

실시간 서지데이터베이스 평가방법에 관한 연구

A Study on Real-time Quality Evaluation Method of Bibliographic Database

노경란

한국과학기술정보연구원 정보시스템부 선임연구원
권오진

한국과학기술정보연구원 정보시스템부 선임연구원
유현종

한국과학기술정보연구원 정보시스템부 선임연구원
문영호

한국과학기술정보연구원 지식정보인프라연구실
홍성화

한국과학기술정보연구원 정보시스템부 선임연구원

Kyung-Ran Noh

Senior Researcher, Div. of Information System, KISTI
Oh-Jin Kwon

Senior Researcher, Div. of Information System, KISTI
Hyung-Jong Yoo

Senior Researcher, Div. of Information System, KISTI
Yeong-Ho Moon

KISTI, Information Infrastructure Support Dept.

Sung-Hwa Hong

Senior Researcher, Div. of Information System, KISTI

중심어 : Database, Quality Management, Quality Classification, Database Evaluation

요약

기존의 데이터베이스 평가방식은 기구축된 데이터베이스를 대상으로 분야별 DB 관리자가 교정·교열을 위한 평가서를 수작업방식(manual method)으로 작성하고 이를 토대로 DB작성자에 대한 평가 및 재교육을 수행하였다. 이로 인해 DB 작성자에 대한 이력정보 및 평가작업에 많은 시간을 소모하여 시간비용측면에서 저효율화 및 DB 작성자에 대한 체계적인 관리가 이루어지지 않아 데이터베이스의 품질향상에 저해요인이 되었다.

본 연구는 시스템을 사용하여 정적·동적 평가 및 DB 담당자에 의한 평가를 결합하여 DB제작의 효율성 및 DB 작성자에 대한 평가결과를 온라인상에서 실시간으로 제공한다. 따라서 DB담당자는 시공간적 제한을 받지 않고 DB 작성자에 대한 재교육을 실시하거나 DB작성자의 지속적인 DB제작여부를 결정하여 DB품질을 향상시킬 수 있다.

Abstract

The conventional database evaluation method is carried out by the way in which the person in charge of each specialty database(DB manager) composes the evaluation sheets for correction·revision on the already-constructed database in a manual method and carries out the measurement and re-education of DB workers based upon it. As a result, that way consumes much time on career information and measurement works about DB workers, causing low time and cost efficiency and lack of systematic management of DB workers, resulting in becoming the hindrance factor of databases' quality improvement.

This research provides on-line, real-time results of measurements about the efficiency of DB production and DB workers by combining the static measurement with dynamic measurement by DB manager, both of which utilize the System. Therefore, the DB manager can contribute to the improvement of DB quality by determining the continuation of DB production by DB workers or carrying out the re-education of DB workers without being affected by time or spacial constraints.

1. 서론

데이터베이스의 품질은 데이터베이스의 바람직한 정도 또는 우수성이라고 정의되며, 데이터 품질과 서비스 품질, 그리고 기타 일반사항에 관한 품질로 구성된다. 데이터베이스의 데이터 품질이란 데이터베이스가 담고 있는 데이터

자체의 바람직한 정도를 의미하며, 데이터 구조, 데이터 값, 데이터 표현에 의해 분석된다. 데이터베이스 서비스 품질은 데이터가 이용자에게 서비스되는 과정에서 매체, 수단, 방법, 그리고 공급처 지원 등의 바람직한 상태를 의미하며, 서비스 정확성, 서비스 편리성, 서비스 친절함, 서비스 신속성, 서비스 비용에 의해 분석된다.

과거 데이터베이스 품질 평가방식은 샘플링, 관찰, 설문 조사방법을 사용하여 데이터베이스 시스템을 대상으로 서비스 시점에서 이용자측면에서 DB평가작업이 이루어졌다. 또한 DB제작기관에서 수행하는 기존의 데이터베이스 평가 방식은 기 작성된 데이터베이스를 대상으로 분야별 DB 관리자가 데이터에 대한 교정·교열을 위한 평가표를 수작업 방식으로 작성하였다. 이로 인해 DB 작성자에 대한 이력정보 관리 및 평가작업에 많은 시간을 소모하였다. 따라서 DB 작성자에 대한 체계적인 관리가 이루어지지 않아 시간 비용측면에서 저효율화를 초래하였고 데이터베이스의 품질 향상에 저해요인이 되었다[1].

본 연구는 온라인으로 주제분야별 DB작성자를 모집·선정하여 데이터베이스를 제작한 후 데이터베이스의 품질을 평가할 수 있는 기준 및 항목을 제시하고, 객관적이고 계량적인 평가방법 및 평가등급을 사용하여 데이터베이스의 품질을 평가분석하는 방법에 관한 것이다. 데이터베이스 품질 평가기준과 평가항목을 이용하여 작성된 데이터를 측정함으로써 데이터베이스의 우수한 부분과 미흡한 부분을 파악할 수 있고, 이 미흡한 부분을 보완하여 품질을 향상시킬 수 있다. 그리고 DB작성자의 성과를 측정하여 평가등급을 부여하여 데이터베이스 제작에 대한 재교육실시 여부 및 교체여부를 판단할 수 있다.

