

데이터웨어하우스 시스템에서 메타데이터 스키마의 설계 및 활용

박 종 모[†] · 조 경 산^{††}

요 약

고객관리에 중요한 정보로 활용되는 웹로그 분석과 의사결정을 지원하기 위한 데이터웨어하우스를 결합하여 데이터웨어하우스가 등장하였다. 그러나 시스템에 분산되어 저장된 정보를 통합하는 데이터웨어하우스는 다양하고 이질적인 정보를 포함하므로 관리적 측면에서 복잡한 작업을 필요로 한다. 본 연구는 데이터웨어하우스 환경에서 데이터를 효과적으로 관리하고 통합하기 위한 메타데이터 스키마를 제안한다. 제안된 스키마는 데이터웨어하우스 개발을 지원하고 기업의 정보 자산에 대한 통합 관리를 가능케 한다. 또한 데이터웨어하우스의 웹로그 추출을 위한 ETL 메타데이터를 사용하여 웹로그 데이터 처리시간을 향상시킬 수 있다.

키워드 : 데이터웨어하우스, 데이터웨어하우스, 메타데이터, 스키마

Design and Application of Metadata Schema in Datawarehouse System

Jongmo Park[†] · Kyungsan Cho^{††}

ABSTRACT

Datawarehouse consists of both web log analysis used for customer management and datawarehouse used for decision support. However, datawarehouse needs complex operations for management in order to transform and integrate data from heterogeneous data sources and distributed systems. We propose a metadata schema in order to enable data integration and data management which are essential in datawarehouse environments. We show that our proposed schema supports datawarehouse development and enables integrated asset management of business information. With ETL metadata for web log extract, we can improve the data processing time of web log.

Key Words : Datawarehouse, Datawarehouse, Metadata, Schema

1. 서 론

급격히 변화하는 시장 경제 체제에서 기업은 내외부에 존재하는 정보자산을 기업 경영에 활용함으로써 경쟁력을 높일 것을 요구받고 있다. 과거 기업의 정보는 단순히 업무운영에 필요한 것으로 취급되어 왔으나, 현대는 모든 경영활동에서 전략적인 의사결정에 활용된다. 따라서 기업의 경쟁력을 높이기 위해서는 보다 적은 비용과 빠른 시간에 기업의 모든 정보를 통합하여 관리하고 기업의 경영활동에 이용되어야 한다. 이에 따라 기업의 정보를 통합하고 관리하기 위한 메타데이터의 필요성이 제시되었다. 메타데이터는 “기업에 의해 활용되는 기술적 절차와 데이터에 관한 모든 물리적 데이터와 지식을 포함하는 정보”이다[1].

메타데이터는 데이터웨어하우스 시스템에서도 이용이 가능하다. 웹과 같은 온라인 기술이 발전하면서 사용자들은 온라인 채널에 자신의 접근 로그를 남기고, 형성된 로그는 기업에서 고객 관리 및 마케팅에 중요한 정보로 활용된다. 그

러나 일반적인 웹로그 분석은 페이지뷰와 같은 일시적이고 평면적인 정보만을 제공한다. 기업에서는 고객과 연관되는 다차원 분석을 요구하므로, 웹로그 분석에 기존의 데이터웨어하우스를 결합한 데이터웨어하우스가 생성되었다[2]. 데이터웨어하우스는 다양하고 이질적인 정보가 포함되므로 관리적 측면에서 복잡하고 효율적인 작업의 수행이 필요하다. 이에 따라 본 논문에서는 메타데이터를 데이터웨어하우스에 적용할 수 있도록 데이터웨어하우스 시스템에서의 메타데이터 스키마를 제안한다. 제안된 스키마를 활용하여 데이터웨어하우스 구축과 운영을 효율적으로 관리할 수 있고, 사용자가 필요한 정보를 효과적으로 검색할 수 있으며, 기업의 정보 자산을 통합하여 관리할 수 있다. 또한 메타데이터를 관리함으로써 데이터 정의를 표준화하고, 데이터 간의 관계를 손쉽게 이해할 수 있다. 제안된 스키마는 웹로그의 데이터 처리시간의 향상을 위해 데이터 추출 부분을 관리하기 위한 ETL(Extraction-Transformation-Loading) 메타데이터를 포함한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터웨어하우스와 메타데이터를 정의하고 이에 관련된 연구를 기술한다. 3장에서는 데이터웨어하우스에서의 메타데이터의 필요성을 제시하고 관리 시스템을 제안한다. 4장에서는 메타데이터 스

[†] 종신회원 : 단국대학교 정보컴퓨터학부 강의전담교수
^{††} 종신회원 : 단국대학교 정보컴퓨터학부 교수 (교신저자)
 논문접수 : 2007년 2월 28일, 심사완료 : 2007년 7월 5일

키마의 설계를 제안하고, 스키마의 특성을 분석한다. 5장의 결론 및 향후 연구로 본 논문을 마무리한다.

