

전자상거래마이닝을 위한 웹데이터베이스시스템의 설계 및 구현

이현호 • 나민영
육군사관학교 전산학과
email : {hhlee, myra}@kma.re.kr

국문 요약

인터넷 사용자의 폭발적인 증가와 더불어 웹을 이용한 전자상거래가 활성화되고 있다. 웹기반 전자상거래시스템은 웹데이터베이스를 이용하여 구축되는데 전자상거래 정보의 효과적인 분석을 위해 데이터마이닝기법이 요구되고 있다. 본 논문에서는 전자상거래 마이닝의 개념을 살펴보고 효과적인 전자상거래 데이터마이닝을 위한 웹데이터베이스시스템을 제안하고 그 프로토타입을 구현하였다. 본 논문에서 제안한 웹데이터베이스는 전자상거래 자체의 내용정보를 저장하는 구조화 데이터 DB와 전자상거래의 사용자 인터페이스를 저장하는 HTML 폼 DB로 나뉘어 전자상거래 자체의 내용정보 뿐 아니라 접속횟수, 접속시간, 원격접속지 등 사용자 인터페이스에서 추출할 수 있는 정보까지 마이닝 대상정보에 포함시켜 효율적인 마이닝환경을 제공할 수 있다.

I. 서론

최근 들어서 인터넷의 발달과 더불어 웹을 이용한 전자상거래가 활성화되고 있다. 1990년대 초반까지만 해도 일부 전문가들만 사용하던 인터넷에서 각종 HTML 용 브라우저의 개발로 비전문가들도 쉽게 인터넷을 통하여 웹문서를 다룰 수 있게 되었다. 웹에서의 정보는 문서들의 집합으로 저장된다. 이 문서는 하이퍼미디어 정보와 하이퍼링크를 포함하여 새로운 문서를 보여주는 기능 등을 제공하며 인터넷의 표준양식인 HTML(Hyper Text Markup Language)로 표현된다. 따라서 웹에서의 데이터는 이 HTML 문서를 통하여 전달되고 유지된다고 볼 수 있다. 전자상거래시스템은 원래 기업간 혹은 기업내 거래업무를 컴퓨터를 통한 정보기술을 이용하여 수행하는 상거래라고 정의할 수 있으나 그 핵심은 역시 HTML 문서(폼)를 이용한 시스템으로 생각할 수 있다. 특히, 인터넷 상에서 지구촌 어디에서든지 쇼핑이 가능하도록 일종의 전자판매시스템을 구축해 놓은 전자쇼핑몰서비스는 전자상거래시스템 분야에서 가장 중요한 관심분야로 부각되고 있다.[9] 이러한 전자상거래시스템에서 전자상거래를 이용하는 고객의 구매행위 분석은 향후 시장확보에 매우 중요한 작업이다. 이러한 전자상거래 정보의 효과적인 분석을 위해서 데이터마이닝 기법이 요구되고 있다.

데이터마이닝 (Data Mining : DM)은 데이터베이스로부터의 지식발견 (Knowledge Discovery in Databases : KDD)이라고도 하는데 대규모의 데이터내에 숨겨져 있는 고급 정보를

추출해서 의사결정, 예측, 예보에 응용하고자 하는 기법[2][8]으로 최근 2000 년대의 데이터베이스 응용기술로 주목을 받고 있는 기술분야이다. 일반적인 데이터마이닝에 관한 연구는 많이 이루어져 왔으나 웹을 기반으로 하는 데이터마이닝에 관한 연구는 아직 초보적인 수준이다.

본 논문에서 다룰 전자상거래마이닝이란 일종의 웹데이터마이닝으로서 HTML 문서(폼)을 통해 전달되는 각종 전자상거래 데이터로부터 필요한 의사결정 정보를 캐내는 것을 말한다. 전자상거래시스템에서 제품정보 뿐만 아니라 전자상거래정보 예를들면, 접속횟수, 접속시간, 원격접속지 등에 따른 고객의 구매행위를 분석하면 더 나은 고객 서비스와 판매실적의 향상을 기대할 수 있게 된다. 전자상거래마이닝을 수행하기 위해서는 우선 웹상에서 전자상거래 데이터가 모아져 웹데이터베이스로 구축되어야 한다.

본 논문에서는 전자상거래마이닝의 개념을 살펴보고 전자상거래마이닝을 수행하기 위한 데이터베이스 시스템 즉 웹데이터베이스시스템을 설계/구현한다. 2 장에서는 전자상거래마이닝에 관하여 살펴보고, 3 장에서는 전자상거래마이닝을 위한 웹데이터베이스시스템 구성 및 설계를 다룬 다음, 4 장에서는 3 장에서 제시된 시스템을 구현한 결과를 보이고 5 장에서 결론을 맺었다.

II. 전자상거래마이닝

2.1 데이터마이닝

데이터마이닝은 대규모 데이터베이스에 존재하는 감추어진 지식을 찾아내는 작업으로써 현실세계에서 쌓여가고 있는 수많은 데이터로부터 유용한 정보를 캐내어 응용하고자 하는 요구에 부응하기 위한 기술이다. 데이터 마이닝의 기본 개념은 새로운 것이 아니라 인공지능 분야의 기계학습(machine learning) 이론에 그 뿌리를 두고 있다. 기계학습에서는 트레이닝 셋(training set)이라 불리는 적은 양의 실험실용 데이터를 사용하여 알고리즘을 만들어 낸다. 그러나 데이터베이스 마이닝은 현실세계의 대규모 데이터베이스를 트레이닝 셋으로 간주해서 이로부터 유용한 지식을 캐내는 작업을 수행하는 것이다. 마이닝에서 얻을 수 있는 지식에는 연관규칙, 분류규칙, 요약규칙, 클러스터링 등 여러 가지가 있다.[1]

- 연관규칙(Association rules) : 연관 규칙은 레코드의 셋에 대하여 아이템의 집합 중에 존재하는 친화도나 패턴을 찾아내는 규칙이다.
- 일반화/요약 규칙(Generalization & summarization rules) :일반화/요약화 규칙은 데이터베이스 내의 사용자가 지정한 부분에 대해 일반적인 특성이나 요약된 고급 뷰를 제공한다.
- 분류규칙(Classification rules) : 데이터 분류란 데이터베이스 내의 객체의 셋에 대하여 그 안에 내재하는 공통 특성을 뽑아내어 이 객체들을 서로 다른 클래스로 그루핑해 내는 작업을 말한다.
- 클러스터링/세그멘테이션(Clustering/Segmentation) : 클러스터링이란 물리적 혹은 추상적 객체를 비슷한 객체군으로 그루핑하는 과정이다.
- 유사성 탐색(Similarity search) : 공간 데이터베이스 등에서 비슷한 패턴을 찾는 일은 위험이나 재난 또는 성향을 발견하고 예측하는데 필수적인 작업이다.
- 순서 패턴(Sequential patterns) : 트렌드를 식별해 내기 위해 일정시간 동안의 레코드를 분석하여 순서 패턴을 찾아낸다.

