碩士學位論文

웹 로그분석을 통한 인터넷 쇼핑몰에 관한 연구

濟州大學校 産業大學院

電子電氣工學科

컴퓨터工學 專攻

元　明　美

2008

碩士學位論文

웹 로그분석을 통한 인터넷 쇼핑몰에 관한 연구

指導敎授 郭 鎬 滎

濟州大學校 産業大學院

電子電氣工學科

컴퓨터工學 專攻

元 明 美

2008

웹 로그분석을 통한
인터넷 쇼핑몰에 관한 연구

指導敎授　郭 鎬 榮

이 論文을 工學 碩士學位 論文으로 提出함.

2008年 8月

濟州大學校 産業大學院
電子電氣工學科 컴퓨터工學專攻

元　明　美

元明美의 工學 碩士學位 論文을 認准함.

2008年 8月

委員長　　印
委　員　　印
委　員　　印

목 차

표 목 차

그 림 목 차

Study on the Internet Shopping Mall Using Web Log Analysis

Myeong Mi Won

Department of Electronic and Electrical Engineering

Graduate School of Industry

Cheju National University

Supervised by Professor Ho Young Kwak

Summary

As users to purchase commodities and search a web page through the internet increases rapidly, many web sites take a growing interest in not a simple site embodiment to gather a user information but a more effective site embodiment to extract a valid date from a user information.

Whenever a user searches a file in website, the server software leaves the recording. The server stores a text information called Acess log, Error log and Referrer log.

We can extract the data of a user like main class, customer purchase pattern, main purchase time, web page search path and so on from the analysis of stored web log file and use various ways as interface design.

I have researched the structure of web log file, the web log server's file creation.

In result, I presented the strategic way of Internet shopping mall with the analysis of the consumer's behavior.

Ⅰ. 서 론

인터넷은 언제부터인가 우리 생활에 깊숙이 파고들고 있다. 인터넷이 급속도로 우리 생활 속에 파급되는 이유는 인터넷이 생활에 필요한 다양한 정보를 제공할 뿐만 아니라 거대한 단일 네트워크로 전 세계와 연결되어 있어 빠르고 신속하게 커뮤니케이션을 가능하게 해주기 때문이다. 또한 효과적인 자원과 지식 발견을 위해 풍부하고 동적인 하이퍼링크 정보와 웹페이지 접근과 사용정보를 포함하여 뉴스, 광고, 소비자 정보, 금융관리, 교육, 정보, 전자상거래 등을 위한 정보 서비스의 중심이 되었기 때문이다.

인터넷 상용화의 중심에 인터넷 마케팅이라는 개념이 생겨났고, 예전에 비해 관심이 고조되면서 점차 정보와 기술기반보다는 고객 니즈를 충족시켜 줄 수 있는 다양한 마케팅이 필요함을 인식하기 시작했다. 이에 로그파일의 중요성은 굉장히 중요한데, 사용자 중심적인 마케팅 전략을 수립할 수 있는지에 대한 길잡이 역할을 해주기 때문이다. 로그파일을 분석하여 방문자수, 방문자 유형, 사용 시간대 별 뿐만 아니라 주요 고객층, 고객 구매패턴, 주 구매시간, 구매 탐색경로 등의 데이터를 추출 할 수 있으며 이러한 데이터를 기반으로 인터페이스 설계나 상품 레이아웃 등의 설계, 고객 서비스의 강화 등의 다양한 고객 마케팅을 펼칠 수 있다.

웹 사이트의 로그파일을 토대로 사이트 방문자에 대한 다양한 데이터 및 방문자의 이동경로를 다각적으로 분석하여 웹 사이트의 문제점과 앞으로 인터넷 비즈니스의 전략적 방향제시를 위한 사이트 전략수립에 대한을 제시하고자 한다.[13]

제 1장은 서론으로 간략하게 정리하였으며, 제2장 관련연구, 제3장 분석시나리오를 바탕으로 제4장 결과를 정리 하였다.

본 연구에서는 방문자의 IP를 중심으로 성별, 성향, 회원 수, 비회원수, 재접속수를 분석하여 방문자의 클릭반응을 통한 이동경로 및 웹 사이트 활용도 측정을 얻어내는데 본 연구의 목적이다.

Ⅱ. 관련연구

1. 인터넷 쇼핑몰

1) 정의

Hoffman과 Novak(1996)은 인터넷 쇼핑몰을 “다양한 영역의 제품들을 포함한 온라인 상점들의 집합”이라고 정의했다. 이는 상품광고 및 전시가 인터넷 쇼핑몰을 통해 행해지고, 서버에 여러 가지 상품에 관한 가격과 특성 등에 관한 자료들을 보유하고 웹페이지를 이용하여 멀티미디어 정보와 함께 상품에 대한 정보를 제공하는 것을 의미한다.

인터파크(1997)는 인터넷 쇼핑몰을 “통신 네트워크에 연결되어 있는 컴퓨터(Server)상에 상품정보를 올려놓고 이 컴퓨터에 접속하는 이용자가 상품을 선택한 후, 온라인 상 에서 결제하면 이용자가 원하는 장소로 상품을 배송해주는 새로운 상품판매형태”로 정의하였다.

소비자와 판매자가 직접 연결되어 전자적으로 상품에 대한 정보를 교환하며 상품을 주문한다. 또한, 주문한 상품에 대한 대금을 지불한다. 상품에 따라서 실시간으로 직접 인도 받을 수 있으며, 시간이 단축되고 지리적, 공간적, 제약이 제거되는 등 기존의 상점이용과 비교할 때 많은 상이한 점을 지니고 있다.[21] 이러한 인터넷 쇼핑은 기업 입장에서는 인터넷 쇼핑몰을 통하여 중간 물류 및 유통단계를 축소할 수 있고, 인적, 물리적 공간자원을 감소하여 비용절감의 효과를 얻을 수 있다. 물건을 사기 위해 차를 타고 백화점이나 상점을 갈 필요 없이 안방이나 사무실에 앉아서 버튼 하나만으로 상품이 집까지 배달되며, 기업은 매장관리비나 여러 유통단계를 거치지 않으며, 소비자와 직거래를 통하여 대금을 현금으로 받는다. 이러한 인터넷 쇼핑몰은 전자매체를 이용한다는 의미에서 기업과 소비자간 전자상거래의 일종이라고 할 수 있다. 매장건축, 상품진열, 판매해위, 대금결제 등이 통신 네트워크에 접속된 컴퓨터 안에서 구현되는 것이 실물 공간에서 상점을 구축, 운영하는 기존 유통시스템과의 차이점이라 할 수 있다.