2. 관련연구

데이터웨어하우스는 OLTP(OnLine Transaction Process) 시스템에서 생성된 데이터로부터 다양한 분석 정보를 추출하여 OLAP(OnLine Analysis Process) 시스템에 사용하기 위한 데이터의 저장고이다[3, 4]. 데이터웨어하우스는 사용자들의 다양한 분석요구를 효율적으로 지원하기 위하여 분석의 기준이 되는 차원들과 분석할 내용인 측정값으로 구성된 다차원 형태의 DFM(Dimension Fact Model)으로 데이터를 표현한다. 데이터웨어하우스의 설계 방법론은 다음과 같이 제시되었다[5]. 운영 데이터베이스로부터 수집된 정보를 분석하고, 사용자의 요구 사항을 명세화하는 단계를 거친 후 개념적 설계를 통해 사실과 차원 구조를 생성한다. 개념적 설계 결과에 대한 작업량과 유효성을 판단하여 테이블 구조로 실체화하기 위한 논리적 설계 단계를 거치고, 물리적 설계를 수행한다.

데이터웨어하우스는 웹 기술에 기반하여 데이터웨어하우스를 위한 도구와 프로세스 집합의 관점으로 정의되었다[6]. 데이터웨어하우스가 웹을 통해 지역의 경계를 벗어나 효율적인 지식 공유가 가능해지고 협동작업이 증가하게 되어 집중화된 정보에 쉽게 접근이 가능하게 되었다[7]. 다른 측면에서는 데이터웨어하우스를 웹로그 분석에 고객과 연관된 다차원 분석을 위한 데이터웨어하우스가 결합된 형태로 제시하였다[2]. 데이터웨어하우스는 오프라인을 통해 획득된 데이터를 저장하여 분석에 활용하지만, 데이터웨어하우스는 웹과 같은 온라인 채널에 남겨진 로그를 의미 있는 분석정보로 만들기 위한 데이터를 저장한다. 이러한 데이터웨어하우스의 발전 단계는 <표 1>과 같다[8].

데이터의 저장소인 데이터베이스에서 데이터 사전과 데이터의 구조 및 의미, 데이터를 조작하는 적용업무 및 처리절차에 대한 설명형 정보로 사용하는 것이 메타데이터이다[1]. 이를 확장하면 메타데이터는 “정보시스템 환경에 대하여 각 구성 요소들간의 관계에서부터 비즈니스와 기술적인 구조에 이르기까지 정의하고 설명하는 데이터”라고 제시된다[9]. 메타데이터를 관리 기능에 따라 구분하면 <표 2>와 같이 데이터 표준을 입력하고 지원하는 표준관리 기능, 데이터모델링 도구와 데이터베이스의 카탈로그 정보를 추출하는 구조관리 기능, 데이터와 프로그램 등의 상호연관성을 조회하는

<표 1> 데이터웨어하우스로의 발전단계

발전 단계	기본 구조	특 성
OLTP, EIS/DSS	메인 프레임 클라이언트/서버	실시간 운영 시스템
데이터웨어하우스	클라이언트/서버	데이터 분석에 기초하여 의사 결정자에게 정보 제공
데이터웨어하우스	인터넷	정보의 분산 정보 소스의 통합

활용관리, 기능과 권한 및 스케줄에 관련된 기본관리 기능으로 구성된다[10].

메타데이터가 데이터웨어하우스에도 적용될 수 있는데 데이터웨어하우스 제품간의 상이한 메타데이터 스키마에 관계없이 정보를 공유할 수 있도록 개념적인 차원에서의 메타데이터 스키마가 <표 3>에서 제시하는 계층을 기반으로 연구되었다[11].

메타데이터를 이용한 응용분야로 메타모델의 생성이 연구되었다. 다른 메타데이터와 정보를 교환하기 위한 표준방식으로 XMI(XML Metadata Interchange format)를 이용한 메타모델을 설계하여 객체지향 모델링의 패턴 및 클래스에 대한 부분을 정형화하고 표준화하였다[12]. 또한 소프트웨어 개발 과정에서 설계의 빈번한 수정으로 설계와 구현 사이의 불일치를 유발하게 되는데 이를 해결하기 위하여 비즈니스 객체 모델을 메타모델로 변환하는 응용프로그램 생성기를 제안하고, 구현코드를 생성시켜주는 도구에 대한 연구도 진행되었다[13].

