

# 데이터 품질평가를 위한 수학적 모델 및 개선을 위한 분석 모형 개발

## The Development of a Mathematical model to evaluate Data Quality and an Analysis model to improve the Quality

김 용 원\*  
Yoeng-Won, Kim

김 중 기\*\*  
Jong-Ki, Kim

### 요 약

컴퓨터와 인터넷 환경의 급속한 변화는 방대한 데이터와 다양한 품질의 데이터를 생산해내고 있다. 이러한 환경의 변화는 데이터의 품질에 따른 수준 평가를 요구한다. 이는 데이터 품질이 기업과 조직에 미치는 영향이 크기 때문이다. 따라서 본 연구는 데이터 품질 기준에 따른 품질 평가 모델을 수학적 모델로 제시하고, 이를 기반으로 데이터의 품질을 분석하고 데이터의 품질을 유지 보완하기 위한 분석 모형을 제안하고자한다.

### Abstract

The rapid change of computer and Internet environments produces a lot of data of various quality. Because this fact affects enterprise and organization, it demands the level evaluation on data quality. Thus, we propose mathematical model for quality evaluation on the base of data quality in this paper. And we propose the analysis model(web evaluation model, DQnA) that analyzes and maintains data quality.

☞ keyword : DQnA, web evaluation model, 데이터 품질, 품질 평가, 데이터 품질 분석, 품질 평가 모델

## 1. 서론

최근 사용자들은 유무선의 네트워크나 인터넷을 통하여 방대하고 다양한 데이터를 접하게 된다. 그러나 사용자들은 그러한 데이터에 대한 신뢰도가 떨어지게 되고, 이는 조직과 기관, 그리고 사회에 부정적 영향을 끼치게 된다. 저 품질의 데이터는 기업과 기관에 막대한 손실을 초래한다. 그러나 현실에서 많은 기업과 기관에서는 데이터 품질 평가를 위한 시스템을 갖추지 못하고 있으며, 데이터 품질로 발생하는 문제와 손실을 인지

하지 못하고 있다[3, 13].

그러나 데이터의 품질(quality)을 평가하고 유지하기가 쉽지 않으며, 이에 유연성 있는 품질 평가 모델의 개발과 개선을 위한 평가 모형의 개발, 즉 데이터 품질을 관리하는 프로세스가 필요하다.

데이터의 품질은 조직의 전략과 목적을 달성하기 위해 필요로 하는 데이터 요구자, 즉 의사결정자 및 관계자의 만족도를 충족시킬 수 있는 수준을 의미하는 것이다. 데이터 품질의 높은 수준을 유지하고 개선하기 위한 데이터의 품질평가는 측정(measure)과 분석(analyze), 그리고 개선(improve)이 반복적으로 이루어져야 한다[8].

데이터 품질의 평가를 위해 필요한 기준은 사용자 요구(needs), 즉 데이터를 사용하는 업무, 사용하는 사람의 목적, 그리고 주관적 판단기준과

\* 정 회 원 : 건양대학교 기업정보관리학과 교수  
ywkim@konyang.ac.kr(제1저자)

\*\* 정 회 원 : 건양대학교 정보보호학과 교수  
jkkim@konyang.ac.kr(교신저자)

[2008/06/18 투고 - 2008/06/19 심사 - 2008/07/04 심사완료]

사용규칙에 따라 달라진다. 따라서 데이터 품질은 주관적인 평가로 이루어지며, 이러한 평가는 품질 향상을 위하여 다면적, 반복적, 그리고 지속적 평가로 이루어진다[11].

본 논문에서는 이러한 평가 환경에서 데이터 품질의 지속적인 평가를 위한 수학적 평가 모델을 제안하고, 분석을 위한 평가 모형을 개발하고자한다.

## 2. 국내외 연구동향

데이터의 품질평가 모델은 초기 현상 분석 중심에서 점차 구조적 분석 중심 모델로 변화해 왔다. 이와 함께 최근에는 지속적인 품질 개선 요인의 분석을 통하여 종합적이고, 다면적인 평가를 위한 통합모델로의 연구가 진행되고 있다[3,7,10].

### 2.1 해외 연구동향

#### 2.1.1 현상 개선 중심모델

현상적 데이터 품질 요인 발견에 중점을 두는 모델이다. 1980년대 말 시작된 모델링 방법으로 몇 가지 연구내용에 관하여 간략히 알아본다.

