2005 한국경영과학회/대한산업공학회 춘계공동학술대회
2005 5월 13일~14일, 충북대학교

메타데이터 마이닝 시스템의 관리효율성의 제고전략

Strategy to Improve the Management Efficiency of Meta Data Mining System

윤 용 운
강원대학교 수리정보학부
yoonwoon@kangwon.ac.kr

Abstract

Many large organizations that have allocated resources to Data Administration(DA) have DA-context meta data mining. Also meta data is an interesting topic in the data warehouse world. This conceptual view gradually cleared up, and recently we have been talking more confidently about the back-room and front-room meta data. We describe the processes and problems that characterize the general architecture of s meta data mining system to do improve management efficiency that require further research and development.

1. 메타데이터 마이닝 시스템의 필요성

현대의 경영정보는 사용자가 주도하는 지능형 시스템을 구현하는데 기업과 조직의 관련자들은 CRM이나 SCM등 기업의 정보시스템에 축적되어가는 데이터의 경향이나 패턴을 발견하여 통합적으로 처리한다. 따라서 정보기술환경이 다변화는 시점에서 정보의 배타성과 공유성 실현을 목표로 데이터베이스와 데이터웨어하우스에 대한 분석이 제고되는 것이다. 지식기반사회가 전략적이고 지능적인 업무방식을 강력하게 요구하므로써 경영정보 시스템에 의한 대용량의 데이터와 다양한 요소기술을 결합한 통합환경을 구성하는 전략적인 접근이 필요하다[1]. 특히 비즈니스 데이터가 어떻게 생성되고 요약되며 사용되었는지에 대한 것과 비즈니스 데이터의 구조 및 의미도 구체적으로 파악하는 메타데이터에 대한 마이닝 시스템의 관리는 고도화된 정보기술환경으로 인식하고 있는 데이터웨어하우스 환경 속에서 애플리케이션 라이프 사이클 상에 중대한 영향을 미친다. 이러한 메타데이터에 대한 마이닝 시스템을 구축함으로써 정보시스템의 환경의 효율성과 최적성을 제고할 목적과 필요성이 제기된다[2]. 이 논문은 메타데이터 마이닝 시스템의 운영에 필요한 소스데이터를 찾아내어 경영요구사항이나 기술적 구조의 설정을 통해 데이터베이스 객체의 동태적인 관리를 가능하게 하는 구조를 구축하고 관리하는 방안에 대한 일반적인 절차를 설명하게 되며, 따라서 운용중인 시스템에 대해 평가하고 앞으로의 데이터와 시스템의 향상을 위한 계획을 수립하고 구축하는 일련의 과정을 정보기술환경의 선택과 운영의 적합성의 기준에 따라 관리하는 관리효율성의 제고전략을 모색하려는 것이다.

2. 메타데이터 마이닝 시스템의 관리 및 운영

정보의 품질과 정확도를 향상시키고, 의사결정 지원 업무를 위한 정보기반을 제공하려는 메타데이터 마이닝 과정은 경영활동의 모델링과 의사결정지원 및 고객관리시스템을 결부시켜 통합환경을 구축하는 전략적인 접근이 제고된다. 메타데이터 마이닝 구조는 데이터웨어하우스 프로젝트를 수행하고 관리하기 위한 업무별, 영역별, 단계별 프로세스 업무들을 유기적으로 결합하여 정보경영환경을 최적화하고, 현재의 정보기술부문의 목표나 전략이나 방침을 경영내부 및 외부환경과 일관성이 유지되도록 설정하는 것이다.

메타데이터의 마이닝 구조의 객체는 개발단계의 설계객체와 구축단계의 설계객체에서 유도되는 형태를 취한다. 이러한 객체의 설계와 구축은 일관된 객체를 유지하도록 정확한 순서 내에서 실행되며, 구축의 효율성과 정확성을 고려한 관계성을 가져야 한다. 또한 웨어하우스 기반의 분산설계 환경에서 효과적으로 설계 및 구축객체를 관리할 수 있도록 수행하는데 분산객체지향 마이닝환경에서 더 많은 사용자들이 데이터를 공유하거나 실시간 형태로 구현하게 된다. 따라서 메타데이터의 마이닝 구조상의 계층적 아키텍처에서 수행하는 몇 가지 속성을 구분한다.