3. 메타데이터 관리 시스템의 제안

3.1 데이터웨어하우스에서 메타데이터의 필요성

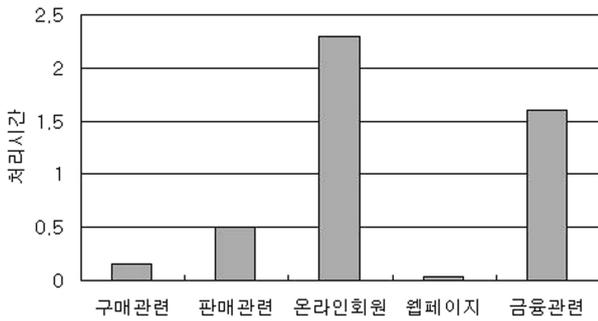
본 연구의 대상이 되는 데이터웨어하우스와 같은 전사적 사결정 지원시스템은 규모와 사용자 수가 계속 증가하고 있어 데이터의 관리와 활용상에 어려움을 겪고 있다. 또한 지속적으로 발전하는 비즈니스 요구를 맞추기 위해 끊임없이 변화하기 때문에 변경 관리가 어렵고, 리포팅 및 분석은 정

<표 2> 메타데이터의 관리기능

관리기능	메타데이터의 유형
표준관리	<ul style="list-style-type: none"> 표준코드, 용어사전, 도메인사전 등의 표준관리 유연성 있는 복수표준 지원
구조관리	<ul style="list-style-type: none"> 다양한 소스로부터 DB, ERD 정보를 추출 모델/구조/표준간 동기화
활용관리	<ul style="list-style-type: none"> 고품질 데이터 유지를 위한 일괄 반영 데이터와 프로그램 구조간 상호연관성 비교
기본관리	<ul style="list-style-type: none"> 사용자 수준/역할별 권한조정 및 규칙관리 다중의 배치 작업을 지원하는 스케줄 기능

<표 3> 데이터웨어하우스의 메타데이터 계층

계층	기능
적용계층	<ul style="list-style-type: none"> 운영시스템관리 적재사업관리 사상관리
핵심웨어하우스계층	<ul style="list-style-type: none"> 데이터베이스 관리 테이블 관리 요소 및 관계 관리
사용자 탐험계층	<ul style="list-style-type: none"> 질의 이력 관리 질의 관리 사용자 관리
비즈니스 계층	<ul style="list-style-type: none"> 비즈니스 정보관리 비즈니스 논점별 질의관리



(그림 1) 데이터 처리시간

확성이 점점 떨어진다. 따라서 이러한 문제의 해결을 위해 데이터처리 과정에 관한 자동화된 변경 추적과 오류 보고가 요구되며, 필요한 정보를 쉽고 빠르게 접근할 수 있는 통합적인 관리가 필요하다.

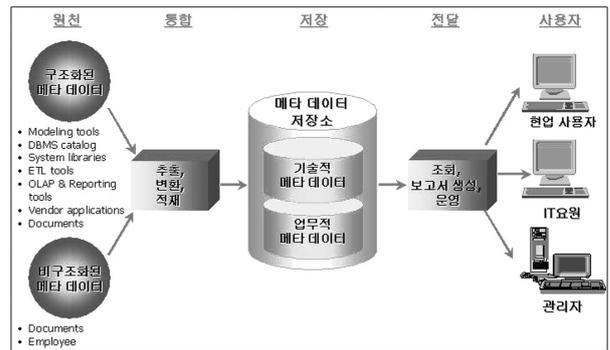
데이터웨어하우스의 메타데이터가 데이터웨어하우스의 메타데이터와 다른 점은 처리되는 데이터의 종류와 데이터 처리의 복잡성이다. 첫째, 처리되는 데이터가 고객 분석을 위한 웹로그이다. 데이터웨어하우스는 의사결정의 지원을 위한 업무 분석용의 데이터를 처리하지만, 데이터웨어하우스는 온라인 상에서 발생하는 웹로그 데이터를 함께 처리한다. 둘째, 데이터 처리의 복잡성과 처리시간의 증가이다. 원천 시스템에서 데이터웨어하우스로의 ETL 작업은 추출작업의 복잡성 때문에 데이터웨어하우스를 구축하는 노력의 80%가 투입된다. 따라서 데이터웨어하우스의 메타데이터는 다수의 데이터의 원천과 목표 테이블간의 대응 정보와 각 작업 단계별 변환 알고리즘 등의 정보, 추출 데이터의 유용성을 확보하기 위한 필터링과 클리징 정보가 필요하다.

데이터웨어하우스 구축 과정에서 해당 데이터를 처리하는 시간은 A사의 사례 (온라인 쇼핑몰에서 의사결정 지원시스템을 개발) 중에서 데이터 처리부분을 표시한 (그림 1)과 같다[14]. 구매와 판매관련 정보는 상대적으로 데이터의 처리시간이 적으나, 여러 웹로그에서 발생한 회원을 통합하여 처리하는 온라인 회원정보는 많은 처리시간이 소요된다.