2.2 전자상거래 데이터마이닝

전자상거래 데이터마이닝은 크게 웹서버에서의 마이닝과 웹클라이언트에서의 마이닝으로 나누어 생각해볼 수 있다. 웹서버에서의 마이닝은 마이닝 모듈이 서버에 있어 각 클라이언트들의 요구를 들어 마이닝을 수행해주거나 클라이언트들의 행위자료로부터 클라이언트 관련정보를 마이닝하는 경우이고, 웹클라이언트에서의 마이닝은 마이닝 모듈이 각 클라이언트에 있어 클라이언트에서 직접 해당 정보를 마이닝하는 경우이다. 본 연구에서 다루는 전자상거래마이닝은 서버에서의 마이닝이나 어느 경우이든 웹데이터마이닝은 HTML 문서에 의한 마이닝을 의미한다.

HTML 문서는 문서에 관한 사항을 포함하고 있는 헤드와 문서의 내용인 정보를 포함하고 있는 본체로 구성된다. 또한 구문적으로는 문서의 형식과 구조를 나타내는 태그에 의해 명시되는데 각각의 동작을 시작하는 선두 태그와 종료시키는 후미 태그의 쌍으로 나타난다. 웹상에서의 정보 전달은 직접 문서를 만들어 전달할 수도 있으나 초보자를 위하여 보통은 폼을 만들어 사용한다. HTML 폼은 보통 입력필드, 선택메뉴, 그리고 부연설명 등으로 구성된다. 사용자는 관련 메뉴를 선택하여 지시된 방법으로 입력 필드를 입력한다. 입력 필드의 내용은 CGI 스펙에 맞게 페러미터화되어 HTTP를 통해 서버로 보내진다. 이 데이터는 서버에서 다시 처리될 수 있는 데이터 형태로 CGI에 의해 복원된다. 실제로 이 데이터는 서버에서 데이터베이스 즉, 웹데이터베이스내에 저장되어 관리된다.

2.3 웹데이터베이스

웹데이터베이스는 일반적인 데이터베이스와 마찬가지로 데이터 또는 자료의 저장소이다.[4] 그러나 웹데이터베이스는 HTML의 백그라운드에서 실행되는 웹 응용 프로그램을 경유하여 액세스하는 데이터베이스이다. 즉 웹 브라우저를 사용하나 사용자가 액세스할 수 있는 응용 프로그램에 데이터베이스를 결합한 형태의 데이터베이스가 웹데이터베이스이다.

웹데이터베이스시스템의 장점으로서 웹데이터베이스시스템과 웹 DBMS 개발물이 여러 플랫폼과 운영체제에 독립적으로 운영되어 글로벌 네트워크인 인터넷 상에서 소프트웨어, 하드웨어와 네트워크 유지보수가 용이하다는 점이다. 예를들어, IBM의 DB2 WWW Connection V1은 DB 구조를 바꾸지 않고 인터넷 상의 모든 사용자가 웹브라우저를 통해 DB2 데이터를 액세스할 수 있다. 또한, 웹 환경에는 사용자와 로컬 인터넷 서버 외에 원격 액세스 서버가 구축될 수 있다. 웹데이터베이스는 원격 액세스 서버에 탑재되어 운영될 수 있는데 이는 회사에서 운영하는 거대한 데이터베이스시스템을 한 장소에서 관리하게 함으로써 데이터베이스의 일치성과 무결성, 데이터관리의 효율성을 높인다.[5]

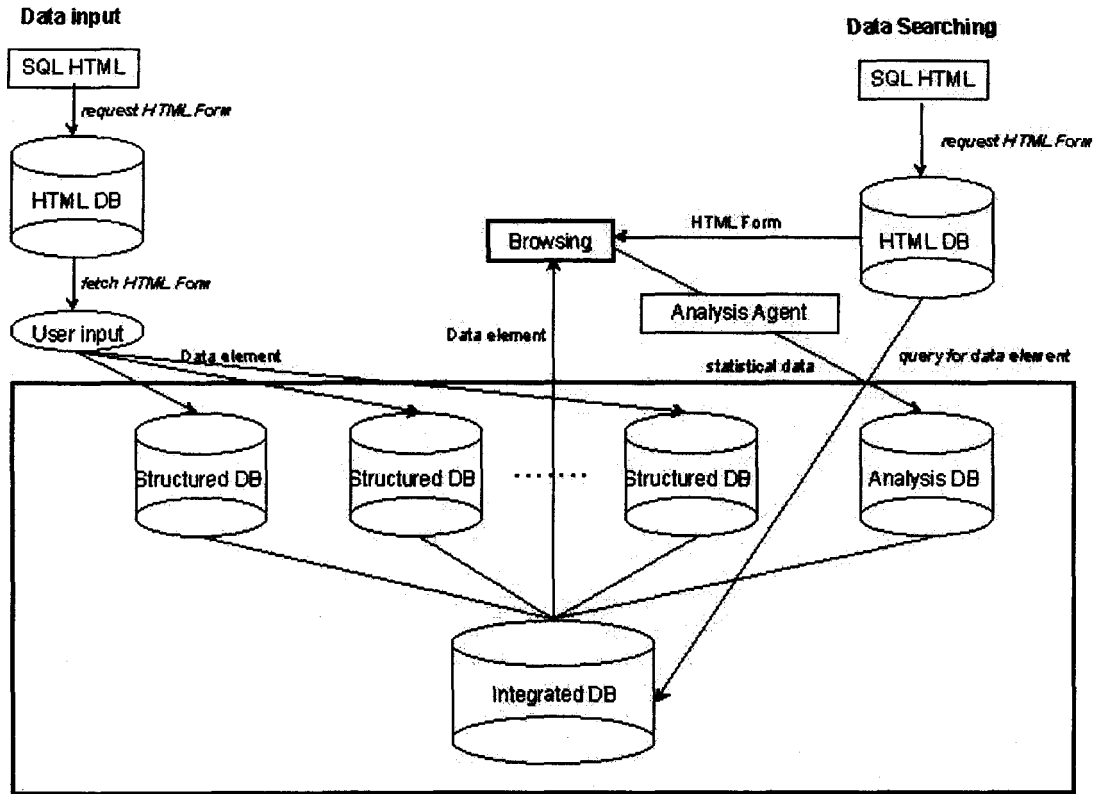
웹데이터베이스는 HTML, CGI, JAVA 등으로 구성된 웹페이지와 연동하여 웹브라우저 상에서 디스플레이 된다. 이는 전세계를 연결하는 글로벌 네트워크인 인터넷 상의 사용자에게 통일된 뷰(View)를 제공하여 정보홍보효과를 극대화시킬 수 있다. 또한 데이터베이스가 저장하는 오디오, 비디오, 그래픽 등 다양한 포맷의 멀티미디어데이터가 웹브라우저라는 한가지 매체를 통하여 디스플레이됨으로써 쉬운 사용자 정보제공환경을 구현한다.