##### 1) Granick의 DB 평가기준

온라인 데이터베이스 사용자 그룹에서 채택된 데이터베이스 평가를 위한 10가지 기준을 다음과 같이 ①일관성 ②수록범위 ③최신성 ④정확성 ⑤이용자 용이성 ⑥통합성 ⑦사용자 지원성 ⑧출력 ⑨내용 ⑩비용 대 가치로 제시하였다. 마지막 카테고리인 비용 대 가치(value-to-cost ratio)란 데이터베이스가 창출하는 가치와 데이터베이스 가격(또는 사용료)과의 비율을 의미하는 개념이다. 그러나 가치와 품질은 서로 다른 의미를 지니고 있으며, 실제적으로 그 가치를 어떻게 평가할 것인가가 문제로 남는다.

##### 2) SCOUG의 평가법

SCOUG 평가법은 서지 데이터나 전문 데이터베이스를 질적인 측면에서 평가할 수 있는 10가지의 측정 기준을 다음과 같이 ①일관성 ②사용용이성 ③범위 ④최신성 ⑤정확성 ⑥통합성 ⑦출력 ⑧도큐멘테이션 ⑨사용자지원성 ⑩비용 비 가치로 제시하였다. 이러한 CD-ROM 데이터베이스 및 온라인데이터베이스의 기준 역시 전통적인 색인 평가 기준의 영향을 받은 것이다.

##### 3) Large의 품질평가

Large는 데이터베이스를 평가하기 위하여 다음과 같이 ①범위와 권위 ②현행성 ③정확성 ④유일성 등의 4가지로 평가 기준을 제시했다. 데이터베이스 품질 중에서도 서비스 부분을 배제하고 데이터 부분에 초점을 맞추고 있다.

##### 4) Herther CD-ROM 평가

Herther는 CD-ROM 평가에 사용되는 평가항목을 8가지로 제시하고 있다. ①정확성 ②접근성 ③사용용이성 ④권위 ⑤내용 ⑥사용자지원성 ⑦최신성 ⑧가치를 평가항목으로 제시하였다. 이 평가 모델은 데이터베이스에 대한 이론적 기반도, 용어의 정의도, 관점간의 구분 근거도 명료하게 나타내고 있지 않다.

##### 5) Doll & Torkzadeh

최종 사용자 환경의 사용자 만족도에 대한 연구에서 요인 분석(factor analyze)을 통해 ①내용 ②정확성 ③형식 ④사용용이성 ⑤신속성 등 다섯 가지 요소(components)를 추출하였다. 이 연구는 수백의 샘플 데이터 분석을 통해 비교적 탄탄한 통계적 근거를 제시하고 있다. 그러나 데이터베이스 데이터와 데이터 서비스 개념을 혼동하는 등의 논리적 취약함을 보이고 있다.

### 2.1.2 구조적 분석 중심 모델

1990년대 중반 이후 제기된 연구 모델로 구조적 데이터 품질 요인의 발견에 중점을 두고 연구된 모델이다.

#### 1) Greg R. Notes의 DB 평가 기준

기존의 데이터베이스 품질 평가 기준을 보완한 데이터 평가 기준을 제시했다. 3가지의 큰 범주의 평가 기준으로 ①데이터베이스의 범위 ②데이터베이스의 구조(레코드 구조) ③레코드의 정확성(cleanliness)을 제시했다. 특히 기존의 연구에서 보다는 진전된 데이터의 구조적 측면을 고려한 평가 기준을 제시함으로써 보완된 기준을 제시하고 있다.

#### 2) Brodie의 데이터 품질평가

Brodie는 데이터베이스 품질 중에서 데이터 부분에 주안점을 두고 있으며, 다음과 같이 ①데이터 신뢰성 ②논리적 무결성 ③물리적 무결성으로 구분하여 제시하였다.

데이터 신뢰성은 데이터베이스가 표현하고자하는 실세계의 구조적 속성을 제대로 표현하는 정도로 정의된다. 그리고 논리적 무결성은 사용자에게 인식되는 개념적 데이터의 일관성 있고 정확함을 의미하고, 물리적 무결성은 컴퓨터에 물리적으로 저장된 상태로써의 품질을 의미한다. 이러한 구분은 데이터베이스가 만들어지고(물리적), 제공되며(논리적), 사용자에게 활용되는(신뢰성) 과정의 단계별 구분에 근거하고 있다. 신뢰성 또는 무결성이라는 용어가 개념적으로는 널리 사용되지만 실제 측정이나 평가에 사용하기 어렵다는 것이 문제점이다.