본 연구에서 DB는 단순히 구조화된 데이터의 집합체로 간주하고, DB의 품질은 DB에 수록되어 있는 데이터의 품질만을 대상으로 제한한다. 그리고 사용된 평가기준 및 평가항목은 문자위주의 데이터베이스를 평가하는데 사용되었으며, 멀티미디어 데이터베이스의 품질평가에 적용하기에는 무리가 따른다. 데이터베이스 관리자가 데이터 품질측면에서 데이터베이스를 측정 평가하는 것으로 제한한다.

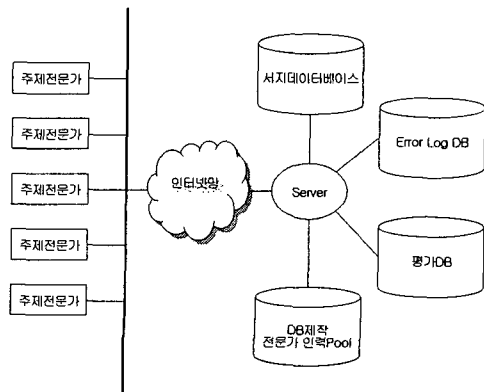


그림 1. 데이터베이스 제작환경

II. 본론

1. 서지데이터베이스 제작

1.1. 전문인력 모집·선정

DB관리자는 적절한 전문인력자원을 상시적으로 신속하고 편리하게 모집할 수 있다. 전문주제지식을 보유한 주제전문가는 자신의 전문지식을 활용하여 경제적인 부가가치를 얻을 수 있어 효율적인 인력자원활용이 이루어진다. 그림 2는 DB제작 주제전문가 모집선정 절차를 보여준다.

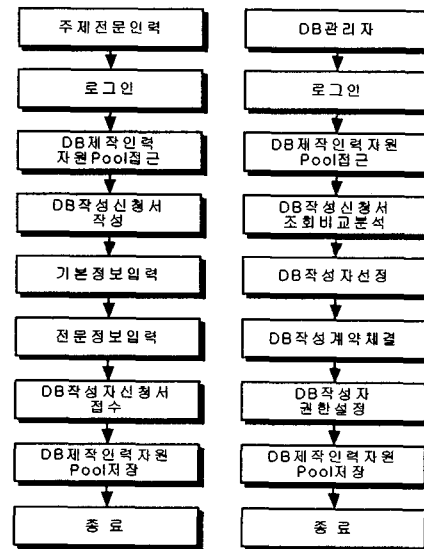


그림 2. DB제작 주제전문가 모집선정

DB관리자는 전문지식을 지닌 주제전문가를 웹상에서 공개 모집하고, 자신이 지니고 있는 전문지식을 제공하고자 하는 개인은 인터넷을 통해 DB제작전문가 인력풀에 자신의 정보를 표준화된 포맷에 따라 데이터베이스에 입력한다. DB작성에 참여하고자 하는 주제전문가가 자신의 기본정보와 전문정보를 DB제작 전문가인력 Pool DB에 저장하면 DB 제작신청 접수가 이루어진다. 전문정보에는 학력, 전문주제분야, 전문분야세부항목, 분야경력, 전문능력, 연구활동 정보, 외국어능력, 특기사항 등으로 구성된다.

DB관리자는 접수된 DB제작 신청서를 저장하고 있는 인력DB에서 주제분야, 학력, 경력년수, 어학점수 등에 우선순위를 부여하여 검색한 후 그 결과를 비교 분석한다. 그리고 적합한 주제전문가를 선별한 후 DB제작 신청자에게 선정결

과를 알린다.

DB관리자와 선정된 DB제작 신청자가 계약을 체결하면 이 주제전문가의 ID와 비밀번호는 DB제작 권한을 가지게 되고, DB제작에 참여하게 된다. DB관리자는 제작된 데이터를 대상으로 품질평가분석을 수행한다. 우수 작성자에게는 상위등급을 부여하며, 불량 작성자에게는 DB제작지침에 관한 교육을 실시하거나 DB작성을 제한한다.

1.2 주제분야별 데이터 가공 및 제작

DB작성자는 네트워크상에서 로그인 후 인증절차를 거쳐 해당 주제분야의 원시데이터에 접근하거나 입수된 DB구축 대상자료를 대출한다. 데이터베이스를 구축하는 과정은 먼저, DB 구축대상자료가 입수되면, DB의 목적에 부합되는 원문헌을 선정한다. 그리고 선정된 원문헌을 분석하여 원문헌을 대체하는 레코드를 작성한다. DB작성자는 DB제작지침에 의거하여 서지레코드를 작성한다. 소장정보, 권호정보, 페이지, 발행년, 자료의 유형, 기사유형, 본문수록언어를 작성하여 기본 서지정보를 생성한다. 논문기사명, 저자명, 수록정보원을 작성한다. 그리고 주제분류코드, 색인어, 초록을 생성한다. DB작성자가 작성한 데이터는 서지데이터베이스에 저장된다.