전자상거래마이닝을 위해서는 먼저 웹데이터베이스를 구축하고 HTML 폼으로부터의 입력되는 데이터를 저장 관리할 수 있는 시스템을 갖추어야 한다. 다음 장에서는 전자상거래 정보를 액세스 DB로 저장하는 웹데이터베이스시스템 구축에 관하여 설명한다.

III. 웹데이터베이스시스템 설계

3.1 시스템 구성

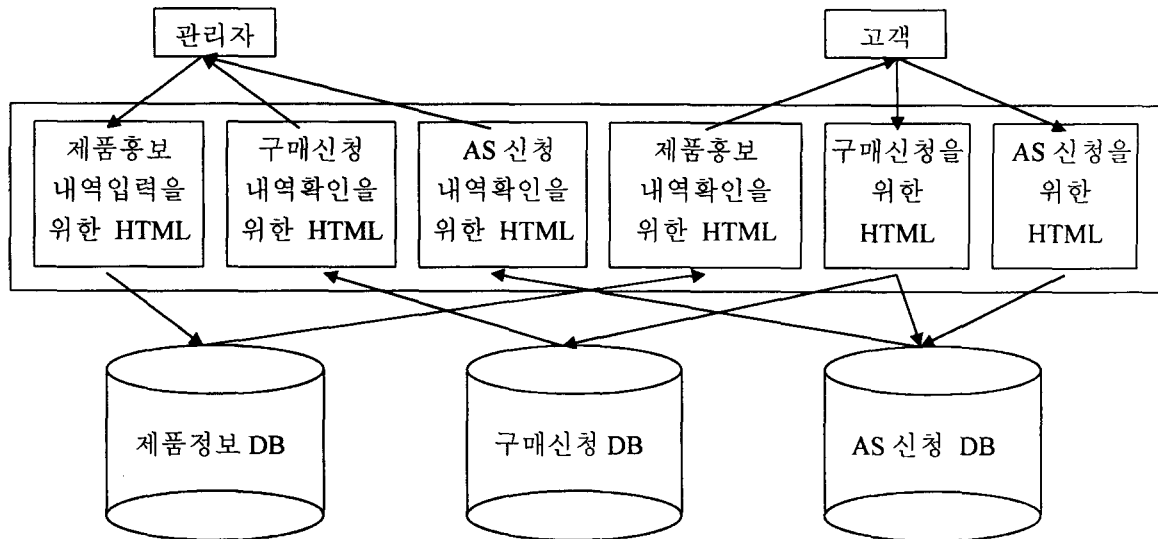
본 논문에서 다루는 전자상거래 마이닝을 위한 웹데이터베이스시스템의 전체적인 구조는 [그림1]과 같다.



[그림 1] 전자상거래 웹데이터베이스시스템의 개략적 구조

웹데이터베이스시스템은 크게 구조화 데이터 DB(Structured Data DB) 부분과 HTML 폼 DB로 나누어 생각할 수 있다. 웹데이터베이스시스템에서의 사용자 인터페이스 부분인 웹브라우저 상의 웹문서는 데이터베이스와 연동되어 있을 경우, 데이터영역과 비데이터영역으로 구분된다. 데이터영역은 데이터베이스에 입력/저장되어 사용자에게 의해 검색되는 부분으로 데이터마이닝의 직접적인 대상이다. 비데이터 영역은 데이터베이스와 직접적인 연관성이 없는 부분으로 사용자에게 데이터의 입력 및 검색에 용이한 환경을 제공하기 위한 폼을 제공한다. 이러한 웹문서와 데이터베이스의 관계적 특성을 고려하여 데이터영역의 데이터베이스화를 위한 구조화 데이터 DB와 비데이터영역의 데이터베이스화를 위한 HTML 폼 DB를 제안한다.[9]

본 논문에서는 컴퓨터 및 주변기기에 대한 전자상거래시스템을 예로 들어 설명한다. 예제 전자상거래시스템은 크게 제품홍보부분, 구매부분 및 AS부분으로 나누어 생각할 수 있다. 각 부분의 정보를 저장하기 위한 데이터베이스를 바탕으로 하여 실제로 전자상거래시스템을 이용하기 위한 고객 및 관리자의 인터페이스가 되는 웹문서에는 제품홍보내역입력을 위한 HTML, 제품홍보내역확인을 위한 HTML, 구매신청을 위한 HTML, 구매신청내역확인을 위한 HTML, AS신청을 위한 HTML, AS신청내역확인을 위한 HTML 등이 필요하다.



[그림 2] 전자상거래시스템에서의 HTML 폼

본 논문이 제안한 구조화 데이터 DB와 HTML 폼 DB를 예제 전자상거래시스템에 적용하면 [그림2]와 같이 제품홍보 DB, 구매신청 DB, AS신청 DB가 구조화 데이터 DB에 해당하고 사용자와 각 부분 DB 간의 인터페이스가 되는 HTML 문서들이 HTML 폼 DB를 구성한다.

구조화 데이터 DB와 HTML 폼 DB라는 데이터베이스 이중구조의 장점으로 첫째, 웹문서의 비데이터영역까지 데이터베이스화하여 데이터베이스에서 전체 웹시스템을 통합관리할 수 있다. 기존의 웹시스템은 웹문서를 구성하는 데이터영역만을 데이터베이스화하고 HTML 문서 자체는 파일시스템을 그대로 이용했다. 이것은 웹구조의 변동이나 유지보수에 어려움을 초래하였다. 이에 비해, HTML DB를 도입함으로써 비데이터영역도 데이터베이스를 기반으로 통합적으로 관리할 수 있다.

둘째, 인터페이스의 일관성을 제공한다. 웹시스템의 구성을 HTML 폼 DB를 통한 웹문서의 비데이터영역 fetch와 구조화 데이터 DB를 통한 웹문서의 데이터영역 fetch라는 두 부분으로 단순화시켜 기존의 하이퍼링크를 이용한 복잡하고 일관성을 잃기 쉬운 웹데이터베이스시스템 구조에 일관성과 통일성을 제공할 수 있다.

셋째, 다양한 HTML 폼템플릿을 제공할 수 있다. 하나의 인터페이스에 정적인 한 형식의 HTML을 매핑시키는 것이 아니라 다양한 형식을 취하는 복수개의 HTML을 매핑시키므로써 동적인 HTML 인터페이스 환경을 구성할 수 있다.

넷째, 분산환경에 능동적으로 대처할 수 있다. 규모가 큰 웹시스템의 경우 데이터베이스가 여러개의 서버에 분산되어 저장된다. HTML 폼 DB를 사용할 경우 각각의 서버에서 관리하는 HTML 폼을 통합관리하여 웹시스템 관리의 용이성을 제공할 수 있다.