첫째, 마이닝 로그데이터를 분석하여 사용자들의 선호도 변화에 대한 지속적인 갱신데이터를 추출하고 분석하게 되며,

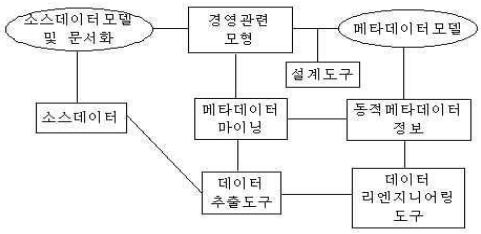
둘째, 메타데이터 마이닝의 처리결과에 대한 데이터 패턴을 추출하여 응용 프로그램의 서식항목이나 쿠키의 상태정보를 유지하도록 하고,

셋째, 메타데이터 마이닝 구조환경의 효율성을 극대화하도록 정보기술의 변환 프로세스를 활용한다.

넷째, 효과적인 정보처리와 정보환경의 양방향성을 최적화하는 메타데이터 마이닝의 아키텍처의 속성을 통합할 수 있어야 하며,

다섯째, 비구조적이고 비규격화된 데이터의 중요성이 증가하고 있는 현재의 기업정보 기술환경에서 고도의 정보서비스라는 메타데이터 마이닝의 본래의 목적을 충족시키기 위해서 마이닝 시스템에 직접 접속하여 시각적인 분석과 통합된 기술 및 마이닝 툴과 엔진을 처리한다[3].

이러한 메타데이터 마이닝의 구조는 조정된 데이터베이스의 갱신의 유효성, 마이닝 결과의 정확성, 더 많이 얻게 되는 지식과 결과의 해석과 수행에 연관된 문제를 해결하는 효율성을 제고하게 되는데 이 과정의 절차는 (그림 1)과 같다.



(그림 1) 메타데이터 마이닝의 절차

한편 메타데이터의 속성상 데이터의 패턴의 추출, 정제 및 업무과정의 수행뿐만 아니라 질의도구와 리포트 작성기능을 조정하는 모든 데이터요소에 의해 표시된 비즈니스 콘텐츠의 기능을 내포하기도 한다. 이 메타데이터가 지향하게 되는 정보생명체화는 메타데이터의 적응성을 모델로 하는 생물학적 개념을 토대로 구축하게 되는 정보시스템인 것이다. 즉 세포가 가진 구조와 환경에의 적응성 등을 모델로 메타데이터 정보시스템이 세포가 할 수 있는 정도의 능력을 발휘하는 생명체적 역할을 하게 된다. 기업의 경영지능지원과 함께 기업의 환경체계내에 자율성을 갖는 메타데이터 마이닝 시스템에서 고도화되고 요약된 정보를 얻을 수 있으므로 정보사용자의 정보요구는 일치성과 정확성을 갖게 된다. 또한 메타데이터의 구조가 핵심적인 데이터와 최적화된 데이터간의 독특한 형태를 취하게 되므로 데이터관리의 효율성을 높이는 데이터분해과정도 고려할 수 있다. 메타데이터의 분해는 데이터 축적, 데이터요약, 순환요약구조의 이용, 상세항목에 대한 선택적 인덱스사용, 표본데이터의 추출 및 데이터 부분집합의 선택적 이용 등 다양한 형태로 이루어지게 된다. 따라서 이러한 메타데이터는 적절한 변환이 필요한데 대개는 데이터의 전환이나 여러 가지 메타데이터 소스중 최선의 것을 선택한다거나 요약이나 핵심적인 구조를 변경하며, 하드웨어의 물리적 구조변경에 따라 데이터의 재 format이나 재연산과정, 데이터의 인코딩과 디코딩 및 내부적인 표현 등이 관계된다[4].

메타데이터 시스템은 자율 시스템으로 데이터의 형태에 있어서 결과 데이터를 최종사용자에게 제공하는데 결과데이터는 최종사용자에게 분석이라는 과정이 필요없게 하므로 실제로 경영자의 의사결정에 결정적인 역할을 한다. 그런데 최종사용자의 질의는 정형화되지 않고 문제의 본질은 같되 형식이 다른 형태로 다양하게 나타난다. 따라서 질의의 고도화에 따라 메타데이터의 요약수준이 다르고, 데이터웨어하우스가 정형화된 질의만을 인식할 수 있기 때문에 특히나 메타데이터 마이닝의 아키텍처가 분명해야 한다.