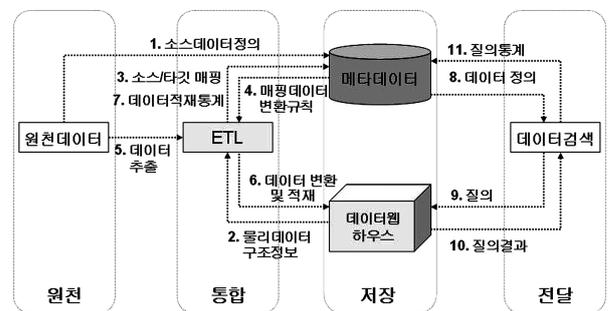
3.2 제안하는 메타데이터 관리시스템

기업 정보의 다양성, 비표준화, 활용의 부족으로 인한 정보시스템 운영상의 문제를 극복하고 정보 자산 관리와 데이터웨어하우스의 구축을 위해 일원화된 메타데이터 관리 시스템이 필요하다. 본 연구에서 제안하는 메타데이터 관리 시스템은 (그림 2)와 같이 현업 사용자와 관리자들이 업무 수행에 필요한 정보를 쉽게 접근하고 공유할 수 있도록 메타데이터 저장소를 구축하고 지속적으로 메타데이터를 원천 저장소에서 통합(추출, 변환, 적재)하여 전달하는 기능을 수행한다.

원천데이터는 통합될 메타데이터가 저장되어 있는 저장소로 구조화된 메타데이터와 비구조화된 메타데이터로 구분된다. 메타데이터 통합은 다양한 원천저장소에 존재하는 메타데이터를 지속적으로 추출하고, 메타데이터 저장소에 저장될 형태로 변환하여 적재하는 과정이다. 메타데이터 저장소는 운영계의 데이터 같은 물리적인 시스템에 관련된 구조화된



(그림 2) 메타데이터 관리 시스템



(그림 3) 메타데이터의 처리흐름

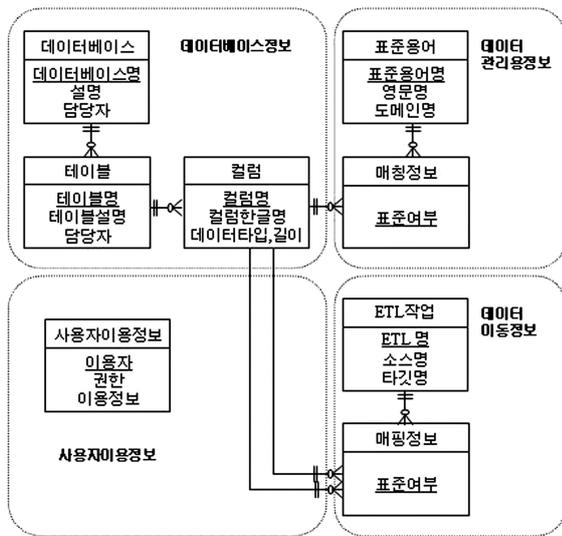
메타데이터와 기업의 정보 자원과 업무 절차 및 지식에 관한 정보를 저장하는 비구조화된 메타데이터를 통합하여 저장한다. 메타데이터 전달은 메타데이터 저장소에 저장된 메타데이터를 실제 업무에 활용할 수 있도록 의미 있는 형태로 가공하여 조회와 보고서를 통해 현업 사용자와 관리자들에게 전달한다.

(그림 2)의 메타데이터 관리 시스템에서의 메타데이터의 처리흐름을 (그림 3)과 같이 제안한다. 원천데이터에 대해 데이터의 추출, 변환, 적재에 해당하는 통합을 수행하여 메타데이터를 저장한다. 사용자가 필요할 경우 메타데이터에 대한 전달과 검색이 이루어진다.

4. 메타데이터 스키마의 설계

본 장에서는 데이터웨어하우스에 활용하기 위한 메타데이터 스키마의 설계를 제안한다. 앞 장에서 데이터웨어하우스의 문제로 지적된 웹로그 처리를 위한 ETL 메타 데이터를 제안하고, 데이터 추출 부분을 효율적으로 관리하여 웹로그의 데이터 처리 시간이 개선됨을 보인다. (그림 4)는 본 연구에서 제시하는 메타데이터 스키마모델의 개략도이다. 제시된 스키마 모델은 <표 4>와 같이 데이터베이스 정보와 관리용 정보를 저장하며, 데이터웨어하우스를 추출하기 위한 데이터 이동정보, 사용자들이 이용하는 리포트와 질의에 대한 사용자 이용정보를 저장한다.

