference Model for the Data Quality Management Process Assessment," *The Journal of Society for e-Business Studies*, pp.83-105, 2013.

[7] Badri, Masood A., "A combined AHP - GP model for quality control systems," *International Journal of Production*

Economics, Vol.72, No.1, pp.27-40, 2001.

[8] ISO/IEC 25012(NEW), "Software Engineering: Software Product Quality Requirements and Evaluation(SQuaRE) - Data Quality Model," ISO, 2005.

[9] ISO/TS 8000-150, "Data quality-Part 150: Master data: Quality management framework," ISO, 2011.

[10] Korea Database Agency, "DQC-V," Korea Database Agency, Seoul.

[11] Korea Database Agency, "DQC-M," Korea Database Agency, Seoul.

[12] Korea Database Agency, "DQC-S," Korea Database Agency, Seoul.

[13] Korea Database Agency, Korea Database Agency [Internet], <http://www.kdb.or.kr>.

[14] Kelly, G., "The psychology of personal constructs," Oxford: Psychology Press, 1955.

[15] Fransella, F., R. Bell, and D. Bannister, "A manual for repertory grid technique," John Wiley & Sons, 2004.

[16] Devi Jankowicz, "The easy guide to repertory grids," Graduate Business School University of Luton, UK, 2004.

[17] Y. J. Lee, "Understanding Factor Analysis," Seoul: Seokjeong, 2002.

[18] Saaty, T. L., "Decision-making with the AHP: Why is the Principal Eigenvector Necessary," *European Journal of Operational Research*, Vol.145, No.1, pp.85-91, 2003.

[19] K. T. Cho, Y. G. Cho and H. S. Kang, "The Analytic Hierarchy Process," Seoul: Dong hyun, 2003.

[20] Seung-Hee Kim and Woo-Je Kim, "Study on the Selection Model CTQ data," *Journal of Korea Society of Computer Information*, Vol.18, No.4, pp.97-112, 2013.

[21] Hwanju Cha and Ja-Hee Kim, "A evaluation model of Transition PMO Competencies using RGT and AHP," *Korea Society of IT Services*, Vol.13, No.2, pp.87-109, 2014.

[22] Soon-yeong Kim, "A Study on the Development of Meta-Evaluation Indicators for Defense R&D Programs by Using FA/AHP Methods," *Korea Technology Innovation Society*, Vol.12, No.1, pp.113-136, 2009.

[23] HyungJun Lee, "Development of Performance Evaluation Indicators of Defense Core-Technology R&D Projects by SMR-based AHP," Seoul National University of Technology, 2010.

[24] Brombacher, Aarnout, et al., "Improving product quality and reliability with customer experience data," *Quality and Reliability Engineering International*, Vol.28, No.8, pp.873-886, 2012.

[25] Wind, Yoram, and Thomas L. Saaty, "Marketing applications of the analytic hierarchy process," *Management Science*, Vol.26, No.7, pp.641-658, 1980.



장 경 애

e-mail : jkalove@hanmail.net
1996년 대구대학교 문헌정보학과(학사)
2014년 연세대학교 컴퓨터공학과(석사)
2014년~현 재 서울과학기술대학교 IT정
책전문대학원 산업정보시스템전공
박사과정

관심분야: 데이터 품질/분석, 최적화 등



김 자 희

e-mail : jahee@seoultech.ac.kr
1995년 KAIST 전산학과(학사)
1997년 KAIST 전산학과(석사)
2003년 KAIST 산업공학과(박사)
2004년 University of Vienna Visiting
Scholar

2005년~현 재 서울과학기술대학교 IT정책전문대학원
산업정보시스템전공 부교수

관심분야: 요구 공학, 소프트웨어 공학, 시뮬레이션, 반도체
스케줄링



김 우 제

e-mail : wjkim@seoultech.ac.kr
1986년 서울대학교 산업공학과(학사)
1988년 서울대학교 산업공학과(석사)
1994년 서울대학교 산업공학과(박사)
1988년~1991년 동양경제연구소 연구원
1999년~2001년 University of Michigan
Visiting Scholar

2003년~현 재 서울과학기술대학교 글로벌융합산업공학과 교수
관심분야: IT서비스, 소프트웨어 공학, 최적화, 스마트그리드 등