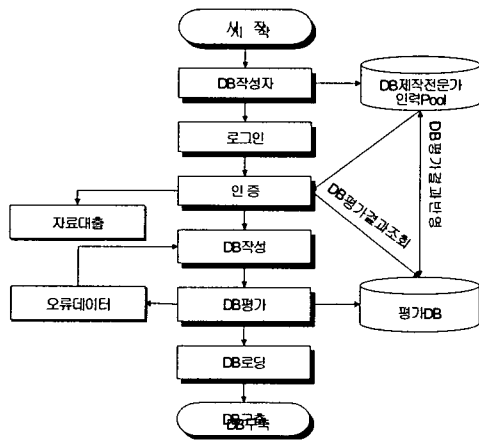


그림 3. 서지데이터베이스 제작과정

DB구축은 데이터입력만으로 끝나는 작업이 아니다. 일단 작성된 DB는 일련의 품질테스트를 거쳐야 비로소 독립된 DB로 생명력을 지니게 된다. DB관리자는 작성 완료된 데이터를 대상으로 시스템을 사용하여 일련의 기준과 방법에 의하여 온라인상에서 실시간으로 데이터의 정확성, 완전

성, 중복성, 일관성을 측정하여 데이터품질을 평가한다. 데이터품질 평가결과는 평가DB에 저장되며, DB작성자가 생성한 데이터의 에러정보는 에러로그DB에 저장된다. DB관리자가 품질평가를 거친 후 오류데이터를 수정하고, DB에 로딩하면 DB가 구축된다. 데이터품질 평가결과는 DB작성자에 대한 평가결과를 포함한다. DB평가결과는 DB제작전문인력Pool에 반영되며, DB작성자에 대한 이력정보에 추가된다. DB작성자는 자신이 작성한 데이터의 품질평가결과를 조회할 수 있으며, DB제작전문가인력Pool에 수록된 기본정보/전문정보를 수정할 수 있다. 서지데이터베이스 작성예와 서비스에는 그림 5와 같다.

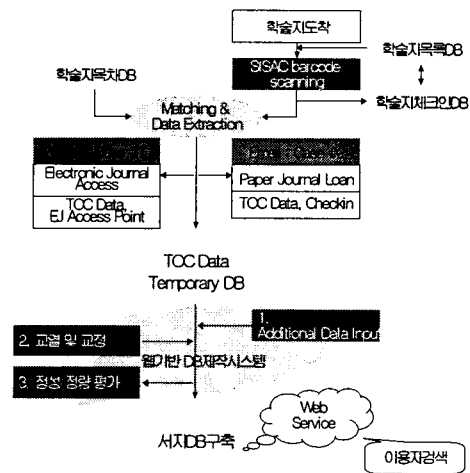


그림 4. 자원연계를 통한 데이터베이스 구축과정

@01K029-/30(4)/423-432/1992
 @03JA1KO
 @04Park, S.h.; Kim, J.H.; Lee, T.J.; Kim, D.H.
 @06Ru/AI O 촉매와 Ni/AI O 촉매상에서 일산화탄소와 이산화탄소의 메탄화 반응속도론
 @07CB0203
 @08촉매; 일산화탄소; 이산화탄소; 메탄화; 반응속도;
 @09루테튬과 니켈촉매에서 CO 와 CO 의 메탄화에 대한 반응속도론적인 연구를 상압과 190-220 °C에서 수행하였다 . CO 와 CO 혼합물의 메탄화반응도 같은 조건에서 조사하였다 . CO 의 메탄화반응에서 0.001-0.08atm 의 CO 분압 범위에서의 반응속도식은 $Nch = kP_{CO}/(1+K_{CO}P_{CO})$ 로 표현되며 , CO 의 메탄화반응에서 0.02-0.2atm 의 CO 분압 범위에서의 반응속도식은 $Nch = K_{CO}$ 의 0 차반응으로 표현되었다 . CO 와 CO 혼합물의 공메탄화에서 CO 는 전적으로 CO 의 메탄화를 저해하였고 CO 는 CO 의 메탄화에 전혀 영향을 끼치지 않았다 . 따라서 혼합물의 공메탄화의 반응속도식은 CO 만의 메탄화속도식과 동일하였다 .
 @110304-128X

등록번호	DOI	EIST199302011791
제작일	IS	199302
자료형태	BT	Journal
기사구분	TC	A (학술논문)
사용언어	LA	Korean
제작	TI	Ru/AI O 촉매와 Ni/AI O 촉매상에서 일산화탄소와 이산화탄소의 메탄화 반응속도론
저자	AU	Park, S.H.; Kim, J.H.; Lee, T.J.; Kim, D.H.;
출처사망	SO	이화여자대학교 (KOR) ISSN: P423-432; 1992; K023;
ISSN	SN	(ISSN) 1548-
분류	CC	CB0203 (반응공학)
색인어	KW	촉매; 일산화탄소; 이산화탄소; 메탄화; 반응속도;
초록	AB	부탄올과 니켈 촉매에서 CO와 CO의 메탄화에 대한 반응속도론적인 연구를 삼광과 190-220℃에서 수행하였다. CO와 CO 혼합물의 메탄화반응도 같은 조건에서 조사하였다. CO의 메탄화반응에서 0.001-0.06atm의 CO 분압 범위에서의 반응속도식은 $Nch = kP_{CO}/(1 + K_{CO}P_{CO})$ 로 표현되며, CO의 메탄화반응에서 0.02-0.2atm의 CO 분압 범위에서의 반응속도식은 $Nch = K_{CO}$ 와 0차반응으로 표현되었다. CO와 CO 혼합물의 공메탄화에서 CO는 전적으로 CO의 메탄화를 저해하였고 CO는 CO의 메탄화에 전혀 영향을 끼치지 않았다. 따라서 혼합물의 공메탄화의 반응속도식은 CO만의 메탄화속도식과 동일하였다.