2) 유형

최근 전자상거래의 형성과정을 통해 다양한 전자상거래 사업모델이 제시되고 있으며, 전통적인 비즈니스 모델이 새로운 형태로 전환되고 있다. 한국전산원에 의하면 인터넷 쇼핑몰에 대한 유형은 다양한 요인에 의해 분류될 수 있는데, 판매하는 상품관점에서 4가지로 분류할 수 있다.

첫째, 판매하는 상품의 성격에 따라 제품을 거래하는 사업과 서비스를 제공하는 사업을 구분할 수 있다. 제품을 거래하는 사업의 경우 다시 물리적 제품을 취급하는 경우와 디지털 제품을 취급하는 경우로 구분되어, 모두 세 가지 유형의 사업으로 분류할 수 있다. 이러한 세 가지 사업유형은 거래 절차의 전자적 수행 측면에서 각각 차별되는 특성을 가지고 있다. 물리적인 제품을 거래하는 사업의 경우 전자적으로 주문되어 물리적으로 배달되는데, 일반적인 인터넷 쇼핑몰이 이러한 유형에 해당된다. 증권, 은행, 전문컨설팅과 같은 서비스는 전자적 주문과 전자적인 배달이 가능하다. 전자적으로 배달되는 상품을 판매하는 유형은 물리적으로 배달되는 상품과 비교하여 많은 경제적 이점을 가지고 있다.

둘째, 취급하는 상품의 종류에 따라 일반 점(종합쇼핑몰)과 전문점(전문쇼핑몰)으로 구분 할 수 있다. 일반 점은 백화점과 같이 다양한 종류의 상품을 취급하는 형태이며, 전문점은 이와 달리 특화된 분야의 제품을 취급하는 형태이다.

대부분의 종합쇼핑몰은 일반 점에 해당하며, 대부분 전자상거래시스템을 독자적으로 운영하고 있다.

셋째, 전자상거래 기업이 가치사슬에서 점하는 위치에 따라 공급자와 중간매개자로 구분할 수 있다. 소비자가에게 직접 상품을 제공하는 공급자는 다시 제조업체가 기존의 유통채널이외에 인터넷을 직접적인 판매채널로 활용하는 형태와 제조업체로부터 상품을 공급받아 판매하는 유통업체 형태로 구분된다. 중간매개자는 고객인 회원에게 상업적인 거래를 위한 전자화된 시장을 제공해 주고 참여자간의 직접적인 거래를 지원해주는 전자경매, 가상공동체 등이 이러한 형태에 해당된다.

넷째, 전자상거래 기업은 인터넷 상에서만 판매하는 경우와 실물세계의 물리적인 매장과 인터넷을 동시에 판매채널로 활용하는 경우로 구분할 수 있다. 전자는 새로운 비즈니스 모델을 시도하는 것이거나 현재 존재하는 시장에 대한 신규 진

입에 해당하며, 후자는 전통적인 기업이 인터넷을 새로운 판매채널로 확장한 경우이다. 최근에 유통, 제조, 서비스산업에서 전통적인 기업이 인터넷으로 전환되는 경향이 두드러지고 있다. 인터넷을 판매채널로 활용함으로써 유통업체의 경우 물리적인 매장이 없어도 판매가 가능하기 때문에 고정비용을 절감할 수 있고, 제조업체의 경우 유통채널을 경유하지 않음으로써 유통채널에 의해 부가되던 원가요인을 절감 할 수 있으며, 은행, 증권업체와 같은 금융서비스 산업도 고정비용과 거리비용을 절감 할 수 있다.[20]

Table. 1. 인터넷 쇼핑몰의 유형

판매하는 상품 관점	상품의 성격	제품을 거래하는 사업 서비스를 제공하는 사업
	취급상품 종류	일반점(종합쇼핑몰) 전문점(전문쇼핑몰)
전자상거래 업체의 성격 관점	가치사슬의 위치	공급자 중간매개자
	판매채널의 복수성	인터넷에서만 판매하는 경우 OFF-Line과 On-Line 모두 판매채널을 가지고 있는 경우

3) 기존 마켓플레이스의 차이점

인터넷 쇼핑몰의 고객 측면에서의 장점은 24시간 시간적 제한을 받지 않고 상품이나 서비스를 구매할 수 있고, 또한 기존의 일반 쇼핑몰에서의 쇼핑보다 더 많은 제품을 접할 수 있어 제품을 믿고 구매할 수 있다.[15] 한편 인터넷 쇼핑몰은 소비자의 신용카드와 같은 개인정보가 유출될 수 있다는 단점을 가지고 있다. Table. 2는 위에서 언급한 인터넷 쇼핑몰의 장・단점을 정리한 테이블 이다.

Table. 2. 소비자 측면에서의 인터넷 쇼핑몰의 장・단점 비교

구분	소비자
장점	24시간 쇼핑할 수 있는 시간의 효율성 국내에서 구하기 힘든 상품 구매 가능성 신속한 고객지원을 받을 수 있음 일반 쇼핑몰보다 가격이 저렴함 다양한 상품정보로 의사결정의 효율성
단점	개인정보의 노출, 악용 위험 불량품 반품의 어려움(특히, 해외거래 시) 거래 사기 빈발

2. 웹 데이터 마이닝

1) 정의

웹 데이터 마이닝은 웹으로부터 얻어지는 정보를 찾아내어 분석하는 것이다. 웹으로부터 얻어질 수 있는 정보인 트래픽, 등록 정보, 거래 정보 등으로 실시간 활용할 수 있는 것들로 실시간으로 웹 데이터를 분석하여 진정한 의미의 개인화 서비스를 가능케 하며 CRM, eCRM, 일대일 마케팅, 개인화 등에 적용할 수 있다.