<표 4>의 데이터 이동정보 중에서 웹로그에 대한 ETL 처리를 위해 필요한 추가적인 특성을 <표 5>와 같이 분석



(그림 4) 메타데이터 스키마

<표 4> 메타데이터 스키마 모델

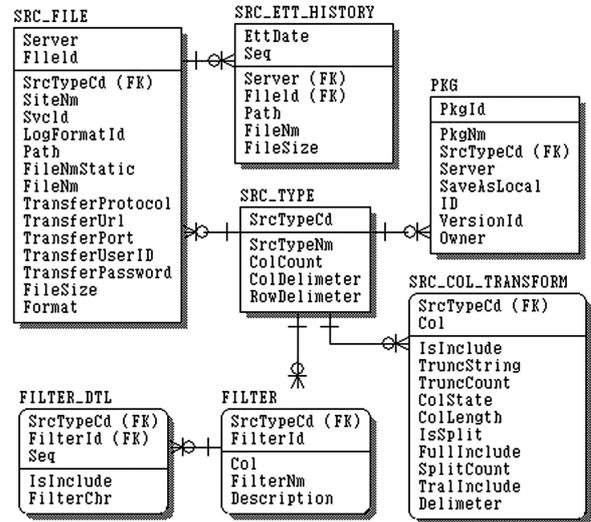
구분	정보내용	탐색 유형
데이터 베이스 정보	데이터베이스 명 테이블, 컬럼명, 인덱스명 사실/자원 테이블 정보 Primary/Foreign키 정보 연결정보, 권한정보	DB를 구성하는 테이블 목록 데이터의 물리적 속성 데이터의 인덱싱 정보 데이터의 갱신과 사용시기 데이터의 사용권한
데이터 관리용 정보	데이터 모델 정보 표준용어, 매칭 정보 엔티티, 속성 정보 도메인, 약어 정보 엔티티별 비즈니스 룰	표준용어 및 표준여부에 대한 정보 비즈니스 엔티티 형태 및 속성 속성들이 가지는 코드 값의 의미 데이터 품질에 대한 책임
데이터 이동정보	소스 및 타겟 정보 소스 및 타겟 매핑 정보 변환 로직, 변환 버전 매핑 값, 예외처리 내역 이동시간, 주기	데이터의 원천 정보 데이터 추출방법과 조건 소스에서 타겟의 변환 과정 정보 데이터 추출작업 상태 정보 예외데이터에 대한 처리 내역
사용자 이용정보	이용자 권한, 보안 정보 리포트, 질의유형 정보 리포트 분배정보 리포트 원천데이터 정보	리포트와 질의의 접근 정보 리포트의 데이터에 대한 원천 질의의 결과로 얻을 수 있는 정보 질의생성자와 사용자에 대한 정보

하였으며 웹로그 처리정보의 내용은 다음과 같다. 첫째, 웹 로그에 남아 있는 고객 정보를 쿠키를 이용하여 추출하고 고객에 대한 데이터웨어하우스와 연관시켜 다차원 분석 정보를 생성한다. 둘째, 웹로그에 포함된 이미지와 같은 불필요한 정보를 제거하는 필터링 정보와 갱신이 안된 노후 데이터, 데이터베이스와 응용 프로그램간의 불일치 데이터, 고객의 중복 데이터에 관련된 정보가 추가되었다. 셋째, 디렉터리 분석을 통한 내부 콘텐츠의 이용 현황 분석, 레퍼터를 이용한 외부 콘텐츠에 대한 채널 추적과 고객의 방문경로 분석(클릭스트림)을 통해 각 채널에서 구매까지의 이동 경로에 대한 다차원 분석정보 등이 추가되었다. 추가된 웹로그 처리 메타데이터를 통해 고객정보를 통합하고 차별화된 고객관리의 전략수립이 가능하다.

(그림 5)는 <표 5>의 메타데이터에서 웹로그 처리정보 중 가장 중요한 필터에 관련된 메타데이터 테이블의 구성도이다. 원천 소스인 웹로그에 관련된 소스 정보테이블과 추

<표 5> 데이터웨어하우스의 ETL 처리를 위한 메타데이터

구분	관리 정보
데이터 변환정보	변환 로직(단순·복합·조건 갱신) 업무코드의 변환(예: 상품코드) 기계코드의 변환(ASCII, EBCDIC) 한글코드 변환(KSC5601 등) 숫자유형 지원 (COMP 모드 등)
운영 관리정보	엑셀 등 외부 데이터 추출 정보 오류의 원인분석 및 재전송 정보 시스템장애에 장애로그 처리정보 시스템 운영변수(최대 로그수 등) 전송상태의 모니터링 정보
웹로그 처리정보	웹로그 연관 다차원 분석 정보 웹로그 필터링, 클린징 정보 디렉터리, 레퍼러, 클릭스트림 정보



(그림 5) 소스와 필터의 메타데이터

출하는데 필요한 필터 테이블로 구성된다. 웹로그 필터링을 처리하는 서비스 사례는 <표 6>에서 제시한다. 해당 서비스는 소스 파일을 분석하여 구분자와 로우 필터 규칙을 선택하여 적용한다.

