
데이터 품질 분석 모델(DQnA)을 이용한 융합적·적응적 품질 분석에 관한 연구

김용원
건양대학교 정보보호학과

A study on Convergent & Adaptive Quality Analysis using DQnA model

yong-won kim
Division of information security, Konyang university

요약 현재 대부분의 기업들이 정보기술을 기반으로 정보 시스템을 이용한 데이터 분석 기법을 활용하고 있다. 이러한 데이터 분석은 기업의 다양한 의사결정에 영향을 미치는 데이터의 품질 평가에 주목하고 있다. 이는 데이터 품질 평가가 기업의 효과적인 운영뿐만 아니라 여러 부분에서 중요한 역할을 하기 때문이다. 본 연구에서는 현재 다양하게 연구되고 있는 데이터 품질 평가 모델에 관하여 기술하고, 이를 기반으로 데이터 품질 분석에 활용되고 있는 융합적이며, 적응적 모델인 DQnA 모델에 관하여 서술하고, 이를 활용한 품질 분석 방법에 관하여 논의하고자 한다.

• **Key Words** : 데이터 품질, 평가 모형, 적응적 모델

Abstract Now, almost enterprise is applying data analysis method using the information systems on based information technology. The data analysis is focusing on the Quality of the data affecting the decision-making of various companies. This is the result of the data quality is due to the important role in the various parts as well as the effective operation of the enterprise. In this study, we describe about the data quality assessment models that are currently being studied. Based on this, we describe about the adaptive DQnA model being utilized for data quality analysis, and discuss about the quality analysis using this method

• **Key Words** : data quality, evaluation model, adaptive model

1. 서론

2000년대 정보화 이후 많은 정보가 다양한 정보 시스템을 통하여 쏟아져 나오고 있는 상황에서 그러한 정보들에 대한 신뢰성과 정확성, 그리고 필요한 용도에 적극적으로 활용될 수 있는 지 등에 대한 많은 의문이 야기되기 시작하였다. 이러한 정보의 품질은 축적되고 보관되

어 제공되고 있는 많은 데이터들의 품질에 많은 영향을 받게 된다. 이러한 데이터들이 높은 품질을 유지하며 제공될 수 있는가의 문제는 우리가 필요로 하는 정보가 얼마나 높은 양질의 정보로 제공되어 질 수 있는가의 문제로 귀결될 수 있다.

그러나 데이터의 품질(quality)을 평가하고 유지하기가 쉽지 않으며, 이에 유연성 있는 품질 평가 모델의 개

이 논문은 2014학년도 건양대학교 학술연구비 지원에 의하여 이루어진 것임.

*교신저자: 김용원(ywkim@konyang.ac.kr)

접수일 2014년 9월 12일 수정일 2014년 11월 26일 게재확정일 2014년 12월 21일

발과 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 [1,13,14].

따라서 데이터 품질의 평가를 위해 필요한 기준은 사용자 요구(needs), 즉 데이터를 사용하는 업무, 사용하는 사람의 목적, 그리고 주관적 판단기준과 사용규칙에 따라 달라진다. 따라서 데이터 품질은 주관적인 평가로 이루어지며, 이러한 평가는 품질 향상을 위하여 다면적, 반복적, 그리고 지속적 평가로 이루어진다[2,15].

또한 데이터 품질은 데이터를 사용하고자 하는 사용자, 데이터를 제공하는 정보시스템, 그리고 정보제공 환경이라는 중요한 요인들의 융합적 결합 요소에 따라 그 품질의 평가 결과는 달라질 수 있다.

본 논문에서는 지금까지 연구해온 데이터 품질 모델 기술에 관하여 그 이론적 고찰과 선행 연구에 관하여 연구를 수행한다. 또한 최근 제시되고 있는 수학적 접근 모델을 통한 품질 평가 방법에 관하여 알아보고, 이를 기반으로 하는 데이터 품질 분석 모델을 활용한 기업의 데이터 품질 분석 방법에 관하여 논의하고자 한다.