그림 5. 데이터베이스 작성 및 서비스 예

2. 데이터베이스 품질평가

기존에 수행된 DB의 품질평가와 관련된 연구는 품질평가의 목적과 평가대상 그리고 평가주체에 따라 다양한 평가기준과 절차 그리고 방법이 개발되어 적용되어 왔다. 데이터베이스에 대한 평가는 제작되는 모든 데이터를 대상으로 DB제작 전단계에 걸쳐 지속적이고 반복적으로 이루어진다. 데이터베이스 평가시 데이터의 품질을 평가할 수 있는 평가기준과 객관적이고 계량적인 측정방법을 사용하여 데이터의 품질을 향상시킨다[1].

품질평가 모델의 핵심 요소는 보편성 있는 평가기준과 합리적인 측정방법 그리고 객관적인 등급결정 메커니즘이며, 이 요소 중에 어느 하나라도 결여될 때 그 모델은 객관적이고 신뢰할 수 있는 평가결과를 기대하는 DB 사용자들의 기대와 요구에 부응하지 못하게 된다[2].

본 연구는 일반적으로 수용되고 있는 평가기준을 설정하고, 각 평가기준을 계량적으로 측정하기 위한 지표를 설정하며, 각 지표의 값을 실제로 측정하기 위한 방법을 개발한다. 그리고 측정결과를 해석하여 DB의 등급을 결정하기 위한 메커니즘을 설계구현한다.

2.1. 데이터베이스의 품질평가 기준

측정(Measurement)과 평가(Evaluation)라는 용어가 혼용되어 사용되고 있는데 명확히 정의를 내릴 필요가 있다. 측정이란 기계적으로 실제 값을 구하는 작업이고, 평가란 측정에 의해 구해진 값을 기반으로 미리 정해진 기준값을 충족시키는지 여부를 판단하는 인간의 지적 작업이다[1].

본 연구는 데이터베이스의 데이터품질을 평가하는데 중점을 두고 DB평가기준·측정방법을 설정하고 이를 자동화

하여 공개적이고 객관적으로 데이터베이스를 품질평가한다.

DB평가시스템은 데이터베이스 평가기준을 데이터의 정확성, 최신성, 완전성, 일관성으로 설정하고[3], 각 평가기준에 속한 측정항목을 측정된 값에 가중치를 부여하여 DB품질평가지표를 산출함으로써 데이터에 대한 품질을 측정한다.

표 1. DB 품질평가기준 및 측정항목

품질평가기준	측정항목
정확성	철자오류
	분류코드기입오류
	특수문자입력오류
	서지항목
완전성	불완전 레코드수
	불완전 필드수
	색인/용어부여건수
	번역수준
	초록의 상세함
최신성	제작까지 소요시간
일관성	동일레코드 중복율
성실성	전월대비 개선정도

데이터의 정확성은 데이터베이스를 구성하는 레코드의 구조의 일관성과 데이터 표기의 일관성·규칙성을 평가하는 기준이다. 즉, 레코드를 구성하는 데이터 필드가 모든 레코드에서 일정한 형태를 유지하는지, 레코드를 구성하는 데이터 필드에 기술된 내용이 모든 레코드에서 동일한 표기방식을 준수하고 있는지를 측정한다[2]. 데이터의 정확성은 데이터 값의 정확성과 데이터 표현의 정확성을 포함한다. 이 정확성은 데이터의 신뢰성에 직접 영향을 미치는 기준이며, 철자오류나 잘못된 데이터 값을 발견하는 것과 관련된다. 데이터 평가시 측정되는 데이터의 정확성은 구문적(syntax) 차원에서 이루어진다.

데이터의 완전성은 실제 데이터베이스를 구성하는 각 레코드, 그리고 이 레코드를 구성하는 각 필드를 대상으로 측정된다. 데이터의 완전성은 데이터 구조의 완전성, 데이터 값의 완전성, 데이터 표현의 완전성으로 구분되며, 논리적 무결성과 관련있다. 데이터구조의 완전성은 데이터베이스 논리구조 설계단계에서 다루어져야 하는 요소이므로 본 연구의 품질평가기준에서 제외한다. 데이터 값의 완전성은 불완전한 레코드수에 의해, 그리고 불완전한 필드 수를 계량화하여 데이터의 완전성을 산출한다. 데이터 표현의 완전성은 인간의 지적요소가 투입되어 내용측면에서 심층적으로 다루어져야 하며, 데이터가공시 색인, 분류, 초록을 적절하게 부여했는지에 대한 전문가의 판단이 투입된다. DB를 구

상하는 레코드는 원자료에 대한 대체물 기능을 수행하기 때문에 DB의 완전성은 단위 레코드들의 내용의 완전성을 평가하기 위한 기준이다.

데이터의 최신성은 데이터의 시기적절성과 관련되며, 원시데이터의 발생부터 가공제작까지 소요되는 시간의 경과 (time lag)를 측정함으로써 구해진다.