### 2.1.3 지속적 개선을 위한 통합 모델

최근에 진행되고 있는 모델로 총체적인 데이터 품질 요인의 발견과 정의를 통하여 품질 개선을

위한 절차를 중시하여 이에 집중하는 연구 모델이다[2,12,13].

#### 1) Larry P. English의 TIQM

TIQM(Total Information Quality Model)은 품질관리를 위한 프로세스를 6단계로 정의하고 있다. 각 단계는 데이터 평가 단계(P1-P3), 데이터 유지보수를 위한 단계(P4), 데이터 프로세스 개선을 위한 단계(P5), 데이터 품질 환경의 개선을 위한 전략 수립 단계(P6)로 구성되어 있다. 이 모델은 데이터의 품질 평가 뿐 아니라 품질 개선을 위한 프로세스를 중시하고자하는 모델이다.

#### 2) MIT의 TDQM

TDQM(Total Data Quality Model)은 2000년 MIT와 정부, 그리고 기업이 협동으로 연구를 수행하였다. 데이터의 품질을 정의하고 데이터 품질이 기업에 미치는 영향과 데이터 품질 개선에 관한 3가지 주제를 중심으로 연구가 진행되었다. TDQM은 정보의 품질을 결정하는 요인을 20가지로 정의하고 있다[2,9].

## 2.2 국내의 연구동향

국내에서는 데이터 품질의 중요성에 대한 인식이 많이 부족했던 것이 사실이다. 최근 10여 년 동안 데이터의 품질이 조직과 기업에 미치는 막대한 손실과 영향에 대한 이해가 넓어지면서 연구가 활발히 진행되고 있다.

국내의 많은 기업들이 데이터웨어하우스를 구축하고, CRM을 통한 고객관계 데이터의 중요성을 인식하면서 데이터 품질의 유지와 지속적인 개선을 위한 노력이 진행되고 있다.

품질평가 모델의 특성에 따라 모델을 분류해 보면 표 1과 같다[1,2,4,5,6,7].

(표 1) 국내 품질 평가모델의 특성별 분류

<b>(1) 현상 개선 중심 모델</b>	
① 한국웹사이트 평가 개발원의 DB품질 평가모델	
② 안계성의 DB품질 평가항목	
③ 이제환의 분산체제로 구축된 DB품질	검증
④ 박준식의 색인평가 기준에 관한 분석	
⑤ DPC의 데이터베이스 표준화연구	
<b>(2) 구조적 분석 중심 모델</b>	
① 정보시스템 유형별 품질측정 모형 개발	
② 분산 환경을 위 중보 데이터 구조 연구	
<b>(3) 지속적 개선을 위한 통합 모델</b>	
① DPC의 DB품질 관리 확장 모델	
② 박주석/박찬수의 데이터 성숙도 모델	
③ 국방 데이터 표준화 연구	

본 연구에서는 데이터의 지속적인 유지와 개선을 위한 평가 모형을 수학적 평가 모델을 기반으로 제안함으로써 데이터의 다면적 특성을 반영하고, 사용자의 뷰에 따른 관리 프로세스가 진행될 수 있도록 한다.

### 3. 데이터 품질 평가를 위한 모델링

#### 3.1 데이터 품질 평가 요소

본 절에서는 다양한 환경에서 필요로 하는 품질 평가 기준을 분석하고, 그 결과를 기초로 분석 기준을 정의한다. 데이터의 품질 평가를 위한 기준 영역을 크게 데이터 영역, 데이터 서비스 영역, 정보 효율도 만족도 영역, 그리고 정보의 안정성 영역 등 4개 영역으로 구분한다[1,4].