다섯째, 효과적인 웹마케팅 환경을 제공한다. 데이터와 폼으로 데이터베이스를 이원화시켜 마케팅의 대상을 명확하게 구분하여 마케팅시스템을 구성할 수 있다. 특히, 인터페이스가 되는 폼을 데이터베이스로 관리하여 웹인터페이스에도 마케팅 기법을 도입할 수 있는 기초를 마련한다.

3.2 HTML 폼 DB 설계

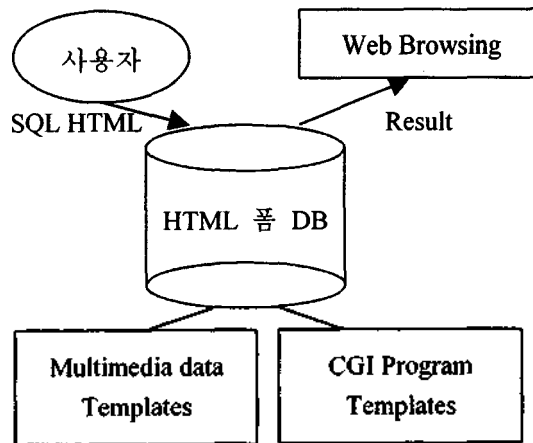
[그림 3]은 본 논문에서 구축한 전자상거래시스템에서의 HTML 품 DB이다.

1	입력	홍보내역입력	ad_input.html	제품의 홍보내역 입력을 위한 폼	8
2	입력	제품정보입력	product_input.html	새로운 제품정보를 입력하기 위한 폼	3
3	입력	취급부서입력	part_input.html	제품정보를 취급하는 부서정보 입력	3
4	리스트	홍보내역리스트	ad_list.pl	제품의 홍보내역 리스트 폼	8
5	리스트	제품정보리스트	product_list.pl	제품정보에 관한 리스트 폼	17
6	리스트	취급부서리스트	part_list.pl	제품정보를 취급하는 부서정보에 관한 리스트 폼	3
8	입력	구매내역입력	sales_input.html	고객의 구매신청내용 입력을 위한 폼	8
9	입력	고객정보입력	customer_input.html	고객정보 입력을 위한 폼	4
10	입력	판매담당자입력	salesman_input.html	제품 판매담당자 입력을 위한 폼	4
11	리스트	구매내역리스트	sales_list.pl	구매내역을 확인하기 위한 리스트 폼	11
12	리스트	고객정보리스트	customer_list.pl	고객정보를 확인하기 위한 리스트 폼	5
13	리스트	판매담당자리스트	salesman_list.pl	판매담당자를 확인하기 위한 리스트 폼	7
14	입력	AS신청내역입력	as_input.html	AS신청 접수를 위한 입력 폼	3
15	입력	AS담당자입력	asman_input.html	AS담당자 입력을 위한 폼	5
16	리스트	AS신청내역리스트	as_list.pl	AS신청내역을 확인하기 위한 리스트 폼	12
17	리스트	AS담당자리스트	asman_list.pl	AS담당자를 확인하기 위한 리스트	5

[그림 3] HTML 품 DB의 예

위의 구성필드 중 주목할 만한 것은 ‘접속횟수’ 필드이다. 이는 관리자나 고객이 얼마나 많이 해당 HTML 품을 액세스했는가를 기록한 필드이다. 이것은 고객의 전자상거래시스템 활용형태를 분석하는데 중요한 자료로서 데이터마이닝의 대상이 된다. 이와 같이, HTML 품 DB를 구축하는 것은 웹인터페이스에도 마이닝기법을 도입할 수 있다는 근거를 마련한다.

그런데, 위와 같은 데이터베이스 구조만으로 HTML 품을 구성하는 모든 요소를 다 포괄할 수는 없다. 왜냐하면, HTML 품에는 이미지, 음성, 동화상 등의 멀티미디어 데이터와 실질적인 데이터베이스 연동을 담당하는 CGI 프로그램이 포함될 수 있기 때문이다.[3] 그러나, 멀티미디어데이터와 CGI 프로그램까지 데이터베이스화하는 것은 데이터베이스 구조를 복잡하게하여 유지보수에 오히려 오버헤드를 줄 수 있으므로 멀티미디어데이터 템플릿과 CGI 프로그램 템플릿을 만들어 파일시스템으로 관리하는 것이 더 효율적이다. 전체적인 HTML 품 DB의 구조는 [그림 4]와 같다.