데이터의 유일성은 데이터베이스내 동일 레코드가 중복되지 않음을 의미하며, 데이터베이스의 절대적 유용성을 평가하는데 중요한 기준이다. 동일 데이터베이스를 구축하는 DB작성자가 다수이며, 분산형으로 제작되는 경우 중복레코드의 발생가능성은 상대적으로 높아진다. 동일 자료에 대해 동일레코드가 중복적으로 출현한다면 이 데이터베이스의 유용성은 저하된다.

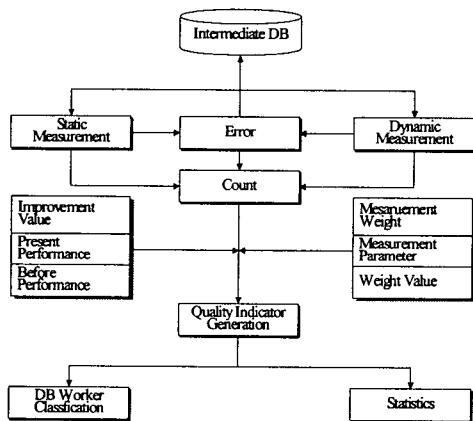


그림 6. 데이터베이스 품질평가방법

2.2. 품질측정방법

DB 품질측정의 가장 큰 어려움은 일부 품질기준이 정량적(quantitative)이 아니라 정성적(qualitative)이라는 데서 찾을 수 있다. 정량적 품질기준이라도 그 정확한 값을 파악한다는 것은 많은 경비와 시간이 소요되며, 또한 비현실적이다[4]. 예컨대 일관성을 파악하기 위해서는 모든 레코드의 대조작업이 이루어져야 하는데 그 작업 자체가 현실적으로 허용될 수 없다고 볼 수 있다. 따라서 일부 샘플 데이터 조사나 사용자 인식도에 의존할 수 밖에 없는데 이 과정에서 객관성 시비와 방법, 도구, 절차상의 문제가 떠오른다.

본 연구에서 사용한 품질측정방법은 정적 평가방법과 동적 평가방법으로 구성된다. 이 두 평가방법은 인간의 지적

분석능력을 어느정도 요구하는지에 따라 분류된 것이다. 정적 평가방법은 인적 요소의 중재없이 기계적으로 데이터의 형태적 무결성을 측정하는 것으로, 서지항목, 주제분야코드, 특수문자, 데이터중복성, DB작성소요시간, 일관성, 전월대비 개선정도가 포함된다.

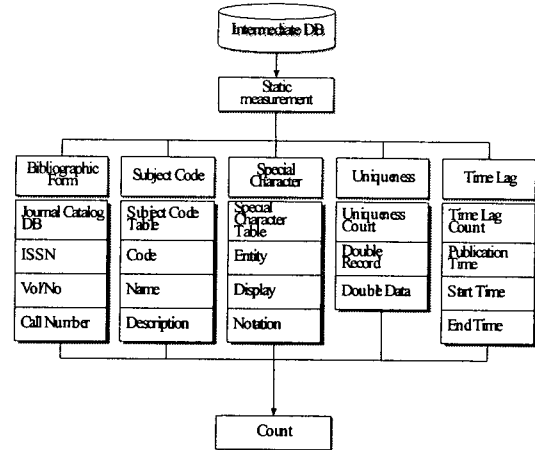


그림 7. 정적 품질평가방법

데이터의 형태적 무결성을 측정하기 위한 첫 번째 방법은 레코드 구조의 일관성을 측정하는 것이다. 이를 위해 필수적으로 입력되어야 하는 기본 데이터필드에 필드명이나 데이터값이 누락되었는지, 아니면 추가된 필드가 있는지 체크한다. 또한 레코드를 구성하는 데이터 필드명과 작성된 데이터 내용이 일치하는지 측정한다.

서지항목에 대한 측정을 통해 정확성과 일관성을 평가할 수 있다. 서지데이터에 대한 일관성 측정은 정해진 형태에 맞추어 데이터가 정확히 기입되었는지 측정하는데 학술지 목록데이터베이스와 연동하여 ISSN, 권호정보, 소장정보의 정확성을 측정한다. 주제분야코드는 주제분야테이블에 의거하여 논리오류를 측정한다. 특수문자테이블을 사용하여 특수문자의 입력오류를 측정한다.

ISSN은 데이터의 입력길이가 고정적으로 정해져 있어 입력길이에 벗어나는 데이터를 찾아낸다. 그리고 ISSN은 학술지를 구분하는 고유식별자라서, ISSN마다 하나의 학술지명을 지니고 있다. 학술지목록데이터베이스와 연동을 통해 작성된 ISSN과 서명이 일치하는지 정확성을 측정한다.

권호정보는 데이터 값이 가변적으로 표현되어 측정하기 어려운 요소이지만, 학술지체크인 테이블을 연동시켜서 입력오류 및 표기의 일관성을 측정한다. 기관 서기번호로 나

타는 소장정보의 정확성과 일관성도 권호정보와 마찬가지로 학술지 체크인 테이블과 연동하여 측정한다.