또한, 데이터 마이닝을 통해서 기업은 웹사이트상의 패턴을 의미 있는 정보로 종합해내고 인터넷상의 고객들과 예상치를 이해하고 연관시킬 수 있게 된다. 데이터와 웹이 제공하는 방대한 사업지식의 흐름에 근거한 웹 마이닝은 온라인 고객과의 관계를 생성하고 유지시키며 생산성 있는 온라인 상점의 최전선을 구축하는데 결정적 열쇠가 되는 것이다.

어느 웹사이트에 심심해서 방문을 했던지 그렇지 않으면 구체적인 목적을 갖고 방문을 했던지 간에 기업의 웹사이트에 한번이라도 들렀던 사람들이 사이트에서 행한 모든 일은 기록으로 남게 된다. 웹사이트 설정이 어떻게 되었는지에

따라서 다르긴 하지만 방문자가 어떤 경로를 통해 들어왔는지 알 수 있다.

예를 들어, 어떤 검색엔진을 통해 들어왔으며, 어떤 키워드로 검색했는지에 관한 정보들에 대한 로그파일이 남게 된다. 또한 쿠키(사용자 인증에 관한 사용자 브라우저와 서버간의 주고받은 기록)는 홈페이지 방문자의 이동경로 혹은 그 사용자가 이전에 한번 들렀던 적이 있는 사용자인지의 여부를 알려주게 된다. 그러나 그 보다도 더욱 중요한 것은 방문한 고객의 정보가 기록된다는 것이다.

웹 사이트에서 수집한 고객등록 정보들과 그 내용들에 대한 마이닝을 통해서 어떤 제품들이 교차판매에 유리한지, 어떤 정보들을 어떻게 링크시킬 것인가, 그리고 어떤 판촉광고가 홈페이지 방문자들에게 인기가 있었는지에 대한 정보를 그들이 가지고 있는 성별, 연령 등의 인구 통계학적인 특성과 일을 바탕으로 제시해준다. 또한 그들의 지리적 분포 또한 우편번호의 집계를 통해 산출해 낼 수 있어 웹 사이트를 활성화시키는데 필요한 주요 정책들을 결정하는데도 도움을 줄 수 있다.

소비자들의 월간 수입이 얼마나 되는지, 어떤 승용차를 몰고 있는지, 구입년도는 언제고, 어느 금융기관의 할부를 이용하는지 등과 같은 정보들을 수집함으로써 현재 보유하고 있는 고객들의 인구 통계학적인 데이터를 비교 분석해 볼 수 있다.

사실 기업의 웹 사이트는 모든 방문자들과 온라인 구매 등을 통해 저절로 구축되는 다이나믹한 소비자 데이터베이스라고 할 수 있다. 웹 사이트를 통해 일어나는 행위들에 의해 생성된 데이터들은 지속적으로 발생되어 무수히 많은 로그파일들이 데이터베이스 기록된다. 만약 데이터웨어하우스에 연결되어 있다면 그 데이터베이스는 풍부한 잠재고객을 마이닝하기 위한 충분한 요소를 갖추게 되며, 기업의 웹 사이트상의 각종 등록 양식들을 통해 목표 시장을 분할해 낼 수 있고, 등록 양식을 마이닝하여 데이터웨어하우스 안의 고객 프로파일들과 비교해 볼 수도 있다.

오늘날 대부분의 데이터 마이닝 툴은 분류나 예측, 고객 프로파일링을 위한 신경망 구조를 가지고 있다. 신경망 구조는 고객들의 온라인 행동을 예측하고 그들의 광고나 배너, 구매제안에 대한 선호도뿐만 아니라 온라인 구매를 유발시키기 위한 목적으로 사용될 수 있다. 신경망 구조를 구축하는 데이터 마이닝 툴은 방

문자들의 남기고 간 흔적들로부터, 긍정적이거나 부정적인 소비자의 행위에 대한 패턴을 고찰하고, 등록된 고객의 특성을 파악하는 데 유용하게 활용될 수 있다. 신경망 구조가 축적된 데이터를 바탕으로 성장해감에 따라, 어떤 방문자들이 어떤 상품들을 선호하고 싫어하는지의 상관관계를 점차적으로 배울갈 수 있는 것이다.

신경망 구조를 거치면서, 소비자들의 행동패턴이 정형화된 컴퓨터 코드의 형태로 기록된 하나의 모델로 형상화되어 가는 것이다. 이런 코드화된 결과물은 처음 사이트에 방문한 사람이 어떤 사람인지를 판단하여 그 사람이 그 웹 사이트로부터 어떤 상품을 구매하고 싶어 할지를 예측할 수 있는 근거자료가 된다.

새로운 방문자에 대한 예측 분류나 행동양식에 대한 예측에 기초하여, 우리는 비로소 그 사람에게 가장 적절한 광고나 마케팅 메시지를 전달할 수 있게 된다. 즉, 그 사람이 어떤 상품을 구매하고 싶어 하는지를 먼저 예측하여 그 사람에게 적절한 상품 판촉메일을 보낼 수 있게 되는 것이다.

정형화된 모델이 신경망 구조로부터 구축된 후에도 민감도 분석 보고서는 끊임없이 갱신되면서 지속적인 온라인 판매의 예측 결과를 생성할 수 있도록 해준다. 이러한 타입의 보고서는 어떤 인구통계학적 특성이, 혹은 소비자들과 인터넷 사이트가 어떤 상호작용을 하는 것이 온라인 판매에 결정적 역할을 하게 되는지를 보여주는 것이다.