메타데이터의 처리에서 이기종에 존재하는 다양한 데이터 소스로부터 데이터를 추출하여 데이터웨어하우스에 저장하는 과정을 메타데이터 스키마를 통해 (그림 1)에서 제시된 2.3 시간에서 1.6 시간으로 30%정도 감축시켰다. 이는 메타데이터를 통한 자동화 기능과 지속적인 모니터링을 통해 처리 과정에서 발생하는 문제점을 제거했기 때문이다.

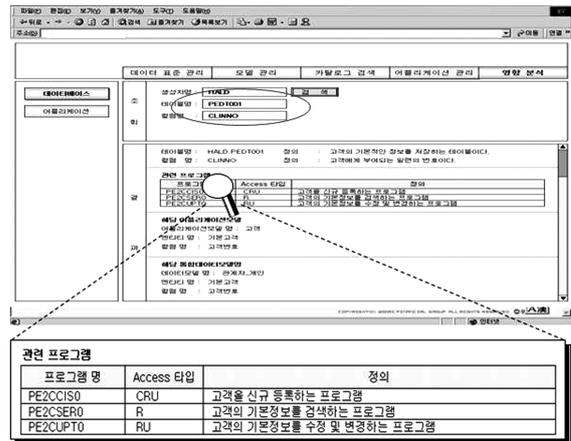
<표 7>은 제안된 스키마의 특성을 분석하여 제시한다. 데이터웨어하우스는 데이터웨어하우스의 메타데이터를 기반으로 구성되었기 때문에 데이터웨어하우스와 메타데이터의 요구사항을 만족해야 한다. 제안된 스키마는 <표 2>에서 제시된 메타데이터의 관리기능을 충족하고, <표 3>에서 제시한 데이터웨어하우스 스키마의 계층을 만족한다. 더불어 데이터웨어하우스는 처리되는 데이터의 특성이 웹로그이며 이로 인해 발생하는 ETL의 복잡성 때문에 메타데이터 데이터 관리 부분, ETL 처리 부분, 웹 데이터 처리 부분이 요구사항

<표 6> 웹로그 필터링 처리 서비스

서비스명	서비스 명세
Creat SQL	MetaData를 Select하기 위한 SQL문 생성 Step1: 컬럼 개수 Select Step2: 구분자 Select Step3: Row 필터 규칙 Select Step4: 컬럼 필터 규칙 Select Step5: Recursive 필터 규칙 Select
서비스 사례	<pre> /* MetaData Column Definition Count */ sQuery ="SELECT Count(*) colCount FROM ColumnDefinition" sQuery+="WHERE CD_IsInclude=1 AND LF_ID="+ m_sLF_ID LF_ID=12 RealTime WWW Apache CLF /* Delimiter */ sQuery += "SELECT IfLF_ColumnDelimiter, IfLF_RowDelimiter" sQuery += "FROM logformat If WHERE If.If_id = " + m_sLF_ID /* Row Filter */ sQuery+="SELECT cd.cd_PhysicalSequence,rfd.rfd_isinclude,rfd.rfd_characters" sQuery += "FROM ColumnDefinition cd, RowFilter rf, RowFilterDetail rfd" sQuery += "WHERE cd.cd_id = rf.cd_id And rf.rf_id = rfd.rf_id and cd.If_id=" + m_sLF_ID; sQuery += "ORDER BY cd.cd_PhysicalSequence; " </pre>

으로 포함되었다. 첫째, 정보의 다양성 및 데이터 관리의 측면에서 데이터웨어하우스 운영상의 문제를 극복하기 위해 메타데이터 관리 부분이 추가되었다. 메타데이터 관리시스템을 통해 데이터를 한 곳에 통합하여 활용도를 높이고, 데이터의 표준화 및 정합성을 높인다. 또한 제시된 메타데이터 처리 흐름을 통해 메타데이터를 한 눈에 파악하고 단계별 처리와 오류 파악 및 수정이 가능하다. 둘째, 웹로그 데이터 처리의 복잡성을 해결하고, 추출 데이터의 유용성을 확보하기 위한 필터링 등의 처리를 위해 ETL에 대한 기능이 보완되었다. 웹의 특성상 이미지 파일과 각종 태그들이 사용되기 때문에 웹로그 처리시 약 90% 이상이 불필요한 웹로그 정보이다. 따라서 웹로그 필터링을 통해 불필요한 웹로그를 제거하여 전체 웹로그를 처리하는 것보다 처리 속도가 개선되었다. 셋째, 온라인 채널에 남겨진 로그를 처리하는 웹 데이터 처리의 요구사항이 추가되었다. 데이터웨어하우스의 특성에 맞게 웹로그와 고객 데이터와의 연관분석을 통해 고객에 대한 다양한 마케팅 지원이 가능하게 된다.