데이터의 유일성은 데이터베이스내 동일 레코드의 중복률이 없는 데이터의 무결성을 의미하며, 불필요하게 중복생성된 레코드의 수, 레코드내 중복 데이터 값의 빈도를 측정한다. 동일자료에 대해 중복생성된 레코드를 중복처리 알고리즘을 사용하여 자동으로 처리한다.

필드 값이 오래된 레코드의 수, 새로운 데이터가 발생하는 시점부터 데이터가공을 거쳐 레코드생성까지 소요되는 시간의 경과를 측정하여 데이터의 최신성을 평가하며, 데이터의 시기적절성을 평가할 수 있다.

동적 평가방법은 이미 구축되어 있는 데이터를 이용하여 각 데이터필드마다 발생빈도수를 계산함으로써 평가를 수행하는 것이며, 정적 평가방법과는 달리 기계적인 오류가능성 판단 후 인간의 지적분석 작업을 최소한 필요로 한다. 주제분야, 색인어/용어, 철자에 대해 오류가능성이 높은 데이터를 추출하고, 주제전문가는 추출된 데이터 항목별로 오류가 있는지 검증한다. 오류 데이터의 경우 전문지식을 활용하여 데이터 수정이 이루어진다. 초록의 길이 및 색인어/용어 부여간수를 계산하여 데이터 완전성을 평가한다. 오류 데이터를 수정입력하는 과정에서 데이터의 품질측정 알고리즘이 작동한다.

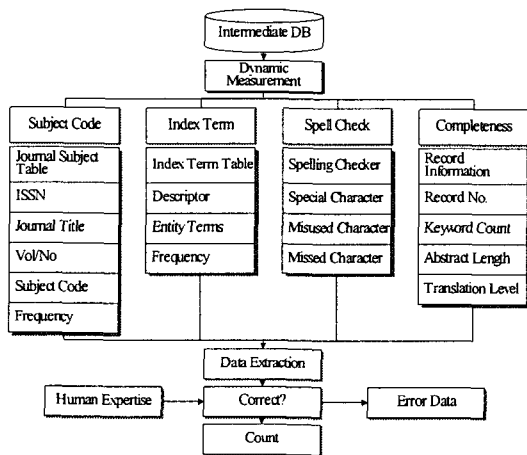


그림 8. 동적 품질평가방법

이 평가시스템은 주제분류코드가 정확하게 부여되었는지 여부를 파악하기 위해 빈도계산을 통해 오류가능성이 높은 데이터를 추출한다. DB대상자료인 각 저널은 대주제분야가

정해져 있으며 세부주제분야로 전개된다. 시스템은 데이터베이스내 모든 데이터를 대상으로 각 저널마다 주제분류코드와 빈도를 계산한 테이블을 생성하여 유지한다. 평가시스템은 저널마다 가지고 있는 분류코드빈도수치와 상이한 분류코드를 지니고 있는 데이터가 오류의 가능성이 있다고 추정하여 자동으로 이 데이터를 추출한다. 주제전문가는 주제분야코드의 정확성을 판단하기 위해 지적 분류작업을 수행한다. 색인어 또는 용어부여작업은 주제전문가의 지적작업으로 내용분석을 통해 중요개념을 추출하고 이를 색인어로 변환하는 과정이다. 평가시스템은 여러 가능성이 높은 주제분야코드가 부여된 데이터를 추출하는 것과 유사한 원리를 사용하여 색인어를 대상으로 오류가능성이 있는 레코드를 추출한다. 시스템은 특정 저널에 대해 기구축된 색인어와는 다른 색인어가 부여되거나, 용어테이블에 존재하지 않은 신조어가 부여된 레코드를 자동추출한다. 그리고 주제전문가는 이 레코드를 대상으로 지적분석작업을 수행한다.

초록은 문헌의 내용을 신속하게 파악할 수 있도록 작성된 문장으로 이용자가 원문헌(original document)을 보지 않고도 이를 충분히 이해할 수 있도록 문헌의 중요한 내용을 정확하고 간결하게 그리고 명료하게 전달하도록 작성된다. 동적 평가방법에서는 초록의 길이를 계산하여 이미 설정한 기준치보다 짧거나 긴 데이터를 추출한다. 그리고 주제전문가의 판단을 통해서 초록데이터의 충실도를 측정평가한다.

3. 품질평가결과

3.1. 측정결과 분석

데이터베이스 품질을 평가하기 위해서는 측정결과를 표준화된 수치 또는 형태로 나타낼 수 있어야 한다. 흔히 DB 품질의 측정결과를 “좋다, 보통이다, 나쁘다”라는 식으로 나타내지만 그것이 정확하게 어떤 의미를 가지고 있는지는 의문이다. 예컨대 “좋다”라는 것이 절대적 의미에서 좋은 것인지 아니면 다른 DB와 비교하여 상대적 의미에서 좋은 것인지 알 수 없다. 본 연구에서는 DB품질의 측정결과를 수치로 나타낼 수 있는 품질지수(Quality Index) 개념을 채택하였다.

데이터베이스의 품질은 정량화된 척도를 통해 0부터 100까지의 품질지수를 통해 표현된다. 그리고 품질지수로 특정 데이터베이스 및 DB 작성자의 품질등급을 제시한다. 제작된 데이터베이스에 품질등급을 부여하여 차별적으로 등급화하고, 각 등급에 맞는 품질수준을 명시함으로써 데이터 품질에 대한 DB작성자의 인식을 제고하고 데이터 품질을

개선시키고자 하는 노력을 유도할 수 있다.