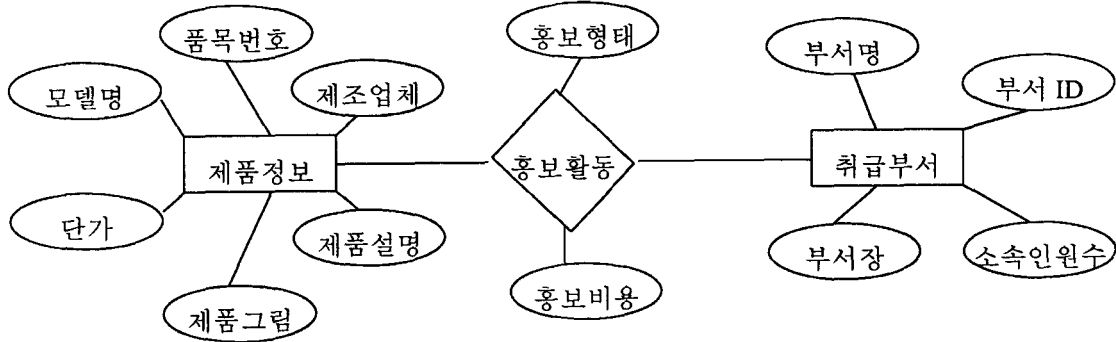


[그림 4] HTML 품 DB의 전체적인 구조

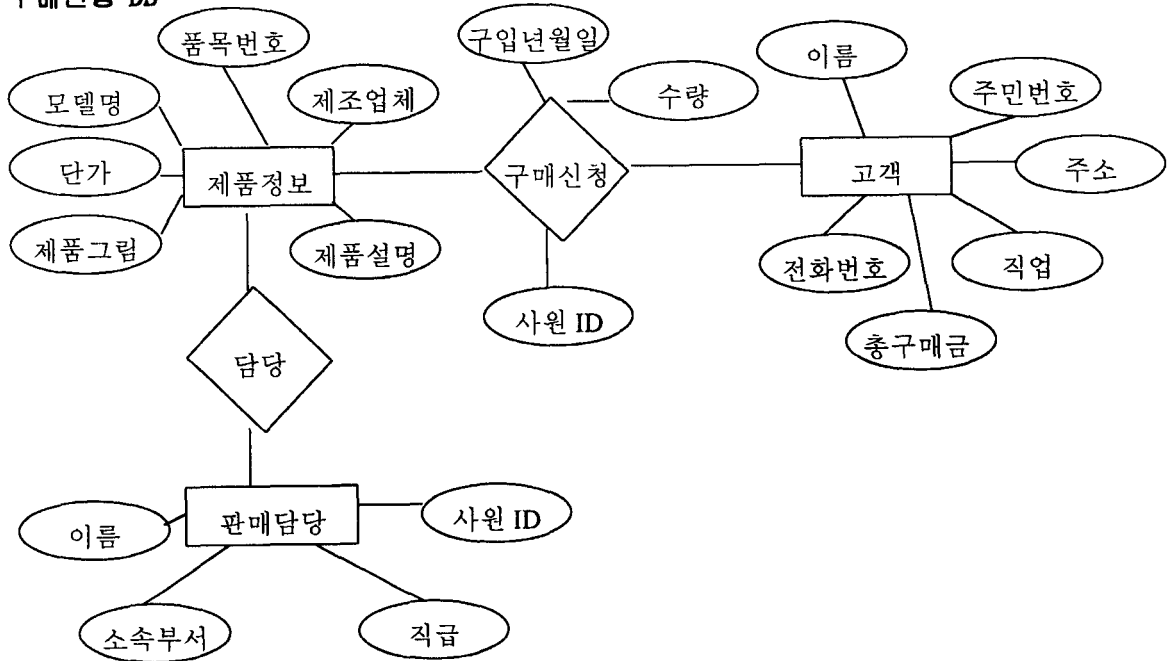
3.3 구조화 데이터 DB 설계

구조화 데이터 DB는 일반적인 DB연동 웹시스템에서의 데이터베이스 구조의 내용을 가지는 데이터베이스로 실제적인 마이닝의 대상이 된다. 예제 전자상거래시스템에서 구조화 데이터 DB를 구성하는 각각의 데이터베이스의 스키마구조는 [그림 5]와 같다.