제안된 스키마를 활용하는 전체 시스템의 일부가 구현되었으며, (그림 6)은 <표 5>의 메타데이터 스키마 모델에서 사용자의 이용정보에 해당한다. 사용자가 시스템 자원정보를 이용하기 위해 메타데이터 시스템에 접속하고, 영향도 분석을 선택하여 변경을 원하는 컬럼명을 입력하고 조회한다. 조회의 결과로 관련 프로그램 목록을 확인하고 수정 후 컴파일 가능하다.



(그림 6) 메타데이터의 영향도 분석

(그림 7)은 <표 5>의 데이터베이스 정보 중 데이터관리 정보를 보여준다. 모델링에 필요한 원천 테이블의 정보를 메타데이터 시스템을 이용하여 파악할 수 있고, 파악된 데이터를 참조하여 데이터웨어하우스 모델링을 진행한다.

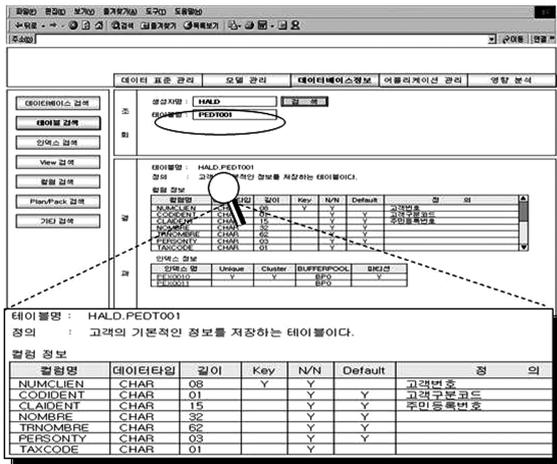
제안된 메타데이터 관리는 추가적으로 다음의 세 가지 장점을 가진다. 첫째, 정의의 일관성을 가질 수 있고, 데이터의 중복과 일치를 최소화할 수 있으며, 동일한 데이터 표현과 정의가 기업전반에 적용된다. 둘째, 관계의 명확성을 가질 수 있다. 데이터웨어하우스 시스템의 모든 요소들인 비즈니스 규칙, 테이블, 열, 변환 등의 관계와 상호작용을 명확하게 할 수 있고, 데이터 변경에 따른 영향을 정확하게 예측하고 관리할 수 있다. 셋째, 정보의 가용성을 가질 수 있다. 메타데이터의 존재는 외부 인터페이스를 통해 노출되지 않지만 데이터의 출처, 데이터를 정의한 사람, 수정된 날짜 등을 담고 있다. 따라서 메타데이터를 이용하여 비즈니스 업무 분석을 효율적으로 수행할 수 있으며, 정보를 자산화하고 정보 활용을 최대화 할 수 있다.

5. 결론

최근에 웹과 같은 온라인 기술이 발전하면서 사용자들은 온라인 채널에 자신의 흔적인 로그를 남기고, 이러한 로그는 기업에서 고객 관리 및 마케팅에 중요한 정보로 활용된다. 웹로그 분석에 고객과 연관되어 다차원 분석을 하기 위한 데이터웨어하우스가 결합되어 데이터웨어하우스가 생성되

<표 7> 제안된 스키마의 특성분석

데이터웨어하우스의 요구사항	메타데이터의 관리기능	데이터웨어하우스의 스키마 계층	제안된 스키마의 특성
데이터 관리	기본관리	핵심웨어하우스	DB 정보
	지원 미비	지원 미비	메타데이터 관리시스템 추가
통합성과 표준화	표준관리	비즈니스 계층	데이터 관리정보
데이터 추출	구조관리	적용계층	이동 정보
ETL의 복잡성	해당사항 없음	지원 미비	ETL 메타 보완
웹 데이터처리	해당사항 없음	지원하지 않음	지원
정보요구 충족	활용 관리	사용자 탐페계층	사용자 이용정보



(그림 7) 데이터베이스 정보

었다. 그러나 시스템에 분산되어 저장된 정보를 통합하는 데이터웨어하우스는 다양하고 이질적인 정보가 포함되므로, 관리적 측면에서 효율적인 작업 수행이 필요하다. 이와 같은 기능을 수행하기 위해 본 연구에서는 데이터웨어하우스 시스템에서 메타데이터 관리 시스템과 메타데이터 스키마를 제안하였다. 메타데이터 관리 시스템은 사용자들이 업무 수행에 필요한 정보를 쉽게 전달받을 수 있도록 원천 데이터에서 추출, 변환, 적재의 과정을 통해 메타데이터 저장소를 구축한다. 메타데이터 스키마는 데이터베이스 정보와 데이터의 관리정보, 이동정보 및 이용정보로 구성된다. 특히 이동정보 중에 데이터웨어하우스의 ETL 처리를 위한 추가적인 스키마를 정의하였다.