그림 9. 데이터베이스 품질등급 부여과정

본 연구에서 사용한 품질평가시스템은 정적 평가와 동적 평가를 거친 데이터의 측정값에 측정항목별로 가중치를 부여하여 데이터베이스 품질평가지수를 정량적으로 산출한다. 품질평가를 위한 기준과 지표들에 대해 차등의 가중치를 부여한다. 여기서 품질평가를 위한 기준과 측정값에 항상 동등한 가중치를 부여하지는 않는다. 데이터베이스의 개발 목적과 유형, 데이터베이스의 사용자 목적에 따라 각 기준이나 지수가 갖는 상대적 가중치는 상이할 수 있다.

이 품질평가방법은 텍스트 위주의 데이터베이스를 대상

으로 데이터 품질을 평가하는 것으로 제한을 두었다. 따라서 포괄성이나 최신성 같은 데이터 내용적 측면과 정확성, 일관성과 같은 데이터의 구조적 측면이 중요한 기준으로 부각된다.

또한 동일 DB작성자에 대해 과거의 DB품질평가분석결과와 현재 품질평가결과를 비교하여 개선된 정도를 수치로 계량화한다. 이 시스템은 평가기준별, 측정항목별 DB작성자의 순위정보를 제공하며, DB관리자가 DB품질 및 DB작성자를 가장 효율적이고 편리하게 관리하도록 한다.

평가분석결과를 텍스트 및 visualization 형태로 제시하며, 평가내용을 기반으로 미리 정한 기준치와 우선순위에 따라 DB작성자에 대한 등급을 부여한다. DB관리자는 주제분야별, DB작성지별, 기간별, 측정항목별로 데이터베이스의 품질분석결과를 조회할 수 있다. 품질분석결과를 토대로 분석함으로써 데이터베이스뿐만 아니라 DB작성자에 대한 품질등급(quality classification)을 제시하고 품질관리(quality management)를 수행한다.

3.2. 품질평가결과 피드백

데이터의 품질평가 결과로써 피드백 활동이 효과적으로 이루어져야 한다. 데이터를 평가하는 본질적인 목적은 평가 대상이 되는 데이터베이스의 수준 및 DB작성 단위별 계획 및 목표, 실행 과정을 점검하여 더 나은 접근방법이나 대안을 모색하는 것이다. 평가분석후, 성과 실현수준 및 목표달성도만이 관심의 대상일 뿐, 그러한 성과 수준을 가져온 근

DB작성방법 등 품질평가표				
작성일자	-	작성지명		Performance Indicator
총 작업건수		작성자ID		Quality Classification
Measurement	Category	Count	Weight Value	Score
Accuracy	Subran Code			
	Index Term			
	Spell Check			
	Tabular Form			
Completeness	Double Record			
	Time Lag			
Completeness	Index Term Number			
	Abnormal Length			
	Transition Level			
	Incomplete Record			
Calculation	Improved Score			

그림 10. 데이터베이스 품질평가지표

본적인 문제점이나 개선 방안 검토/피드백은 간과되는 경향이 많다. 평가는 점수 매기거나 측정 그 자체가 아니다. 아울러 평가 대상에 대한 순위매김은 더욱 아니다. 평가 결과로 나타난 DB작성자의 역량 특성을 파악하고 고품질의 DB작성에 기여토록 해야한다.

평가시스템은 작성된 데이터를 측정함으로써 데이터베이스 품질의 우수한 부분과 미흡한 부분을 파악한다. DB관리자는 미흡한 부분에 대한 원인파악과 개선방안을 강구하여 보완하고 DB작성자에게 추가교육 또는 보완교육을 실시하여 품질을 향상시킨다.

평가시스템은 품질평가분석결과를 토대로 작성된 데이터에 대해 등급을 매기며, DB관리자는 이를 토대로 DB작성자에 대한 평가를 내린다. 품질평가지표가 높은 등급에 속하는 DB작성자에게는 권위와 인센티브를 부여할 수 있다. 또한 DB관리자는 품질평가결과가 기준치에 미달하는 DB작성자에 대해 데이터베이스 제작과 관련한 계약중단결정을 내린다.

DB작성자는 자신이 작성한 데이터의 품질평가결과 및 성과지표를 조회할 수 있다. 이를 통해 DB작성자는 자신의 작업효율성을 파악하며, DB 품질을 향상시키도록 노력한다.

III. 결론

과거의 서지데이터베이스 제작에는 주제전문지식을 지닌 활용가능 인력이 제한되고, DB관리자가 주제분야별 세부 DB제작 전문가를 관리하고, 제작된 데이터베이스의 품질을 측정평가하는데 한계가 있었다. 따라서 데이터베이스에 수록된 데이터의 부정확성, 중복레코드 및 오류데이터의 발생, 데이터의 일관성 결여, DB구축까지 많은 시간이 소요되는 등 여러 가지 문제가 발생하였다.