2. 이론적 고찰 및 선행연구

2.1 데이터 품질의 정의

데이터란 컴퓨터에 저장된 형태의 의사결정을 위한 근거로써 사용되는 사실(facts)이나 값(value)을 의미하며, 품질이란 어떤 것의 등급 혹은 일반적인 기준(general standard)이다. 따라서 데이터 품질(data quality)이란 특정 사실이 특정 활용 목적에 부합하는 지 결정하기 위해 사용되는 주관적인 기준(subjective standard)이라 정의할 수 있다.

데이터 품질에 대한 평가는 원하는 시점에서 활용 목적에 부합하는 데이터를 제공하고 있으며, 업무규칙을 준수하는지, 그리고 단위업무가 아닌 전사차원에서 공유될 수 있는 품질수준의 데이터를 제공할 수 있는지에 관한 것이다. 그러므로 환경과 목적, 그리고 업무 프로세스

의 변화에 따라 데이터는 끊임없이 변환하고 생성되므로 데이터품질의 관리를 위한 프로세스 역시 지속적이고 반복적으로 수행되어야 한다.

2.2 데이터 품질 평가 모델

데이터의 품질평가 모델은 초기 현상 분석 중심에서 절차 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].

2.2.1 현상 개선 중심모델

현상적 데이터 품질 요인 발견에 중점을 두는 모델이다[4,5,6].

1980년대 말 시작된 모델링 방법으로 온라인 데이터베이스 사용자 그룹에서 채택된 데이터베이스 평가를 평가하는 Granick의 DB 평가 모델, 서지 데이터나 전문 데이터베이스를 질적인 측면에서 평가를 설계하는 SCOUG 평가 모델, 데이터베이스 품질 중에서도 서비스 부분을 배제하고 데이터 부분에 초점을 맞추는 Large의 품질평가 모델, 그리고 최종 사용자 환경의 사용자 만족도에 대한 요인 분석(factor analyze)을 통해 평가하는 Doll & Torkzadeh의 평가 모델 등이 있다.

2.2.2 구조적 분석 중심 모델

1990년대 중반 이후 제기된 연구 모델로 구조적 데이터 품질 요인의 발견에 중점을 두고 연구된 모델이다 [2,8,9].

기존의 연구에서 보다는 진전된 데이터의 구조적 측면을 고려한 평가 기준을 제시함으로써 보완된 기준을 제시하는 Greg R. Notes의 DB 평가 모델, 데이터베이스 품질 중에서 데이터 부분에 주안점을 두고 있으며, 다음과 같이 ①데이터 신뢰성 ②논리적 무결성 ③물리적 무결성으로 구분하여 제시하는 Brodie의 데이터 품질평가 모델이 있다.

2.2.3 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[3,5,10].

데이터의 품질 평가 뿐 아니라 품질 개선을 위한 프로

세스를 중시하고자하는 모델로 Larry P. English의 TIQM(Total Information Quality Model) 모델, 그리고 데이터의 품질을 정의하고 데이터 품질이 기업에 미치는 영향과 데이터 품질 개선에 관한 3가지 주제를 중심으로 연구한 MIT의 TDQM(Total Data Quality Model) 모델이 있다.

3. 웹 평가 모델

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 <Q, W, S, Φ, F>로 구성된다 [2,11,12].

① W는 각 영역의 품질 평가 기준으로 계산되는 영역의 가중치 집합이다.

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

집합 W의 원소 w_i 는 정의되는 영역의 수 n의 값에 따라 0 부터 1/n 사이의 값을 갖는다.

② Q는 데이터의 품질을 요소들의 평가를 통하여 계산되는 전사적 품질에 대한 가중치를 말한다. Q는 함수 F에 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 S는 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_{si} 로 표현하면, 영역의 가중치 w_i 는

$$w_i = w_{s1} + w_{s2} + \dots + w_{sm} \text{ 이다.}$$

S에 속하는 세부 평가기준 a_i 의 가중치를 w_{a_i} 라 할 때,

$$s_i = \{(a_1, w_{a1}), (a_2, w_{a2}), \dots, (a_l, w_{al})\} \text{와 같이 표현하고,}$$

w_{a_i} 는 0에서 1 사이의 값으로 정의한다.