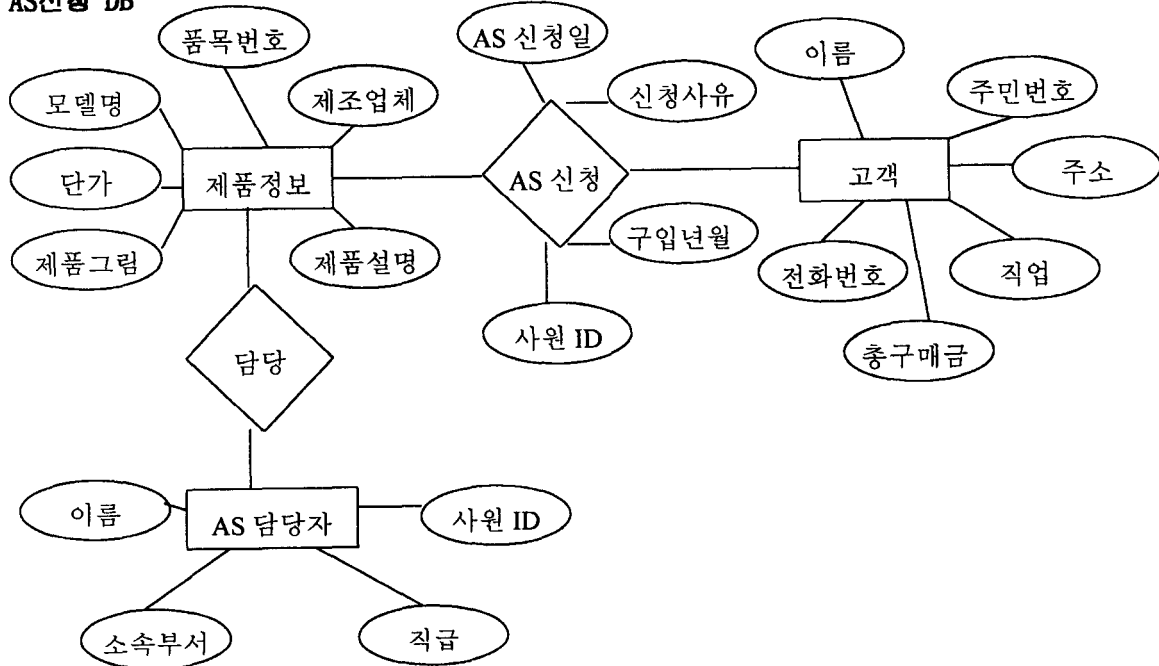
제품홍보 DB



구매신청 DB



AS신청 DB

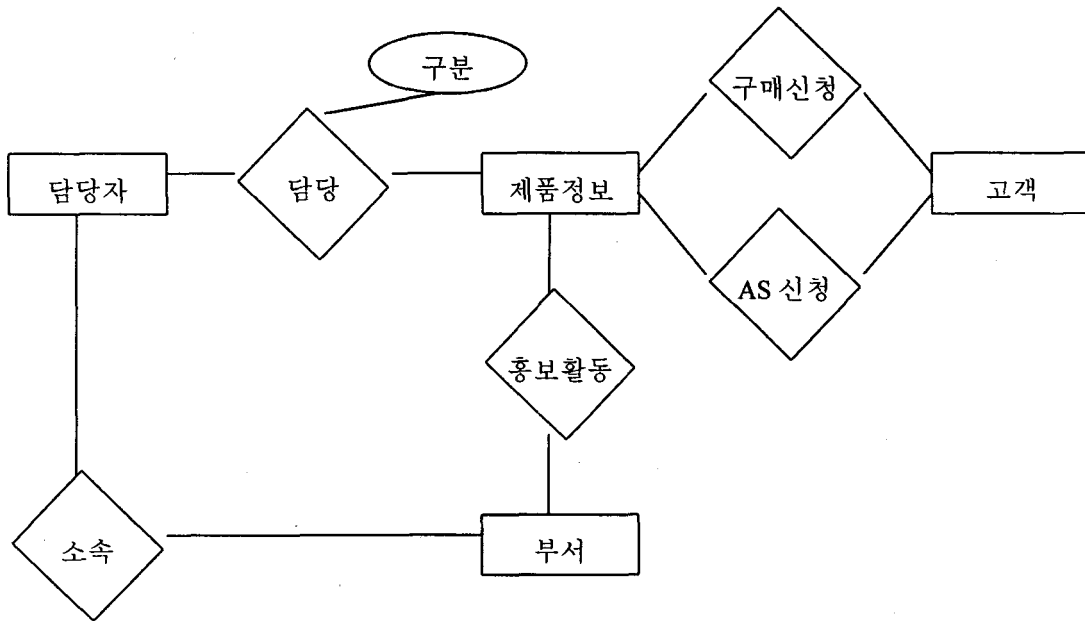


[그림 5] 개별 구조화 데이터 DB의 스키마구조

3.3.1 통합데이터베이스(Integrated DB) 설계

위의 데이터베이스를 기반으로 하는 데이터마이닝을 위해서는 각각의 데이터베이스가 통합된 통합데이터베이스가 필요하다. 전자상거래시스템에서 사용하는 각각의 데이터베이스 즉, 제품정보 DB, 구매신청 DB, AS신청 DB의 스키마는 모두 ER(Entity Relationship)모델을 기반으로 하므로 스키마 통합을 통하여 하나의 통합된 스키마구조를 생성할 수 있다. 스키마 통합은 사전통합(preintegration), 스키마 분석(이름충돌분석, 구조충돌분석), 합병, 재구성 및 문서화 등의 단계를 거쳐서 이루어진다.[7] 제시한 세 개의 스키마를 살펴보면 제품정보 DB에서 '취급부서'는 엔티티로 표현되어 있고 구매신청 DB와 AS신청 DB에서는 '소속부서'라는 이름의 애트리뷰트로 표현되어 있어 구조충돌이 일어난다. 이를 해결하기 위해서 구매신청 DB와 AS신청 DB의 '소속부서' 애트리뷰트를 전역 레벨에서 '부서'라는 이름의 엔티티로 표현하는 것이 적합하다.

또한, '판매담당자' 엔티티와 'AS담당자' 엔티티는 같은 애트리뷰트로 구성된 동일 세트에 보아 이것을 일반화한 '담당자' 엔티티를 생성하고 판매담당자와 AS담당자의 구분을 위해 '구분' 애트리뷰트를 생성시켜 최종적인 통합스키마를 생성하면 [그림 6]과 같다. 본 연구에서 통합데이터베이스는 앞으로 데이터웨어하우스로의 발전을 위하여 물리적으로 유지된다.



[그림 6] 통합된 구조화 데이터 DB

3.3.2 웹데이터마이닝 대상정보(target information) 추가

웹데이터베이스 마이닝의 특성을 가지기 위해서는 일반 데이터베이스와는 구별되는 웹환경에서 마이닝 대상정보로서 유용하게 사용될 수 있는 정보가 데이터베이스에 추가되어야 한다. 이는 웹환경이 가지는 특수성에 기인하여 추출되는 정보로서 다음과 같은 것들이다.

- 웹접속횟수(the number of web connections) : 사용자 인터페이스인 HTML 폼 별로 접속횟수를 계산하여 저장하는 것은 추후 인터페이스 개선을 위한 의사결정과 전자상거래 마케팅 전략에 결정적인 도움을 줄 수 있으므로 웹접속횟수는 유용한 마이닝 대상정보이다.
- 웹접속시간(web connection time) : 웹은 서버가 동작을 멈추지 않는 한,

접속시간의 제한이 없다는 특징이 있다. 전자상거래시스템의 경우, 웹접속시간을 분석하여 시간대별 고객의 구매행위 형태를 파악하고 다양한 마케팅 전략을 구사할 수 있으므로 웹접속시간은 유용한 마이닝 대상정보가 된다.

- 원격호스트이름(remote host name) : 웹의 구동기반이 되는 인터넷은 글로벌 네트워크의 성격을 지닌다. 전자상거래시스템의 경우, 원격호스트이름은 지역별 구매행위를 파악하는데 필수적인 정보로서 웹만이 가질 수 있는 유용한 마이닝 대상정보이다.
- 원격사용자이름(remote user name) : 웹시스템에서는 접속 ID를 사용자에게 부여하여 제어할 수 있다. 전자상거래시스템의 경우, 원격사용자이름 분석을 통하여 고객의 나이별, 직업별 구매행위 형태 파악이 가능하므로 원격사용자이름 또한 마이닝 대상정보로서의 가치가 충분하다.

이와 같이, 웹환경에서 추출할 수 있는 유용한 정보를 마이닝 대상에 포함시킴으로써 일반 데이터베이스 마이닝과 구별되는 웹데이터베이스 마이닝의 독자적 영역을 구축하고 웹환경의 장점을 충분히 활용할 수 있다.

본 논문에서는 마이닝 대상정보로서 웹접속시간과 원격호스트이름을 채택하여 시스템에 구현하였다. 원격호스트이름은 숫자주소로 되어 있으나 DNS를 통하여 문자주소의 변환이 가능하다.

[그림 7]은 본 논문에서 구현한 구매신청내역확인을 위한 리스트 폼의 일부이다.

번호	모델명	사양	구분	수량	구매일	구매시간	구매IP
3	SENS 520 PA	610912-1643526	김영석	구본근	1	4/11/98, 19:11:3	210.103.9.25
6	SENS 520	610912-1643526	장현진	장수정	1	4/14/98, 9:32:11	120.130.2.32
7	HP DesignJet 60K	600104-1630211	민정철	김현진	5	4/16/98, 11:56:22	165.132.34.32
2	HP NetServer 1100	610912-1630211	장영식	알진희	1	4/20/98, 10:16:41	165.135.23.45
6	SENS 520	610912-1643526	김영석	배승수	1	4/26/98, 10:41:16	210.103.9.34

[그림 7] 구매신청내역확인을 위한 리스트 폼

IV. 웹데이터베이스시스템 구현

4.1 개발환경

본 논문에서 개발한 전자상거래시스템 프로토타입의 개발환경은 다음과 같다.

- 플랫폼 : Microsoft 95/NT 4.0
- 웹서버 : Microsoft Personal Web Server
- 데이터베이스 : Microsoft Access 97
- 웹문서 저작도구/브라우저 : 나모 웹에디터/Netscape Communicator 4.0
- 개발기술 : Perl CGI, ODBC(Open DataBase Connectivity)

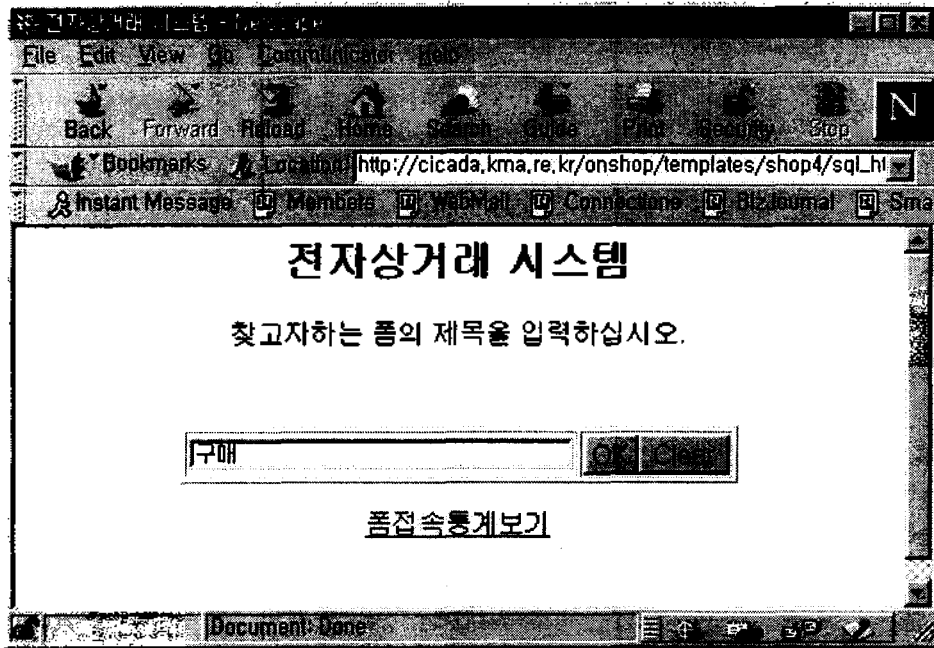
본 논문에서 구현한 시스템은 ODBC라는 개방형 데이터베이스 연동기술을 이용했으므로

연동 데이터베이스를 서버급으로 쉽게 교체할 수 있다.