웹 사이트 마이닝을 통해 기업은 인구통계학적 소비자 선호도를 발견하여 특정 광고나 배너를 포지셔닝 할 수 있도록 하는 기초자료를 추출할 수 있다. 새로운 데이터나 정보가 웹 사이트를 통해 수집되면 이 정보들은 지속적으로 데이터 웨어하우스로 통합되어 향후의 의사결정에 도움을 주는 분석결과를 제공하고, 데이터베이스 마케팅과 전략기획을 위한 자료로 활용되는 것이다. 또한 웹 사이트 데이터 마이닝을 통해서 온라인상에서 제공하는 서비스와 제품 간의 연관관계를 밝혀내어 적절한 제품이 적절한 서비스와 함께 판매가 되고 있는지의 여부를 밝혀내게 될 수도 있다.

데이터 마이닝을 통해서 기업은 웹 사이트상의 패턴을 의미 있는 정보로 종합해내고, 인터넷 상의 고객들과 예상치 들을 이해하고 연관시킬 수 있게 된다. 데이터와 웹이 제공하는 방대한 사업지식의 흐름에 근거한 웹 마이닝은 온라인 고

객과의 관계를 생성하고 유지시키며 생산성 있는 온라인 상점의 최전선을 구축하는데 있어 결정적 열쇠가 되는 것이다. 웹은 소비자와 판매자간의 밀접한 관계를 형성하는 새로운 매체라는 특성으로 인해 아주 독특한 채널로 인정받고 있다. 소비자를 대상으로 한 전자상거래를 통하여 판매자는 고객의 취향과 소비패턴, 그리고 그들의 물건을 구매할 때 질과 가격과 볼륨 중 어떤 것을 중요하게 생각하는지에 대한 정보를 수집할 수 있다.

다른 채널과는 다른 웹은 일대일 판매환경을 조성하며 고객과 소매상간의 밀접한 상호작용을 구축할 수 있게 한다. 인터넷을 통한 판매는 쌍방향 상호작용을 유발하여 상품을 생산하고 마케팅활동을 전개하고 판매행위를 하며 이는 다시 소비자와의 상호작용 속에서 마이닝을 진행하게 되는 일련의 사이클 과정이 되는 것이다. 전자상거래는 데이터 웨어하우스와 결합되었을 때 비로소 완전한 소매 사이클의 한 부분을 제공할 수 있으며, 소비 프로세스와 성향, 고객의 선호도, 새로운 잠재 고객을 볼 수 있는 시각을 제공할 수 있다.

데이터 마이닝은 인지와 습득에 관한 것이라는 점을 잊어서는 안 된다. 데이터 마이닝은 수시로 변하는 그 어떤 비즈니스 환경에서도 패턴들을 찾아내는데 사용되는 인위적인 지식기술이며, 보다 경쟁력 있는 지식전략을 위해 인위적으로 이를 사용하는 것이다. 또한 데이터 마이닝은 점차적으로 양산되어지는 강력한 패턴을 인식하는 기술이자 도구로써, 디지털화와 네트워크화가 확장되고 가속화되어가는 시장변화의 흐름에 재빠르게 적용할 수 있도록 해준다. 수백만의 방문자들이 세계 도처의 웹사이트들과 상호 행위를 주고받는 과정에서 방대한 양의 데이터가 매일매일 쌓여가고 있다. 그러나 웹 사이트가 현재와 미래의 고객들이 기업과 가장 가깝게 관계를 맺을 수 있는 도구가 된다는 놀라운 사실을 알면서도, 매일 매일의 이 중요한 정보들을 마이닝하는 기업은 많지 않다. 웹 사이트를 통해 수집되는 정보를 기업의 데이터웨어하우스와 통합하는 것은 가까운 미래에 그 회사에 매우 커다란 기회를 제공하게 된다. 웹상에서 일어나는 일들이 수년 이내에 수십억 건을 돌파할 것으로 예상되고 있다. 곧 전자상거래와 비즈니스 지식은 온라인 고객을 끌어들이고 확보하기 위한 웹 콘텐츠의 중심축이 될 것이며, 데이터 마이닝은 행동양식의 패턴과 방문자들의 프로파일을 통해 콘텐츠의 근거자료들을 제공할 수 있게 될 것이다.[4]

2) 분 류

웹 데이터 마이닝은 웹 기반의 구조화되지 않은 데이터로부터 비슷한 형태를 찾기 위한 것으로 관계형 데이터베이스를 활용하기 위한 데이터 마이닝 기법인 Web content Mining, 웹 사이트나 웹 페이지에 대한 요약된 구조를 생성시키는 것을 목적으로 하는 Web Structure Mining, 사용자들이 웹브라우징 했던 기록을 남기는 웹 서버 로그로부터 접속 유형을 발견하는 것을 목적으로 하는 Web Usage Mining으로 분류 할 수 있다.

⑴ Web content Mining

Web content Mining은 웹 사이트의 콘텐츠, 자료, 정보 등의 관계를 분석하여 사용자의 요구에 가장 잘 부합하는 내용을 보여줄 수 있도록 자동으로 찾아주는 프로세스를 말한다.

Web content Mining은 웹 사이트에서 사용자에게 제공하고자 하는 모든 콘텐츠(일반 텍스트 문서, 동영상, 그림, 각종 데이터), 사용자의 만족도, 기호도, 관심도를 측정하는 데이터로 서치엔진(자연어 처리 시스템), 데이터베이스의 축적된 데이터 중에서 원하는 자료를 쉽게 찾을 수 있게 해주는 구조화된 질의 언어에 활용 할 수 있다.

⑵ Web Structure Mining

Web Structure Mining은 웹 사이트와 웹 페이지의 하이퍼링크를 데이터 마이닝 과정을 통해 정보를 구조화, 표준화 시키는 프로세스를 의미한다. 웹 사이트 문서, 하이퍼링크 등의 원천 자료를 가지고 웹 사이트 사이의 관련된 페이지, 웹 페이지 사이의 관련된 링크 등에 활용할 수 있다.

⑶ Web Usage Mining

Web Usage Mining은 웹 서버로부터 사용자의 접근 패턴을 발견하는 자동화된 마이닝이다. Web Usage Mining은 Server Level, Client Level, Proxy Level 등에서 발생시키는 접속 로그들이며 이 밖에 Referrer log, CGI 스크립트 등에 의해 얻어낸 고객 등록 정보, 설문 데이터 등으로 개인화 맞춤 서비스, System Improvement, Site Modification 등에 활용 할 수 있다.