(표 2) 데이터 품질 평가 영역별 평가 요소

영역(Wi)	평가 요소(WSo)
데이터 영역	정확성, 현재성, 포괄성, 일관성, 적합성, 유일성, 객관성
데이터 서비스 영역	접근성, 검색성, 표현성, 비용, 시간의 범위
정보 효율도 만족도 영역	정보의 적합도, 정보의 효율도, 정보의 비용대비 만족도, 정보 제공방법 충족도
정보의 안정성 영역	보안성, 웹(네트워크)의 안정성, 접근 인터페이스의 안정성, 하드웨어적 요소의 안정성

표 2에서는 4개의 각 영역에서 평가될 수 있는 평가 요소를 정의하여 그 기준으로 제시하고 있다. 데이터 영역에서는 데이터의 활용 가치를 평가할 수 있는 요소로 구성되며, 데이터 서비스 영역의 평가 품질은 어떤 형태(any presentation), 어떤 시간(any time), 어떤 장소(any where)에서도 서비스가 제공될 수 있는가가 평가 척도라 할 수 있다. 정보 효율도 만족도 영역에서는 사용자의 정보 활용 만족도 품질을 평가하는 항목으로 구성된다. 마지막으로 정보의 안정성 영역에서는 제공되는 정보의 안정적 제공 능력을 평가하는 영역이다.

표 2는 데이터 품질 평가를 위한 기준을 4개 영역으로 정의하여 각 영역별 평가 항목을 정리하여 보여주고 있다. 각 영역에서 정의된 품질 평가 변수들의 품질 기준을 평가함으로써 데이터의 품질이 정의되고 이에 따른 유지 및 개선방안이 연구될 수 있다. 이러한 변수들의 평가는 사용자의 다양한 측면, 즉 다면적 평가를 통하여 이루어지는 것이다. 따라서 평가 결과는 사용자의 성향, 데이터 사용 환경, 그리고 데이터의 사용 목적 등에 따라 다양한 평가가 이루어진다.

각 영역에서 정의된 평가 변수들은 각 변수를 평가할 수 있는 평가 세부 기준을 정의하여 평가할 수 있으며, 평가 기준은 데이터의 평가 환경과 사용자 다면적 측면을 고려하여 정의한다. 표 2는 평가 영역별 세부 항목과 세부 평가기준의 예이다 [1].

#### 3.2 품질 평가를 위한 수학적 모델링

데이터 품질의 평가를 위한 수학적 접근법을 제시하고자한다. 수학적 모델링을 통한 품질의 정량화함으로써 일관된 평가 시스템을 유지할 수 있으며, 이를 통하여 데이터의 많은 평가 요소들의 강점과 약점을 도식화하여 제시할 수 있다. 이는 데이터의 유지와 개선 전략을 정의하는데 활용된다.

평가 시스템은 다음의  $\langle Q, W, S, \Phi, F \rangle$ 로 구성된다.

- 1)  $W$ 는 각 영역의 품질 평가 기준으로 계산되는 영역의 가중치 집합이다.

$$W = (w_1, w_2, \dots, w_n)$$

집합  $W$ 의 원소  $w_i$ 는 정의되는 영역의 수  $n$ 의 값에 따라 0 부터  $1/n$  사이의 값을 갖는다.

- 2)  $Q$ 는 데이터의 품질을 요소들의 평가를 통하여 계산되는 전사적 품질에 대한 가중치를 말한다.  $Q$ 는 함수  $F$ 에 의해 0과 1 사이의 값으로 사상된다.

$$F(w_1, w_2, \dots, w_n) = Q,$$

$$Q = \sum_{i=1}^k w_i, \quad 1 \leq k \leq n$$

- 3)  $S$ 는 각 영역별 세부 평가 요소로 구성되는 평가 항목들의 집합으로 정의하고,

$S = (s_1, s_2, \dots, s_m)$ ,  
 각 세부 항목  $s_i$ 의 가중치를  $w_{s_i}$ 로 표현하면, 영역의 가중치  $w_i$ 는

$$w_i = w_{s_1} + w_{s_2} + \dots + w_{s_m}$$

이다.

$S$ 에 속하는 세부 평가기준  $a_i$ 의 가중치를  $w_{a_i}$ 라 할 때,

$s_i = \{(a_1, w_{a_1}), (a_2, w_{a_2}), \dots, (a_l, w_l)\}$ 와 같이 표현하고,

$w_{a_i}$ 는 0에서 1 사이의 값으로 정의한다.