정보기술과 인터넷의 발달로 다양한 메타데이터 사용자의 욕구와 정보력이 커지고 그 욕구가 표준화된 획일적인 형태가 문제가 되는 이유 때문에 메타데이터에 대한 마인드와 가치관의 변화가 표출되므로 메타데이터 마이닝은 사용자의 변화가 아니라 기업전체의 프로세스 리엔지니어링이 된다. 따라서 메타데이터마트에서 데이터관리을 위한 기본단위를 최소단위로 구축함으로써 최소단위의 데이터분석 및 branch속성분석이 가능하다. 또한 메타데이터의 이용 속성을 계층별 최종사용자 등을 분석하여 계층별 경쟁력을 확인하고 장기적인 데이터마이닝전략을 수립하는데 필요하게 된다.

메타데이터의 다중 플랫폼이 전형적으로 데이터웨어하우스 프로젝트 내에 포함되므로 필요한 메타데이터 문서는 내포된 마이닝 시스템에 따라 지수적으로 증가한다. DW-최종사용자중심의 메타데이터마이닝의 플랫홈은 다음의 사항들을 포함해야 한다.

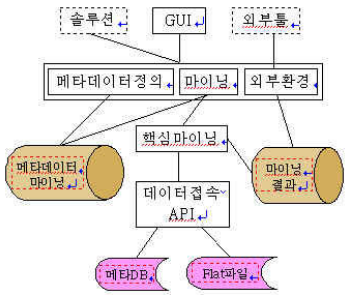
첫째, 데이터구조형 메타데이터 형식에서 DA-콘텐츠 문서화의 소스시스템의 데이터구조와 웨어하우스 및 메타데이터마트의 데이터구조에 관한 사항을 고려되어야 하고,

둘째, 메타데이터의 매핑(mapping)절차로서 패턴추출과정에 관한 소스와 변환데이터에 해당된다.

셋째, 프로세스 메타데이터의 추출, 변환 및 부하과정에 관한 문서화를 통제하고 관리하도록 한다. 이것은 메타데이터의 통합과 준비를 위해 소스시스템에서 데이터저장으로 이동하는 과정과 웨어하우스 데이터구조와 일치하는 형식으로 데이터를 통합하려는 것이다.

메타데이터 마이닝 솔루션은 고도의 정보서비스라는 데이터마이닝의 본래 목적을 충족시키도록 데이터베이스에 직접 접속하는데 시각적인 분석능력과 대용량의 메타데이터를 처리하는 해석력이 우선되어야 한다. 특히 메타데이터의 품질에 대한 민감도가 높고, 통합된 기술과 다양한 마이닝 툴과 엔진을 활용해야 하는 필요성이 있으며 이러한 전제사항들은 기업의 데이터베이스의 역할과 기능의 효율성을 제고하는 요인들로 작용하게 된다. 데이터 라이프의 변화에 따른 설계와 구축을 적절하게 수행하므로써 데이터마트에 대한 기업전략의 분명한 목표, 구축시스템의 선택 및 고객정보의 구성항목등을 선택한다. 특히 고객정보의 구성함목은 기업의 정보이용범위에 따라 다양하게 선택하는데 대개는 고객의 라이프 스타일을 기준한 분류가 명확성이 높아 많이 활용된다.

(그림 2)는 메타데이터 마이닝 시스템의 시스템 아키텍처를 보여주는 그림이다.