본 연구는 데이터베이스 품질평가기준과 방법에 의거하여 계량적으로 산출된 품질평가결과를 분석함으로써 객관적이고 신뢰성있는 작업통계를 실시간으로 온라인상에서 산출하는 것을 목적으로 한다. 데이터베이스 제작과정에 참여하는 주제분야별 전문가들을 차별적으로 등급화함으로써 데이터 품질에 대한 DB작성자의 인식과 책임감을 제고시키고, 제작되는 데이터베이스의 품질을 향상시킨다.

지금까지 수행된 데이터베이스 품질평가에 관한 여러 연구들은 대부분 데이터베이스 제작자측면이 아니라 이용자측면에서 서비스품질을 중심으로 직관적인 방법, 경험적 방법, 이론적 방법을 사용하여 수행되었다. 그리고 이와같은

연구들은 데이터베이스 평가기준만을 제시하고 있으며, 평가지표 설정 및 이들 지표의 측정방법에 대한 구체적인 설명이 결여되어 있다.

향후 데이터품질은 이용자 입장에서 접근하고 평가하는 방법이 필요하다. 뿐만 아니라 서비스 품질을 제고하기 위하여 서비스 품질의 기준이 무엇이며, 서비스 품질을 평가하는 항목이 무엇인지, 서비스 품질을 어떻게 계량화하여 측정하는지 그 방법에 관한 연구 및 시스템 설계가 요구된다.

참 고 문 헌

- [1] 안계성, 함정훈, "데이터베이스 품질평가항목", 서울 : 데이터베이스진흥센터, 2000.
- [2] 이용호, 이제한, 노지현, "KERIS 서지 DB의 품질관리를 위한 평가모델 개발 및 개선방안 수립", 서울 : 한국교육학술정보원, 2001.
- [3] Herden, Olaf., "Measuring Quality of Database Schemas by Reviewing : Concept, Criteria and Tool," 5th International ECOOP Workshop on Quantitative Approaches in Object-Oriented Software Engineering, 2001.
- [4] 김선형, "과학기술정보 데이터베이스의 품질평가에 관한 연구", 석사학위논문, 서울여자대학교 대학원, 문헌정보학과, 1997.
- [5] 이응봉, 조현양, 류범중, 최재창, "과학기술분야 데이터베이스의 품질향상을 위한 품질평가연구", 한국문헌정보학회지, 제35권 제2호, pp.109-132, 2001.
- [6] 이제한, "과학기술분야 서지DB의 품질관리 및 평가방안 : KORDIC의 KRISTAL DB를 중심으로", 한국문헌정보학회지, 제31권 제3호, pp.109-134, 1997.

노 경 란(Kyung-Ran Noh) 정회원



1990년 8월 : 전남대학교 문헌정보학과 졸업 (도서관학사)
1994년 2월 : 연세대학교 문헌정보학과 졸업 (도서관학석사)
2001년 9월 ~ 현재 : 연세대학교 문헌정보학과 박사과정중

1994년 ~ 2000년 산업기술정보원 책임연구원
2001년 ~ 현재 : 한국과학기술정보연구원 정보콘텐츠개발실 선임연구원
<관심분야> : 문화, IT 콘텐츠, 지식구조시각화

1985년 1월 ~ 1990년 12월 : 산업연구원 연구원

1991년 1월 ~ 2000년 12월 : 산업기술정보원 책임연구원

2001년 ~ 현재 : 한국과학기술정보연구원 지식정보평가연구실 책임연구원

<관심분야> : 정보, 가치 평가, 과학기술 콘텐츠

홍 성 화(Sung-Hwa Hong) 정회원



1986년 2월 : 부산대학교 금속공학과 졸업 (공학사)

1989년 2월 : 부산대학교 금속공학과 졸업 (공학석사)

1989년 6월 ~ 2000년 : 산업기술정보원 책임연구원

2001년 ~ 현재 : 한국과학기술정보연구원 정보콘텐츠개발실 선임연구원

<관심분야> : 문화, IT 콘텐츠

권 오 진(Oh-Jin Kwon) 정회원



1990년 2월 : 광운대학교 전자계산학과 졸업 (이학사)

1994년 8월 : 광운대학교 전자계산학과 (이학석사)

1994년 ~ 2000년 : 산업기술정보원 책임연구원

2001년 ~ 현재 : 한국과학기술정보연구원 정보콘텐츠개발실 선임연구원

<관심분야> : 전문정보콘텐츠, 검색시스템, 산업재산권 데이터표준화, 객체지향소프트웨어, 병렬컴파일러, 시각언어

유 현 중(Hyon-Jong Yoo) 정회원



1976년 2월 : 서울대학교 섬유공학과 졸업 (공학사)

1977년 ~ 1980년 : 방림방직(주)

1981년 ~ 1990년 : 산업연구원

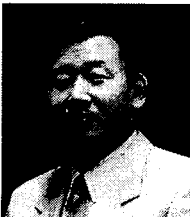
1991년 ~ 2000년 : 산업기술정보원

2001년 ~ 현재 : 한국과학기술정보

연구원 정보서비스실 선임연구원

<관심분야> : 문화, IT 콘텐츠

문 영 호(Yeong-Ho Moon) 정회원



1983년 2월 : 부산대학교 토목공학과 졸업 (공학사)

1986년 2월 : 한국과학기술원 건설환경공학과 졸업(공학석사)

1998년 2월 : 한국과학기술원 건설환경공학과 졸업(공학박사)