- 4)  $\Phi$ 는 세부 평가 항목( $S$ )을 평가할 수 있는

세부 기준들의 집합으로 데이터의 평가 환경에 따라 다양하게 정의될 수 있다.

$$\Phi = (a_1, a_2, \dots, a_l),$$

$$l = 1, 2, 3, \dots$$

- 5)  $F$ 는 데이터의 품질을 결정하는 데이터 품질 평가 함수이다. 함수  $F$ 는 다음과 같이 정의된다.

$$F : W \rightarrow [0, 1]$$

함수  $F$ 는 영역별 가중치에 따라 품질을 평가하여 데이터의 전사적 품질을 0과 1사이의 값으로 반환한다. 이는 데이터의 품질 수준을 총체적으로 평가하는 값으로 향후 유지 및 개선 계획 수립, 운용할 수 있는 중요한 요소가 된다. 즉, 현재의 품질 수준을 의미한다.

품질 평가를 위하여 본 연구에서 정리하여 제시하고 있는 4개 영역, 데이터 영역, 데이터 서비스 영역, 정보의 활용 만족도 영역, 정보의 안정성 영역은 각각,  $w_i$ 로 영역별 가중치를 정의한다.

$$w_i = \left( \sum_{j=1}^m w_{s_j} \right) / m, \quad 1 \leq i \leq n$$

$$w_{s_j} = \left( \sum_{l=1}^l w_{a_l} \right) / l$$

이때,  $m$ 는 영역별 평가 요소들의 집합  $S$ 의 원소 수를 의미하며,  $l$ 은 각 평가 항목별 평가 기준인 집합  $\Phi$ 의 구성 원소의 수로 정의한다. 각 영역의 가중치를 의미하는  $w_i$ 의 범위는

$$0 \leq w_i \leq \frac{1}{n}, \quad n \text{은 영역의 수로 한다. 한편,}$$

$w_i$ 의  $i$ 값은 본 연구에서는 영역을 4개의 영역으로 분류하고 있으므로 4이다. 따라서  $w_i$ 와  $Q$ 는

$$w_1 = (w_{s1} + w_{s2} + w_{s3})/3,$$

$$w_{s1} = (\sum_{i=1}^6 w_{a_i})/6,$$

$$w_{s2} = (\sum_{i=1}^2 w_{a_i})/2,$$

$$w_{s3} = (\sum_{i=1}^3 w_{a_i})/3$$

$$\delta_{w_{s_i}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (w_{s_i} - \overline{w_{s_i}})^2}$$

$$\delta_{w_{a_i}} = \sqrt{\frac{1}{l} \sum_{i=1}^l (w_{a_i} - \overline{w_{a_i}})^2}$$

$Q = (w_1 + w_2 + w_3 + w_4) / 4$ 이다.

각 영역의 평가항목을 평가하기 위하여 제시된 평가 기준 집합  $\Phi$ 의 항목  $a_i$ 의 값은 조사를 위하여 제시되는 선택 항목들의 평가 척도를 5로 정의한다면 가중치를 (0.2 0.4 0.6 0.8 1.0)으로 부여한다. 이때 선택 항목의 평가 가중치는 0과 1의 범위 내에서  $\pm 0.1$ 의 값을 부여하여 평가할 수 있도록 한다. 이는 평가 결과에 대한 표준 편차를 줄이고 평가의 변위를 유연하게 제시함으로써 현실을 정확하게 반영하고자 한다.

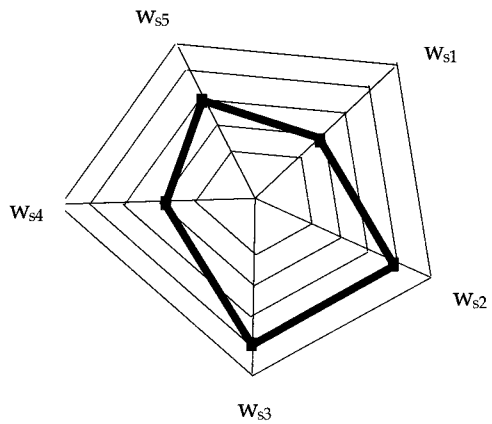
#### 4. 품질 평가 분석 모형과 해석

품질 평가 모델로써 제시된 평가 모델에 의한 평가 결과를 이용하여 영역별, 평가 기준별 품질을 분석하는 유형에 관하여 알아보자. 먼저 각 집합  $S, \Phi$ 의 품질 평가를 위한 방법으로 구성하는 항목간의 표준 편차를 이용한다.

집합  $\Phi$ 를 구성하는 평가 기준 항목들의 평가 값들에 대한 평균의 반영 가중치를 표준편차 범위 내에서 허용하도록 함으로써 특정 기준  $a_i$ 의 가중치가 전체 가중치에 미치는 영향을 판단하게 한다. 이는 세부 평가 항목  $s_i$ 의 개선을 요구할 수 있도록 한다.