3) 단계

⑴ 자료수집(Resource finding)

웹 사이트를 구성하고 있는 웹 페이지 및 웹과 연동되는 데이터베이스, 웹 사이트의 구조 정보, 웹 사이트 사용자가 웹 서버 로그 파일, 사용자의 프로필 등의 정보를 직접 또는, 에이전트를 이용하여 수집한다.

⑵ 전처리과정(Information Selection & Pre-processing)

준비된 데이터들에 마이닝 작업을 수행할 수 있도록 필터링 한다. 정제(cleansing), 변형(transforming) 등의 작업을 걸쳐 데이터를 재구성한다.

⑶ 일반화(Generaliazation)

각 목적에 맞는 다양한 데이터 마이닝 알고리즘을 적용하여 하나의 혹은 여러 개의 웹 사이트에 대해서 일반적인 패턴을 자동으로 발견한다.

⑷ 분석(Analysis)

데이터 마이닝을 통해 발견된 패턴과 결과를 적절한 검증 기법을 이용하여 해석, 검증한다.

4) 방법론

일반적으로 데이터 마이닝의 절차는 “데이터 획득→ 샘플의 추출→ 데이터 정제 및 변환→ 변수선정→ 모형구축→ 모형평가”의 단계를 거친다. 그리고 이러한 일반적인 절차를 기반으로 데이터 마이닝 제공업체별로 다양한 방법론을 제시하는데 간략하게 각 업체의 SEMMA 방법론을 소개하겠다.

SAS Software에서 제공하는 데이터 마이닝 분석 방법론으로 5단계인 Sampling, Exploration, Modification, Modeling, Assessment로 구성되어 있다.

⑴ Sampling 단계

분석의 목적에 따라 적절하게 추출된 표본의 활용은 비용과 시간의 절약, 보다 효율적인 모형화 작업을 위해서 매우 중요하다. E-Miner는 전체를 잘 대표 할 수 있는 표본을 추출하기 위해서 단순임의 추출로부터 복잡한 층화추출에 이르

기까지 다양한 표본추출 방법을 제공한다.

⑵ Exploration 단계

여러 측면에서의 데이터 탐색을 통해서 기본적인 정보를 검색하고 유용한 정보를 추출하는 기법들을 제공한다.

⑶ Modification 단계

탐색 단계에서 얻어진 정보를 기반으로 모형화 단계에서 모형의 성능을 향상시키기 위해, 데이터가 가지고 있는 정보를 효율적으로 사용할 수 있도록 변수변환, 수량화, 그룹화 같은 방법을 통해서 데이터를 변형하고 조정한다.

⑷ Modeling 단계

모형화를 통해 얻어진 결과의 신뢰성, 타당성, 유용성 등을 평가 할 수 있다. 평가단계에서는 리포트 표, ROC곡선, 이익도표, ROI곡선 등 다양한 평가도구가 제공된다.

3. 로그분석

1) 개념

로그 파일이란 웹 서버를 통해 이루어지는 모든 작업들에 대한 기록이라고 표현할 수 있다. 우리가 웹 서버에 접속을 하게 되면 그 이후의 모든 작업들은 웹 서버에 접속하고, 미리 정해 놓은 위치에 데이터로 남게 된다. 일반적으로 특정 웹 페이지를 보기 위한 사용자의 요구로 웹 서버는 해당 웹 페이지와 관련된 여러 파일 등에 접근하게 된다. 따라서 사용자가 요청하는 특정 웹 페이지 뿐만 아니라 웹 페이지와 관련된 이미지 파일, 이미지 데이터, 인쿨루드 파일 등에 대한 정보가 로그파일에 저장된다.[6]

즉, 사이트의 방문객이 남긴 자료를 근거로 웹의 운영 및 방문 행태에 대한 정보를 분석하는 것을 말한다. 방문객이 웹 사이트에 방문하게 되면 웹 서버에는 액세스 로그, 에러 로그, 리퍼럴 로그, 에이전트 로그 등의 자료가 파일 형태로

기록된다. 액세스 로그는 누가 어떤 것을 읽었는지를, 에러 로그는 오류가 있었는지를, 리퍼럴 로그는 경유지 사이트와 검색 엔진 키워드 등의 단서를, 에이전트 로그는 웹 브라우저의 이름, 버전, 운영 체계, 화면 해상도 등의 정보를 제공한다. 이러한 기본적 분석 외에도 실시간 분석을 위해 분석 태그를 웹 사이트에 삽입하여 분석하는 방법도 있다. 웹 로그 분석에 의해 얻은 방문자 수, 방문 유형, 각 웹 페이지별 방문 횟수, 시간·요일·월·계절별 접속 통계 등의 자료는 웹의 운영 및 마케팅 자료로 유용하게 이용된다.[12] 웹 로그분석의 필요성은 바로 측정에 근거한 E-Biz의 진행에 있다. E-Biz의 경우 IT적 요소가 결합되어 있으므로 측정이 보다 수월하며, 로그분석 등 측정에 근거한 웹사이트의 과학적 운영과 그 측정결과에 근거한 마케팅 활동 프로세스는 E-Biz 전체결과에 중요한 영향을 미친다.

인터넷에 기반을 둔 E-Biz의 성과측정 및 원인분석을 위한 다양한 요소가 존재하며, 정리하면 다음 6단계와 같다.

1단계, 트래픽분석 : 방문자수, 페이지뷰 등 트래픽분석

2단계, 매출/수익 분석

3단계, 회원(인구통계학적)

4단계, 방문자(익명) 성향분석 : 컨텐츠/상품 관심도, 행동분석

5단계, 경영정보시스템(MIS) 및 고객관계관리(CRM) 등에서 측정 및 분석결과 제공

6단계, 웹 로그분석 및 방문자관계관리 등에서 측정 및 분석결과 제공

측정에 근거한 E-Biz진행을 위해 마케팅/운영 활동의 1차 결과에 대한 분석이 선행되어야 하며, 이와 연계하여 2차 결과에 대한 평가가 이루어져야 한다.