4.2 전자상거래시스템

개발된 시스템은 3장에서 제시된 설계를 따라 구현되었다. 본 논문에서 다루고자 하는 초점은 전자상거래를 위한 웹인터페이스 설계가 아니라 웹데이터베이스를 구축하는 것이므로 개발된 시스템의 웹인터페이스는 그 기능이 간략하게 축소 구현되었음을 미리 밝힌다. 이 절에서는 전자상거래시스템을 통한 고객의 구매신청이 실제로 어떻게 이루어지는지를 단계적으로 살펴본다.

1) 품질의 실행(SQL HTML)

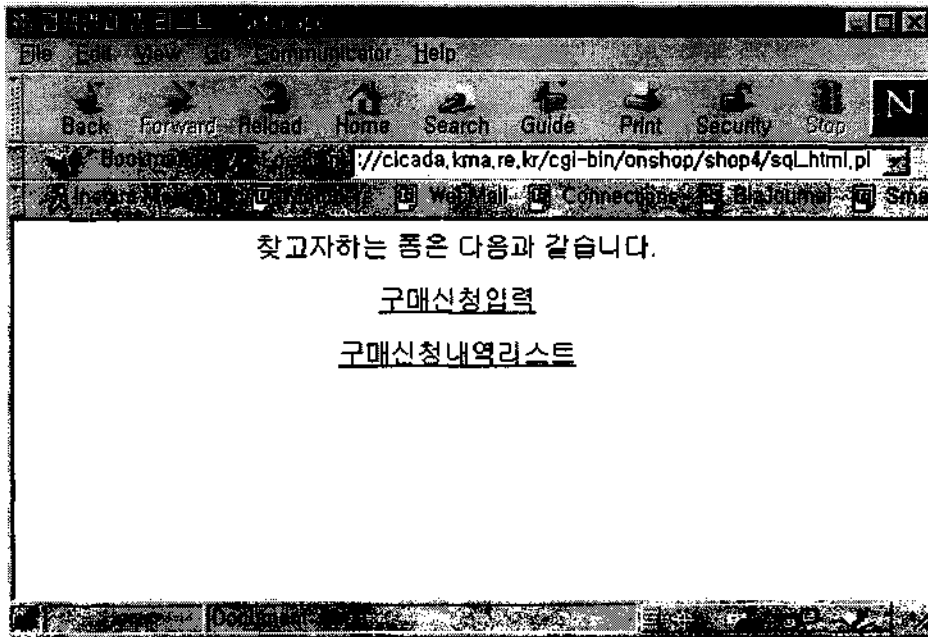


[그림 8] 전자상거래시스템의 초기화면

고객은 구매신청을 위해 메인(main) 페이지에 접속하여 구매신청에 필요한 품을 찾는다. 이것은 본 논문의 2장에서 설명한 시스템의 전체적인 구조에서 SQL HTML에 해당하는 부분으로 찾고자하는 품제목의 일부를 입력하고 'OK'버튼을 누르면 품을 찾기위한 SQL이 실행된다. 다음은 SQL 실행을 위한 프로그램코드이다.[6]

```
.....  
# sql을 통해 사용자가 입력한 keyword에 match되는 html 품을 찾는다.  
if ($runtime eq "search")  
{  
    $sql = " select *";  
    $sql .= " from HTML_FORM";  
    $sql .= " where 제목 Like '%$keyword%'";  
  
    $query->Sql($sql);  
  
    print "<head>";  
    print "<title>검색결과 품 리스트</title>";  
.....
```

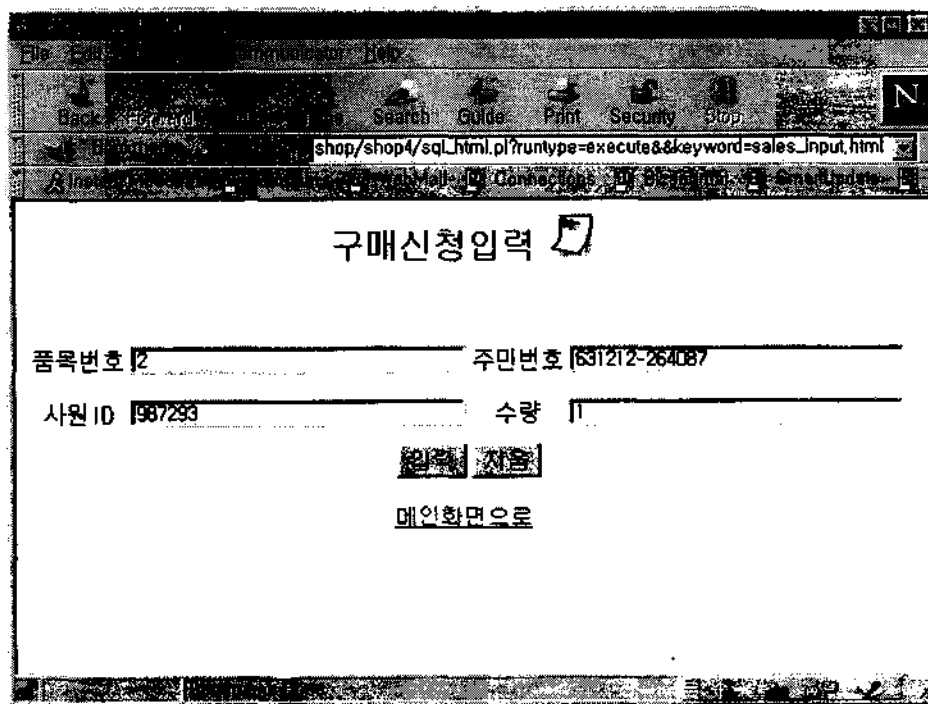
2) 품질의 실행결과 리스트



[그림 9] 품질의 실행결과 리스트

질의에서 서브스트링비교(substring match)가 가능하므로 질의결과 품은 여러 개일 수 있다. 여기서는 고객이 구매신청을 한다고 가정했으므로 구매신청내역을 선택한다.