$w_{s_i}$ 와  $w_{a_i}$ 의 표준 편차를 정의하면 다음과 같다.

Q를 구성하는 값 영역별 가중치  $w_i$ 를 이용하여 분석 모형을 구성하면 (그림 1)과 같은 web 평가 모형이다. web 평가 모형은 영역별 품질 평가 결과를 평가항목을 이용하여 관계를 표현함으로써 정량적으로 평가된 데이터의 품질 수준을 통하여 사후 관리 시스템의 평가와 관리 방법, 그리고 개선방법을 모색할 수 있도록 한다.

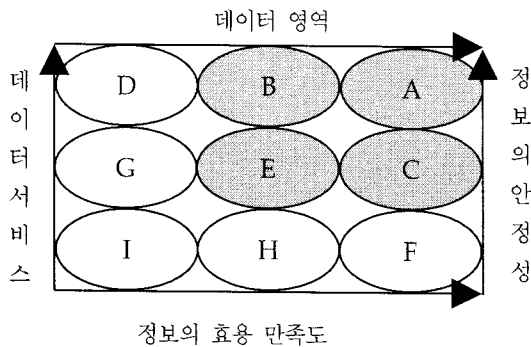


(그림 1) 품질 수준의 웹 평가 모형

그림 1은 웹 평가 모형을 이용하여 영역별 세부 평가 구성 항목 간의 관계를 도식적으로 표현한다. 특정 영역의 평가 세부 항목 ( $w_{s1}, w_{s2}, w_{s3}, w_{s4}, w_{s5}$ ) 간의 수준을 거미줄 모형으로 표현하여 보여줌으로써 항목별 품질 수준의 균형을 판단할 수 있으며, 유지 개선 분야를 도식화하고 있다. 그림 1은 5가지 항목으로 구성된 특정 영역의 세부 항목 균형을 나타내고 있다. 이것은 각 항목의 세부 기준이 다르기 때문에 데이터의 품질에 영향을 미치는 세부

기준을 파악하고, 이에 따른 프로세스를 결정하게 한다.

웹 평가 모형에 표현된 다각형은 가장 큰 정다각형 모양을 만들 때 최상위 품질의 데이터 수준을 유지하는 것이다. 이 모양은 사용자의 뷰와 환경의 변화에 따라 영향을 받으며, 따라서 고급 수준의 품질을 항상 유지하기 위한 지속적인 평가 프로세스가 반복적으로 진행되어야 한다.



(그림 2) 전사적 품질평가 지수모형(DQ4A모형)

그림 2는 데이터 품질을 전사적으로 평가하는 Q를 구성하는 영역 품질 지수에 따라 해석하고, 현재의 데이터 품질을 확인할 수 있다. DQnA(Data Quality n Area)의 하나인 DQ4A(n=4) 모형은 모두 9가지 영역으로 품질을 분석하고 있다. 그림 2의 음영 부분의 평가 지수는 모든 영역에서 평균 이상의 품질을 유지하고 있는 부분으로 해석할 수 있으며, 영역별 기대치를 평가하여 수정 개선에 활용할 수 있다. DQ4A 모형은 데이터 품질이 영역별로 어느 위치를 유지하고 있는지를 정량적 지표를 통하여 도식적으로 보여준다. 따라서 9개의 평가 지수 영역에 따라 어떤 관점에서 데이터의 품질 수준을 어느 영역에서 최적화해야 할 것인지를 판단할 수 있게 한다. DQnA 모형은 품질을 유지하고, 지속적인 개선을 위한 최적화 전략을 현재의 관점에서 평가할 수 있도록 하고 있다. 또한 미래 환경의 변화를 예측하고, 이를 반영함으로써 상위

수준의 품질을 꾸준히 유지할 수 있도록 한다.

## 5. 결론

다변화되어가는 정보통신 환경에서 데이터의 품질은 그 평가 기준이 엄격해야 하며, 평가 기준도 환경에 맞게 변화되어야 한다. 또한 변화하는 평가 기준 뿐 만아니라 이에 따른 평가 방법도 유연하고 적응적인 평가 모델을 필요로 한다.