측정분석은 결과중심의 측정/분석(마케팅/운영의 2차 결과)과 원인중심의 측정/분석(마케팅/운영의 1차 결과)으로 나눌 수 있는 데 원인중심의 1차 결과에 대한 분석을 위해 웹 로그분석/방문자 분석을 통한 접근이 가장 쉬운 방법이다. 기초적인 1차 결과를 소홀히 한 2차 결과만의 판단은 전체가 아닌 일부만으로 판단하는 위험성을 내포하고 있다. 특히 E-Biz는 IT를 기반으로 하여 운영되므로 마케팅 및 운영활동의 1차 결과에 대하여 대부분의 기록을 남길 수 있으며, 분석이 가능한 일종의 블랙박스를 갖고 있습니다. 이를 분석하지 않는 것은 경영에

필요한 중요 자료를 버리는 것과 같다. 웹 로그분석 및 방문자 분석은 이와 같은 블랙박스 부분을 분석함으로서 측정에 근거한 E-Biz를 수행할 수 있도록 해준다. 웹 로그분석 및 방문자분석을 통해 운영의 상태와 결과에 대한 원인을 파악할 수 있으며, 이를 통해 과학적인 E-Biz의 진행이 가능하게 된다.

웹 로그분석을 통해 얻을 수 있는 정보는 다음과 같으며, 스탠다드 서비스기준과 커버스 서비스 기준으로 분류하였다.

- 방문자는 평균 몇 페이지를 둘러보고 떠나는가? 그리고 그 추세는?
- 이벤트/캠페인 후 어느 정도의 방문자가 증가하였는가?
- 재방문을 하는 방문자의 비율은 어느 정도 되는가?
- 어떤 페이지/컨텐츠를 방문자들이 가장 선호하는가?
- 방문자는 어떤 페이지에서 웹사이트를 떠나는가?
- 어떤 검색엔진 및 검색단어로부터 웹사이트를 방문하는가?
- 어떤 검색엔진에 어떤 키워드광고를 해야겠는가 ?
- 어떤 웹사이트가 우리 웹사이트를 링크하고 있는가?
- 방문자는 어떤 시/도, 회사/조직이며, 어떤 ISP를 통해서 접근하는가?
- 방문자의 웹 브라우저, 언어 등 사용 환경은 어떻게 되는가?

◎ 스탠다드 서비스 기준

- 제휴사와 교환한 배너는 어느 정도 노출되고 클릭되는가? 그리고 그 추세는? (이하 스탠다드 서비스)
- 홍보메일, 뉴스레터발송에 대하여 어느 정도 메일이 읽혀지고, 클릭되어 접속하는가?
- 언제 배너의 이미지 등을 바꿔서 관심도를 높여야 하는가?
- 특정 페이지 내 어떤 링크를 선호하여 클릭하는가?(micro action)
- 방문자는 웹 사이트 내에서 어떠한 경로로 서핑을 주로 하는가?
- 방문자는 페이지별 몇 분정도의 머무르고 떠나는가?
- 어떤 컨텐츠/상품 카테고리를 가장 많이 둘러보는가?
- 어떤 상품이 가장 많이 보여 지는가? 그리고 그 상품의 관심도 추세는?
- 어떤 공급사의 상품이 가장 많이 보여 지는가? 그리고 그 추세는?
- 방문자중 활성화된 구매 가능성 있는 로그인한 고객의 비율은 어느 정도인가?

- 방문자중 어느 정도의 수(비율)가 주문을 완료 하는가?
- 방문자중 어느 정도의 수(비율)가 회원등록/이벤트등록을 하는가?
- 장바구니에 상품이 담길 비율과, 최종 구매될 비율은 어떻게 되는가?
- 어떤 구매단계 중 방문자가 구매를 포기하는가?

◎ 커머스 서비스 기준

- 각종 마케팅활동(이메일, 광고, 제휴 페이지 등)의 통합된 캠페인 분석과 관리가 가능 한가?
- 어떠한 캠페인이 실제 매출, 회원확보, 주문발생에 기여 하였는가?
- 캠페인별 집행비용대비 실제 효과의 비율은 어떻게 되는가?
- 캠페인별 통합된 매출, 회원획득, 주문의 추세는 어떻게 되는가?
- 매출에 기여한 상품은 어떤 것 이며, 방문 당 평균 구매액은 어떻게 되는가?
- 참조도메인(Referrer Domain)별 방문자중 실제 매출, 주문, 회원획득에 기여하는 참조도메인은 어디인가 ?
- 다양한 제휴사중 실제 매출, 회원획득에 기여하는 제휴사는 어디인가?

2) 종류

로그 파일은 사용자가 처음 사이트를 방문하면 웹 서버에 자동으로 자신의 로그 파일이 웹 서버에 저장되어 로그파일이 생성되게 된다.

로그 파일은 웹 서버가 지정하는 곳에 위치하며, 웹 서버 관리자는 웹 서버를 설치할 때 로그파일의 위치와 기록방법 등을 지정할 수 있으며, 웹 서버에 따라서는 한 개가 아닌 여러 개의 로그파일을 만들 수도 있는데, 이는 트랜스포(Transfer), 엑세스(Assess), 로그파일, 에러(Error) 로그파일, 리퍼럴(Referer) 로그파일 및 에이젼트(Agent) 로그파일 등으로 크게 4가지로 분류 할 수 있다.

⑴ 엑세스 로그(Access_Log)

엑세스 로그파일은 트랜스퍼 로그파일이라고도 한다. 일반적인 사항을 모두 기록하며, 접속자가 들어와서 웹 서버에서 한 행동을 그대로 보여줄 수 있다. 그러므로 이 정보는 차후에도 많은 이용가치가 있기 때문에 아주 중요하다.