④ Φ는 세부 평가 항목(S)을 평가할 수 있는 세부 기준들의 집합으로 데이터의 평가 환경에 따라 다양하게 정의될 수 있다.

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

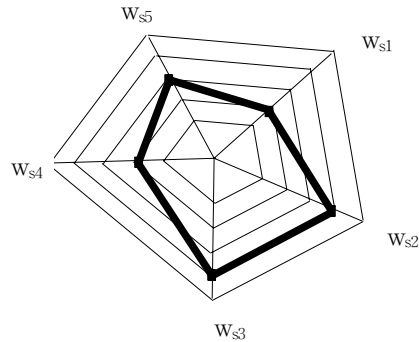
$$l = 1, 2, 3, \dots$$

⑤ F는 데이터의 품질을 결정하는 데이터 품질 평가 함수이다. 함수 F는 다음과 같이 정의된다.

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

함수 F는 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.

Q를 구성하는 값 영역별 가중치 w_i 를 이용하여 분석 모형을 구성하면 그림 1과 같은 웹 평가 모형이 된다.



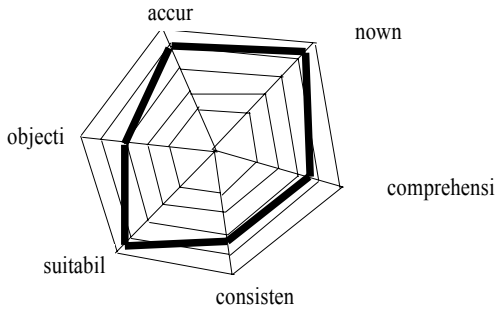
[Fig. 1] Web Evaluation Model of quality level

웹 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 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 2는 6가지 항목으로 구성된 특정 영역의 세부 항목 균형을 나타내고 있다. 이것은 각 항목의 세부 기준이다

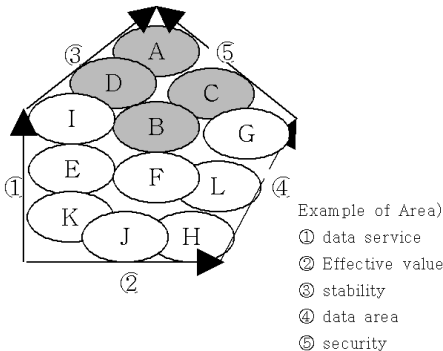
르기 때문에 데이터의 품질에 영향을 미치는 세부 기준을 파악하고, 이에 따른 프로세스를 결정하게 한다[4,12].



[Fig. 2] Example of Web model on Data Area

4. 전사적 품질 평가 모델을 이용한 품질 분석

전사적 품질 평가는 DQnA(Data Quality n Area)를 이용하여 데이터의 적응적 평가 분석이 이루어진다. 그림3은 데이터 품질을 전사적으로 평가하는 Q를 구성하는 영역 품질 지수에 따라 해석하고, DQ5A(n=5) 모형은 데이터를 모두 12가지로 분류하여 각각의 데이터 품질을 분석하고 있다. DQ5A(n=5) 모형은 n의 값을 5로 정의하고 5개 분야에서 데이터 품질을 평가하고 있다.



[Fig. 3] Total Quality Index Model(DQ5A)

데이터의 품질 평가 영역에 따라 그 수는 증가 또는 감소될 수 있으며, 각 영역은 데이터의 품질 평가 중요도에 따라 그 영역을 정의할 수 있다. 그림 2의 음영 부분의 평가 지수는 모든 평가 영역에서 평균 이상의 품질을 유지하고 있는 부분으로 해석할 수 있으며, 기타(E-K) 영

역에 속하고 있는 데이터의 경우는 현재 데이터 품질 영역별 약점과 강점이 분명하게 분석되어 나타남으로써 정보시스템의 데이터 품질 지수를 평가할 수 있다. 이러한 약점에 대한 분석을 통하여 품질 개선을 위한 수정 개선 방안에 활용할 수 있다.