제안된 메타데이터 관리 시스템은 데이터웨어하우스의 구축 단계에서부터 운영 단계까지 중요하게 관리되어야 하는 메타데이터를 효율적으로 관리할 수 있는 체계를 제공한다. 메타데이터 스키마는 데이터웨어하우스의 내용을 정의해서 사용자에게 업무 분석에 관련된 정보를 제공하고, 전사적인 의사결정에 필요한 정보를 쉽고 빠르게 접근할 수 있는 통합적인 관리를 가능케 하며, 운영계 원천 시스템으로부터 데이터웨어하우스까지 데이터의 흐름과 다른 주제 영역에 있는 관련 데이터까지도 추적이 가능케 한다. 또한 본 연구에서는 데이터웨어하우스의 문제로 지적된 데이터 추출 부분을 효율적으로 관리하기 위한 ETL 메타데이터를 제안하였고, 이로 인해 웹로그의 데이터 처리 시간이 향상됨을 보였다.

향후에는 본 연구에서 설계된 스키마를 바탕으로 메타데이터 데이터베이스를 구축하고 사용자가 필요한 메타데이터를 입력하고 조회할 수 있는 프로토타입 시스템의 완전한 구현이 필요하며, 이를 통해 메타데이터 스키마의 실용성을 검증하고자 한다.

참고 문헌

[1] David M., Building and Management the MetaData Repository: A Full LifeCycle Guide, John Wiley & Sons, 2000.

[2] 박종모, 조정산, “데이터웨어하우스의 성공적인 개발 전략,” 경영정보학회, 2004년 추계학술대회, pp.766-772, 2004.
 [3] Immon W., Building the Data Warehouse: 3rd Ed., John Wiley & Sons, 2002.
 [4] Kimball R., Reeves L., Ross M. and Thornthwaite W., The Data Warehouse LifeCycle Toolkit, John Wiley & Sons, 1998.
 [5] Golfarelli M. and Rizzi S., “A Methodological Framework for Data Warehouse Design,” Proc. of DOLAP, pp.3-9, 1998.
 [6] Tan X., David Y. and Fang X., Web warehousing: Web technology meets data warehousing, Elsevier Science, 2003.
 [7] Moeller R., Distributed data warehousing using Web technology, AMACOM, 2001.
 [8] Agosta L., The essential guide to data warehousing, Prentice Hall, 1999.
 [9] Devlin B., Data Warehouse from architecture to Implementation, Addison & Wesley, 1997.
 [10] 손동성, “전사 메타데이터 관리의 중요성과 활용 방안,” SAS Magazine, 2000.
 [11] 이희석, 조남철, 손명호, 김태훈, “데이터웨어하우스 메타데이터 스키마 구현,” Information System Review, 제12권 1호, pp.51-69, 2000.
 [12] 이돈양, 송영재, “XML 기반 객체지향 메타모델 생성,” 정보처리학회논문지, 제11권 2호, pp.397-406, 2004.
 [13] 김치수, “메타데이터를 이용한 응용프로그램 생성기의 개발,” 정보처리학회논문지, 제13권 1호, pp.97-102, 2006.
 [14] 위세아이텍, “CRM: B2 분석 시스템 구축 및 현행 시스템 분석서,” 2003.

박종모



e-mail : jmpark@dankook.ac.kr
 1995년 단국대학교 전자계산학과(학사)
 1997년 단국대학교 전산통계학과(석사)
 2007년 단국대학교 정보컴퓨터학과(박사)
 1997년~2000년 한국정보시스템 ERP개발부
 2000년~2003년 위세아이텍 CRM사업부

2003년~2006년 이지엠텍 전산실
 2007년~현 재 단국대학교 정보컴퓨터학부 강의전담교수
 관심분야 : 소프트웨어공학, 데이터베이스, 컴퓨터시스템

조경산



e-mail : kscho@dankook.ac.kr
 1979년 서울대학교 전자공학과(학사)
 1981년 한국과학원 전기전자공학과(공학석사)
 1988년 텍사스 대학원(오스틴) 전기전산공학과(Ph.D.)

1988년~1990년 삼성전자 컴퓨터부 책임연구원
 1990년~현 재 단국대학교 정보컴퓨터학부 교수
 관심분야 : 네트워크시스템 및 이동통신보안, 컴퓨터시스템