웹투비(WebtoB)는 NODE, SVRGROUP, VHOST 등에 모두 설정이 가능하며

각 절에 모두 설정했을 경우 접속자의 접속서비스를 요청에 따라 그 우선순위가 SVRGROUP, VHOST, NODE 순서로 존재하며 가장 우선순위가 높은 절에 해당하는 로그파일에 그 내용이 기록된다.

⑵ 리퍼럴 로그(Referer_Log)

리퍼럴 로그는 화살표로 표시되며, 방문자가 사이트를 방문하기 위하여 어떠한 검색엔진을 활용하였으며, 사이트를 들어오기 위해서는 어떠한 키워드를 검색하여 방문하였으며, 방문자가 사이트를 방문하기 위하여 거친 URL 경로는 어떠한 것이 있는지를 알 수 있기 때문에 검색된 키워드를 통해 고객들이 원하는 콘텐츠를 구성할 수 있고, 검색엔진과 링크페이지를 통해 인터넷 광고 매체 선정 및 서치 엔진 키워드 구성 등의 프로모션 전략 방안을 설정하여 타깃 화 된 웹 프로모션 전략 방안을 설정하여 웹 프로모션을 전개할 수 있다.

⑶ 에이전트 로그(Agent_Log)

에이전트 로그는 사이트를 접속하는 방문자의 웹 브라우저 타입 및 버전, OS의 종류, 화면해상도 애플리케이션 프로그램 종류 등에 관한 정보를 제공해 최적화된 웹 사이트를 구성할 수 있는 단서를 제공해주고 있다.

⑷ 에러 로그(Error_Log)

에러 로그는 웹 서버에서 발생하는 모든 에러와 접속 실패에 대하여 에러가 발생한 시간과 에러의 내용을 기록한다. 이는 시스템에서 발생할 수 있는 에러에 대한 기록이기 때문에 웹 서버에 문제가 생긴 경우 문제 해결을 수행하는 경우는 이를 저장하여 시스템에 문제가 생길 경우 이를 통하여 문제 해결을 쉽게 할 수 있다. 대부분 에러 로그는 상태코드에 404나 505등이 기록된다.[25]

3) 방식

⑴ 태그(TAG)방식

태그 방식은 웹 서버의 부담을 줄이고 사용자의 트랜잭션 정보를 실시간으로 수집하기 위하여 웹문서에 사용자들이 주고받는 정보를 인식하는 태그 코드를 삽입하여 사용자의 정보를 수집하는 방법이다.

⑵ TCP/IP 패키스니핑(Packet Sniffing)방식

패키스니핑 방식은 해킹공격 시 네트워크에서 주고받는 패킷데이터에 담긴 사용자의 로그인 정보를 빼내는 방법이나 이러한 방법을 응용하여 사용자 분석을 위한 패킷데이터에 담긴 방문자의 트랜잭션을 수집하여 분석하는 방법으로 실시간 정보수집 및 분석이 가능하다. 그러나 부가적인 시스템이 필요하여 웹사이트가 대규모가 될수록 높은 사양의 설비가 필요한 한계가 있다.

⑶ 서버 애드인(Server-Add In)방식

서버 애드인 방식은 서버에 로그데이터를 실시간으로 수집하고 분석할 수 있도록 서버에 내장되어 정보를 수집하는 방법으로 실시간 의사결정과 분 단위의 보고를 수행할 수 있는 반면에 사용자의 트래픽에 따른 웹 서버와 퍼포먼스에 영향을 미쳐 속도가 느려지거나 서버가 다운될 수 있기 때문에 ISP 또는 호스팅 업체에서는 채택하기가 곤란하다.

4) 측정 단위

⑴ 접속(Hits)

히트는 방문자가 웹 사이트를 접속했을 때 연결된 파일의 숫자를 말하는 것으로 한 페이지를 전송할 때 그 안에 포함된 그래픽, HTML 등의 모든 파일을 히트로 계산하고 있다.

⑵ 페이지뷰(Page View)

페이지뷰는 하나의 HTML 문서를 보는 것을 말한다. 현재 대부분의 인터넷 광고측정 시 페이지 개념을 많이 도입하고 있는데, 배너 광고가 포함되어 있는 웹 페이지가 한번 전송되면 일단 방문자가 페이지에 접속하여 광고에 노출된 것으로 간주한 것으로 여겨 효과적으로 측정할 수 있는 장점이 있기 때문이다. 광고 측정단위에서 이러한 페이지뷰를 임프레션 이라는 단위로 기록하고 있다. 페이지뷰는 현재 웹 사이트를 평가할 수 있는 단위 기준으로 가장 많이 사용하고 있는 단위이다.

⑶ 방문시간(Duration Time)

체류시간은 한 방문자가 특정 웹 페이지에 얼마나 오래 머물렀는가를 시간을 기준으로 기록하는 것을 말한다. 체류시간이 길다는 것은 그 페이지에 관한 관심이 많기 때문에 콘텐츠 분석과 관리 등의 효과적인 분석을 할 수 있다.

⑷ 방문(Session)

세션은 한 방문자가 특정 웹 사이트에 접속해서 연속적으로 페이지를 본 후 다른 사이트로 이동하는 과정을 하나의 방문으로 기록하는 것을 말한다. 세션의 측정은 보통 하나의 IP 어드레스를 통해 접속한 경우, 서버에 로그가 기록되며 기록된 IP 어드레스를 통해 파악하고 있다.

⑸ 방문자(Visitor)

방문자는 특정 웹 사이트를 한번 이상 접속한 사용자들의 수를 파악하는 방법으로서 방문자의 증가추이 및 충성고객 등을 파악하는 중요한 요소이다. 이러한 방문자 측정은 쿠키나 사용자 인증을 통해 방문자의 방문 경로 및 웹 사이트 방문 형태를 분석하여 웹 사이트의 콘텐츠 관리 및 전략 수립에 이용할 수 있다.

5) 웹 로그 분석의 방법론

웹 로그 파일을 분석하는 방식은 웹 서버에 쌓여진 웹 로그파일을 FTP 등을 이용해 다운로드 받은 후에 분석기를 실행시켜 분석 리포트를 받아보는 구조이다.