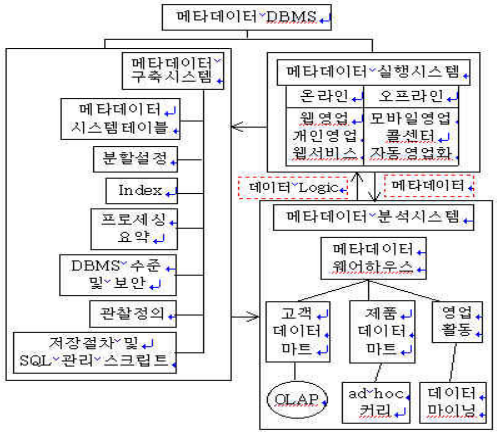
(그림 2) 메타데이터 마이닝시스템 아키텍처

서버 상에서 실행되는 마이닝 알고리즘은 데이터 소스로 close된다. 최종사용자는 동일한 워크스테이션이나 상이한 클라이언트 기계 상에서 실행할 수 있는 GUI를 통한 시스템에 관심을 갖는다. 사용자가 마이닝 운영의 외부결과를 소프트웨어 선택으로 해석할 수 있도록 open된 API가 있는데 메타데이터 마이닝 시스템 아키텍처의 중요한 관점은 이의 I/O 아키텍처이다. 모든 접속 상태를 입력으로 정의한 표준적인 stream 인터페이스가 있으며 데이터 저장명세의 알고리즘 코드를 격리시키는데 데이터접속 API에서 보호된다. 그러므로 새로운 데이터저장형태를 메타데이터 마이닝 시스템으로 추가하기가 쉽다.

핵심마이닝은 구체적인 소스형태와 copy형태의 내역과 부분 소스데이터 내역의 특성을 파악하거나 검사된 메타데이터 마이닝에 대한 형식과 관계형 소스데이터의 명세를 구분하게 된다. 또한 시각적인 그라프 소스내역의 표현과 URL 소스의 식별이 가능토록 한다.

따라서 소스 데이터의 이용에 관한 부분적인 제한과 각 소스의 갱신된 이용정도를 통해서 자동화된 추출도구를 설정하여 시스템의 비용을 최적화하게 된다.

메타데이터 마이닝의 활용은 기업전체의 경영데이터 마이닝의 영역에 적용되는데 대개 메타데이터의 확보, 개발 및 유지의 형태에 따라 DBMS의 검토사항과 분석기법을 이용한다. DBMS는 시스템 테이블과 프로세싱의 요약, 수준 및 보안 등으로 세분되어지며, 메타데이터의 확보나 개발에 대한 커리와 마이닝분석은 데이터의 세분화 및 성향과 선호도가 이용되는 반면에 마이닝 실행시스템은 데이터 사용자의 인터페이스를 자율적으로 하여 데이터관리를 효율화시킨다. 최종사용자에게 차별화된 웹 콘텐츠를 제시하는 개별화시스템, 현장에서의 효율성을 높이는 모바일 컴퓨팅과 전화를 통한 CTI 콜센터 및 영업활동 자동화시스템 등을 운영하는 것이다. DBMS의 이 과정을 정리한 절차가 (그림 3)과 같다.



(그림 3) DBMS 시스템 과정

3. 메타데이터 마이닝시스템의 효율성 제고

메타데이터의 운영환경은 비정형적 환경이므로 기업의 효율성의 문제가 제기된다. 따라서 체계화된 메타데이터 지원환경을 구축하므로써 기업환경의 능동적인 변화에 적응하도록 정형성과 자율성 및 창조성이 제고되어야 한다. 소스데이터의 다양성, 메타데이터의 갱신가능성 등은 새로운 운영시스템의 효율성을 높이는 전략이 될 수 있는데 특히 다음의 요소들이 고려됨이 바람직하다.

첫째, 외부환경의 분석전략은 분자식 유동 데이터베이스 모델로 전환한다. 이 모형은 내부에 메타데이터의 운영체계 명령어를 미리 입력하여 그 운영환경에 맞게 메타데이터를 조작하는 것인데 다차원 데이터베이스와 관계
형 데이터베이스의 특성을 통합한 베이스체계가 된다.

둘째, 전략적 데이터 지향성의 분석전략을 제고한다. 특히 메타데이터의 스코어링과 세분화에 따른 효율성의 제고는 데이터마이닝 솔루션의 이용가능성과 활용성이 높아진다.

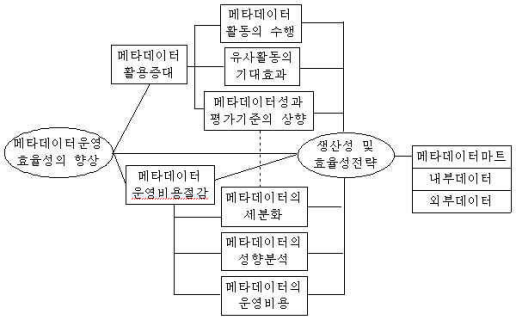
셋째, 메타데이터의 애플리케이션 라이프사이클을 적절히 분석하는 전략이 필요하다. 메타데이터의 운영과 정보환경의 중요성에 따라 정보시스템의 최종사용자들에게 활용되는 역할을 수행하는 것이다. 따라서 메타데이터를수집하여 유용한 상태로 변환시키거나 이동에 관한 조정문제를 고려하게 되고, 환경변화에 따른 메타데이터의 관리상의 기술적인 불일치문제를 해결한다.