따라서 본 연구는 다양한 평가 항목의 변화와 사용자의 다면적 특성을 고려한 평가 모델을 수학적으로 제시한다. 한편 수학적 도구를 이용한 평가 결과에 따른 분석 모형을 함께 제시함으로써 현재의 환경에 따른 데이터 품질을 평가하고, 미래의 변화에 유연하고 적극적으로 대처하도록 한다.

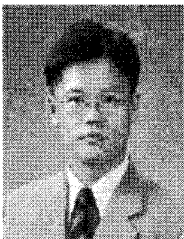
따라서 이러한 다양하고 다면적인 환경에 유연하게 적응하여 평가할 수 있는 모델을 필요로 하게 되었다, 본 연구가 이를 제시하고, 이를 분석하기 위하여 평가 기준 ( $\Phi$ ), 각 기준에 따른 영역별 항목(S)과 각 영역(W)에서 관계를 통하여 정량적으로 해석하고 분석할 수 있는 웹 평가 모형과 평가 영역들의 관계를 9가지 부분으로 구분하여 해석할 수 있는 DQnA 모형을 제안하였다. 본 연구는 데이터 품질의 정량적 평가와 이를 이용한 분석 도구를 제시함으로써 다양한 환경과 변화에 적응적으로 대처하고 미래의 변화를 예측할 수 있도록 하고자한다.

## 참고 문헌

- [1] 김용원, “데이터베이스 품질 평가를 위한 모델 개발에 관한 연구”, 건양논총, 15권, pp.51-65, 2006.
- [2] 김찬수, 박주석, “데이터 품질관리 성숙 모델에 관한 연구”, 한국정보과학회, 20권, pp.249-275, 2003.

- [3] 나관상, 백두권, “단계적 데이터 품질 모델링 방법론과 스키마”, 한국정보과학회: 학술대회지, pp.118-120, 2001.
- [4] 박미영, 승현우, “CMM을 적용한 데이터베이스 품질평가 항목 개발에 관한 연구”, 한국정보처리학회:학술대회지, pp.153-161, 2003.
- [5] 이용봉, “CD-ROM 데이터베이스의 품질 평가 및 개선 방안에 관한 연구”, 한국문헌정보학회지, 33(4), pp.29-46, 1999.
- [6] 이제환, “공동목록 DB의 품질평가와 품질관리 : KERIS의 종합 목록 DB를 중심으로”, 한국문헌정보학회지, 36(1), pp.61-90, 2002.
- [7] 장혜란, “데이터베이스 품질평가를 위한 모형 개발-텍스트 데이터베이스 내용을 중심으로”, 한국정보관리학회지, 17(4), pp.83-97, 2000.
- [8] 정혜정, “데이터 품질평가에 관한 연구”, 한국인터넷정보학회 논문지, 8(4), pp.119-128, 2007
- [9] 함정훈, 안계성, “데이터베이스 품질 평가 항목” 데이터베이스 표준화연구, 한국데이터베이스진흥센터, 2000.
- [10] 홍현진, “웹 기반 데이터베이스의 품질평가 기준 개발에 관한 연구”, 한국문헌정보학회지, 39(2), pp.211-235, 2005.
- [11] Howard, P., “Data Quality Products :an evaluation and comparison,” Bloor Research, 2004.
- [12] Kahn, B., Strong, D., and Wang, R., “Information Quality Benchmarks : Product and Service Performance,” Communications of the ACM, Vol.45, Apr, 2003.
- [13] <http://dbq.dpc.or.kr/assessment/research.php>

## ○ 저 자 소 개 ○



### 김 용 원 (Yoeng-Won, Kim)

1986년 고려대학교 수학과 졸업(학사)  
 1988년 고려대학교 대학원 수학과 졸업(석사)  
 1995년 고려대학교 대학원 전산학과 졸업(박사)  
 1995~현재 건양대학교 기업정보관리학과 교수  
 관심분야 : 데이터베이스 모델링, 데이터 검색, 데이터베이스 품질, 인터넷 데이터베이스  
 etc.

E-mail : ywkim@konyang.ac.kr



### 김 종 기 (Jong-Ki Kim)

1977년 인하대학교 수학과 졸업(학사)  
 1982년 인하대학교 대학원 수학과 졸업(석사)  
 1992년 인하대학교 대학원 수학과 졸업(박사)  
 1993~현재 건양대학교 정보보호학과 교수  
 관심분야 : 데이터베이스, 센서 네트워크, 보안, 정보보호 및 암호 etc.

E-mail : jkkim@konyang.ac.kr