3) 구매신청폼에 구매신청내역



[그림 10] 구매신청폼에 구매신청내역

고객은 구매신청내역 폼에서 제품품목번호, 고객주민번호, 담당사원ID, 수량을 입력한 다음 '입력' 버튼을 누름으로써 구매신청을 하게 된다.

이로써 고객의 구매신청은 끝나게 된다. 예에서, 백그라운드에서 데이터베이스와의 연동부분은 두 곳에서 나타난다. 첫번째는 앞의 1), 2)에서 나타나는데 원하는 품을 찾는

부분으로 HTML 폼 DB와의 연동이다. 두번째는 앞의 3)에서 나타나는데 고객의 구매신청정보가 구조화 데이터 DB에 입력된다. 고객의 구매신청은 다음과 같은 결과를 낳는다.

① 웹데이터베이스에 신청정보가 입력된다.

구분	상품명	모델명	수량	가격	합계	비고
1	SPC 8000 VA	10000-10000	1	41100	41100	
2	SENS 120	11111-11111	1	41000	82100	
3	HP DESIGN SUK	12222-12222	1	41000	123100	
4	HP DESIGN SUK	13333-13333	1	42000	165100	
5	SENS 120	14444-14444	1	42000	207100	
6	SENS 120	15555-15555	1	52000	259100	

[그림 11] 웹데이터베이스에 신청정보가 입력된 결과화면

제일 마지막 레코드의 정보가 입력된 신청정보이다. 구매시각과 구매서버이름은 백그라운드 프로그램에서 자동적으로 찾아 입력된다.

② 구매신청입력 품의 접속횟수가 1 증가한다.

구매신청입력 품의 접속횟수를 주목하면 접속횟수가 1증가 되어 있음을 알 수 있다. ([그림 12])

신청전 :

신청후 :

구분	상품명	모델명	수량	가격	합계	비고
1	SPC 8000 VA	10000-10000	1	41100	41100	
2	SENS 120	11111-11111	1	41000	82100	
3	HP DESIGN SUK	12222-12222	1	41000	123100	
4	HP DESIGN SUK	13333-13333	1	42000	165100	
5	SENS 120	14444-14444	1	42000	207100	
6	SENS 120	15555-15555	1	52000	259100	

[그림 13] 구매신청행위 전후 구매신청입력 품의 접속횟수

V. 결론

본 논문에서는 전세계적 글로벌 네트워크로 자리를 잡은 인터넷 상에서 전자상거래 데이터마이닝 환경구축을 위한 웹데이터베이스시스템을 설계/구현해 보았다. 웹환경에서 전자상거래를 실현하기 위해서는 웹데이터베이스의 구축이 필수적이다. 나아가 보다 효율적인 전자상거래가 이루어지기 위해서는 사실데이터의 축적/검색 뿐 아니라 사실데이터를 바탕으로 고급 정보를 추출하여 의사결정, 예측 등에 이용할 수 있어야 한다. 이를 가능하게 하는 것이 데이터마이닝 기술이며 마이닝 기법을 효과적으로 적용하기 위해서는 웹환경에서 유용하게 이용될 수 있는 마이닝 대상 정보를 잘 정의하고 이러한 정보들을 효율적으로 구조화 시킬 수 있는 웹데이터베이스 스키마 설계가 필수적이라 하겠다.

본 논문에서는 이러한 시도로서 데이터베이스를 HTML 폼 DB와 구조화 데이터 DB로 구분하여 설계하였다. 구조화 데이터 DB는 주로 전자상거래 자체의 내용정보를 저장하였고 HTML 폼 DB는 웹환경에서 전자상거래의 사용자 인터페이스가 되는 HTML 폼을 저장하였다. 이는 사용자 인터페이스를 데이터베이스화 하여 데이터마이닝의 대상에 포함시킴으로써 사용자 인터페이스가 중요한 의미를 갖는 웹환경에서 보다 효율적인 마이닝환경을 제공할 수 있다. 또한, 시스템 구축 예에서 보인 '구매시각, 구매서비스름' 등 웹환경에서 전자상거래 이용형태분석에 중요한 자료로 활용될 수 있는 정보를 구조화 데이터 DB에 저장함으로써 보다 다양한 각도에서의 데이터마이닝을 가능하게 하였다.

본 논문은 전자상거래 마이닝을 위한 환경 구축에 중점을 둔 것이므로 이를 바탕으로 실제적으로 마이닝을 적용하기 위한 마이닝 대상정보의 다양화, 데이터웨어하우스의 구축 및 마이닝 기법의 도입/적용 등이 향후과제로 남는다. 데이터웨어하우스는 기존의 운영 데이터베이스와 분리하여 최적화된 형태로 의사결정의 지원 및 사용자들의 요구사항을 충족시키고자 하는 의도에서 제시된 데이터베이스 기술이며, 현재 상태의 데이터뿐만 아니라 과거의 데이터도 유지하여 누적된 통합 데이터를 분석하여 필요한 정보를 추출하는데 효과적이다. 본 논문에서 생성 제시된 통합데이터베이스는 물리적인 통합데이터베이스이므로 이를 기반으로 하면 데이터웨어하우스로의 발전이 용이할 것이다. 향후과제로서 또한 생각할 수 있는 것은 본 논문에서 구현한 시스템은 로컬 시스템 환경에서 프로토타입의 형태로 개발한 것이므로 실용화를 위해서는 향후 서버급 데이터베이스시스템을 도입하여 실질적인 클라이언트/서버 환경을 갖춘 시스템으로 업그레이드하는 것이다.

참고문헌

- [1] Chen, M. S., Han, J., and Yu, P. S., "Data Mining: An Overview from Database Perspective," IEEE TKDE, Vol. 8, No. 6, 1996.
- [2] Fayyad, U. M. et al. Advances in Knowledge Discovery and Data Mining, MIT Press, 1996.
- [3] Jeffrey Dwight, Michael Erwin, Using CGI, p. 70 – p. 90, Que Corporation, 1995.
- [4] Korth, H. F., and Silberschatz, A., DataBase System Concepts, McGraw-Hill, 1997.
- [5] Marianne Winslett, "Databases and the World Wide Web", <http://cdr.cs.uiuc.edu/pubs/de97tutorial/index.htm>, University of Illinois , Version of April 15, 1997.
- [6] Mike Glover, Ardan Humpbreys, Ed Weiss, PERL 5 How-To, Waite Group Press, 1996
- [7] 나민영, 데이터베이스 설계, p.198 – p.214, 기한재, 1996년 7월.
- [8] 데이터 마이닝, 1998 년도 춘계 튜토리얼, 데이터베이스연구회, 1998.
- [9] 손주찬, 김중배, 함호상, "전자거래정보 검색 플랫폼 모델링에 관한 연구", 한국 CALS/EC 학회지, 제1권1호, 1996.