넷째, 데이터베이스에 수록된 개별 데이터에 통계적 분석과 모델링기법 등을 적용하여 최종사용자와 기업간의 쌍방연계형태를 취하는 메타데이터 체계를 수립한다.

다섯째, 메타데이터 정보컨셉이 정보기술 하드웨어와 소프트웨어가 뒷받침해 주지 못하는 메타데이터 정보기술환경의 전략을 수립하여 질적 및 양적 정보 활용도를 제고한다. 전략적 자원배분의 우선 순위화의 최적화는 효율적인 경영가치의 관리가 되므로 투자수익률 분석을 반드시 고려해야 한다.

일반적으로 효율적인 데이터마이닝 제고전략을 수립하기 위해서는 먼저 데이터의 최종사용자의 요구조건을 적절하게 정의하는데 기업특유의 데이터웨어하우스 성격을 미리 규정하는 것이다. 예를 들어 메타데이터의 주제영역을 어느 정도 설정할 것인지, 데이터의 통합은 어떻게 할 것인지, 또한 어느 정도의 기간에 해당하는 데이터를 가져올 것인지 여부 등을 결정하는 것이다. 다음으로 하드웨어의 요소기술과 관련된 DBMS나 데이터마트에 대한 통합환경의 구성요소 및 도구를 선택한다. 데이터웨어하우스 설계의 사항으로 데이터모델링과 운영기능과제를 분석해야 한다. 데이터의 추출 및 변형간격, 데이터의 재정리간격과 자원관리와 보안문제, 시스템 장애시 해결 및 복구방안 등의 장애대비책도 수립한다.

이러한 메타데이터의 효율성 제고전략을 설정하는 과정을 정리하면 (그림 4)와 같다.



(그림 4) 마이닝 효율성 제고전략 과정

4. 결론과 과제

메타데이터 마이닝 시스템의 활용은 데이터웨어하우스에 대한 최종사용자의 만족도나 데이터의 모니터 및 유틸리티의 실행, 메타데이터의 신선도와 안정적인 질의 및 개발된 데이터모델의 호환성 등을 제고시킬 수 있을 뿐만 아니라 실제 데이터 정보기술환경의 전략수립이나 효율성을 향상시키게 된다. 특히 데이터 자산의 효율적인 사용에 따른 정보마인드의 확산과 정보해석권의 평등화, 데이터 업무협력 수행에 의한 시너지효과를 기대한다.

이러한 기대효과는 고객관계관리나 공급망관리, 전사적 품질관리, 제품관리 및 영업망관리 등의 경영관리를 포함한 비즈니스 프로세스 리엔지니어링의 최적적인 운영을 수행하므로써 기업경제적 가치의 극대화를 꾀할 수 있다. 또한 웹 애플리케이션의 객체지향기술을

기반으로 하여 확장성과 재사용성뿐만 아니라 관련기술의 표준화나 프로세스 관리기준으로 핵심적인 사업추진과 관련된 메타데이터 마이닝기법의 성과와 향상을 도모하게 되며, 지속적인 경제성이 반영되도록 경영전략과 정보를 생명체화하는 시스템에 활용하게 된다.

참고문헌

[1] 윤용운, “동적 데이터마이닝 시스템의 구조”, 한국경영과학회, 2004년 하계학술대회, 2004.8

[2] 윤용운, “ 전략적 CRM 데이터마이닝 시스템의 관리”, 한국경영과학회 2004년 추계학술대회, 2004.10

[3] R. Ng. and J. Han, Efficient and Effective Clustering Method for Spatial Data Mining, In Proc. of the @20th VLDB Conference, pp.144-155, Santiago, Chile, 1994

[4] W.H. Inmon, Building the Data Warehouse, 2nd. ed., John Wiely & Sons, Inc., 1994