또한 활용되는 데이터의 특성에 따라 그 가중치를 부여함으로써 데이터의 중요한 품질 지수가 무엇인지를 평가하여 업무의 특성과 환경에 맞는 데이터를 추출해 낼 수 있다. 또한 제시하고 있는 모델은 정보 보안 모델로서도 활용될 수 있을 것이다.

5. 결론

다양한 평가 항목의 변화와 사용자의 다면적 특성을 고려한 평가 모델인 수학적 도구를 이용한 평가 모델은 분석 모형과 함께 제시되었으며, 이러한 평가 모델의 활용은 현재의 환경에 따른 데이터 품질을 평가하고, 미래의 변화에 유연하고 적극적으로 대처하도록 한다. 또한 DQnA 모형을 이용한 데이터 품질의 분석은 다양한 분야에서 활용될 수 있는 데이터의 적응적 분석을 위한 기본 모델로서의 역할을 할 수 있을 것으로 판단된다. 또한 다양한 분야에서의 융합적 접근을 통하여 제안된 모델의 활용은 더 극대화될 수 있으며, 차후 정보 보안 품질의 평가 및 분석 분야에서의 단계적 활용 방안에 관한 연구가 지속될 것으로 판단한다.

References

[1] Hye-jung Jung, "A Study of the Data Quality Evaluation", Journal of KSII, Vol. 8, No.4, pp. 119-128, 2007.
 [2] Howard, P., "Data Quality Products :an evaluation and comparison," Bloor Research, 2004.
 [3] Hyun-jin Hong, "A Study on the Development of the Quality Evaluation Standard of Web-Based Databases", Journal of KSLIS, Vol. 39, No. 12, pp. 211-235, 2005.
 [4] Yoeng-won Kim, Jong-ki Kim, "The Development of a Mathematical model to evaluate Data Quality and an Analysis model to improve the Quality",

- Journal of KSII, Vol. 9, No. 5, pp. 109-116, 2008.
- [5] Chan-soo Kim, Joo-seok Park, "A Study of Data Quality Management Maturity Model", Journal of KIISE, Vol. 20, pp. 249-275, 2003.
- [6] <http://dbq.dpc.or.kr/assessment/research.php>
- [7] Yoeng-won Kim, "Study on development model for the Quality Evaluation of Database", Journal of konyang, Vol. 15, pp. 51-65, 2006.
- [8] Kahn, B., Strong, D., and Wang, R., "Information Quality Benchmarks : Product and Service Performance," Communications of the ACM, Vol. 45, 2003.
- [9] <http://dbq.dpc.or.kr/assessment/research.php>
- [10] Kahn, B., Strong, D., and Wang, R., "Information Quality Benchmarks : Product and Service Performance", Communications of the ACM, Vol. 45, 2003.
- [11] Yoeng-won Kim, "A Study on the Use of Web Evaluation & Analysis Model for Data Quality Evaluation", ICCT2013 Proceeding, 2013.
- [12] Yoeng-won Kim, "The Design for Quality Evaluation of Web Service considered on User Aspects", Journal of KIIT, Vol. 6, No. 5, pp. 138-144, 2008.
- [13] Mi-young Park, Hyon-woo, Seung, "A Study On Database Quality Evaluation Item Development Applied CMM", 10th Proceeding of KIPS, pp. 153-161, 2003.
- [14] Jae-whoan Lee, "Quality Evaluation and Management of a Shared Cataloging DB: the Case of KERIS UNICAT DB", Journal of KSLIS, Vol. 36, No. 1, pp. 61-90, 2002.
- [15] Jung-hoon Hahm, Gye-sung Ahn, "The Standardization of the Quality Evaluation item on Database", A study on the standardization of the database, KDA, 2000.

저자소개

김 용 원(Yoeng-Won Kim)

[정회원]



- 1988년 고려대학교 대학원 수학과 졸업(석사)
- 1995년 고려대학교 대학원 전산학과 졸업(박사)
- 1995 ~ 현재 건양대학교 정보보호학과 교수

<관심분야> : 데이터베이스 모델링, 데이터베이스 품질, 디지털 포렌식