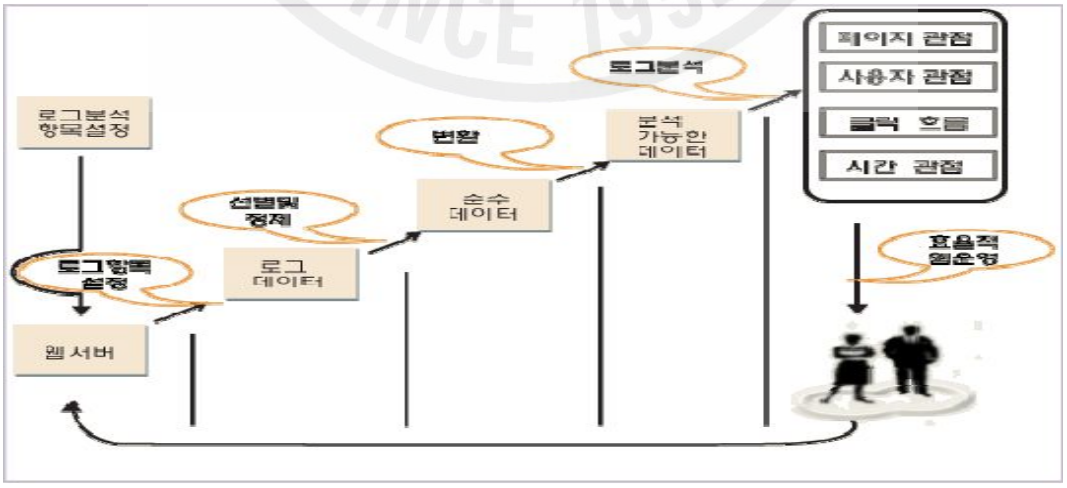


Fig. 1. 웹 로그 분석의 단계

⑴ 로그 분석 항목 선정

웹 로그 분석에 있어 가장 먼저 해야 할 작업은 분석할 대상을 명확하게 하는 것이다. 예를 들어, 회원 관리를 해야 하는 경우 사용자별 분석은 필수적이다.

로그 항목 설정

앞 단계에서 분석할 항목들을 정했다면 분석에 필요한 항목들만 로그에 남겨지게 해야 한다.

⑵ 선별, 정제 및 변환

필요한 데이터가 로그 파일에 저장된 후에는 분석 가능한 형태로 데이터를 변환한다.

데이터 분석에 있어서 가장 중요한 요소 중의 하나는 분석에 사용 될 데이터의 질이며, 정제 작업은 바로 데이터의 질을 높여 주는 역할을 한다. 로그 분석에 가장 시간이 많이 걸리지만 중요한 과정이 바로 이 과정이다.

⑶ 로그 분석

이 과정은 실제 로그 분석에 필요한 알고리즘이 적용되는 단계이다. 연관 규칙 탐사, 연속 탐사, 의사결정나무, 신경망 모형 등 데이터 마이닝에 쓰이는 알고리즘을 적용할 수 있다.

⑷ 결과 반영

마지막으로 분석결과를 웹 사이트 운영에 적용시킨다.

사용자들의 성향을 분석하고 사이트 항목 중 어떤 것에 주로 관심을 나타내는지 아닌지를 알아내서 사이트 개편 작업 등에 반영시킬 수 있다.

6) 웹 로그분석의 한계

로그분석 데이터는 서버에 기록된 사이트 방문자 데이터를 로그분석 들을 활용하여 각 조건(월별/일별/시간별)에 의하여 정리한 것으로 방문자 개개인에 관한 정확한 데이터가 아닌 단순한 방문자의 방문 기록 정보이기 때문에 사이트 현황에 관한 기본적인 분석만 할 수 있다. 로그분석의 방문데이터 또한 다양한 사용자 환경이나 기술적 환경에 의해서 정확한 결과를 얻어낼 수 없는 한계가 있다.

첫 번째, 로그분석의 사용자에 대한 구분이 접속 IP 주소 단위로 이루어지기 때문에 변동 IP 주소를 사용하는 경우 서로 다른 방문자를 동일한 세션으로 인식할 우려가 있으며 사람들이 많이 이용하는 공공장소인 PC방이나 학교에서 사용하는 경우 동일한 컴퓨터를 이용하여 사용자가 다중으로 인식하기 때문에 정확한 측정이 어려운 점이 있다.

두 번째, 웹 페이지에 접속할 때 프록시 서버나 웹 브라우저의 캐쉬를 이용하여 접속하는 경우 방문자의 프록시 접속이나 캐쉬 접속에 대한 트래픽을 측정하기 어려운 한계가 있다.

세 번째, 사용자가 어려 개의 브라우저를 동시에 띄워 놓은 상태에서 작업을 하는 경우 방문자의 웹 사이트 체류시간을 측정하기 어렵다.

웹 서버 로그 분석 방식도 역시 로그파일을 분석하는 가장 보편적인 방식으로 서버에 쌓인 로그 정보를 로그분석 프로그램을 활용하여 분석하는 방법으로 객관적인 증빙 자료로서의 가치가 있는 반면에 실시간 분석이 어렵다는 한계가 있다.[7]

Ⅲ. 분석 시나리오

1. 로그 데이터 구성

1) 환경

본 연구에서는 통계 프로그램인 SPSS WIN 10.0 웹 로그분석기를 사용하였으며, 그 구성요소인 프로그램 언어로 ASP를 사용하였다.

사용된 데이터베이스는 안정성 면에서 우수한 성능을 보여준 ACCESS를 사용하였다. 또한, 원시 로그파일인 Clementine를 이용하여 로그분석을 해봤다.

이를 도표화 하여 나타내면 아래 Table. 3 과 같다.

Table. 3. 로그분석 환경

분석 사이트	http://on-pedder.com/
분석 기간	2008.05.20 - 2008.05.28
분석 환경	SPSS WIN 10.0
프로그램 언어	ASP
사용된 데이터베이스	ACCESS

2) 구성

Fig. 4.는 소비자가 접속하였을 때 내부에 삽입된 로그 파일이다.

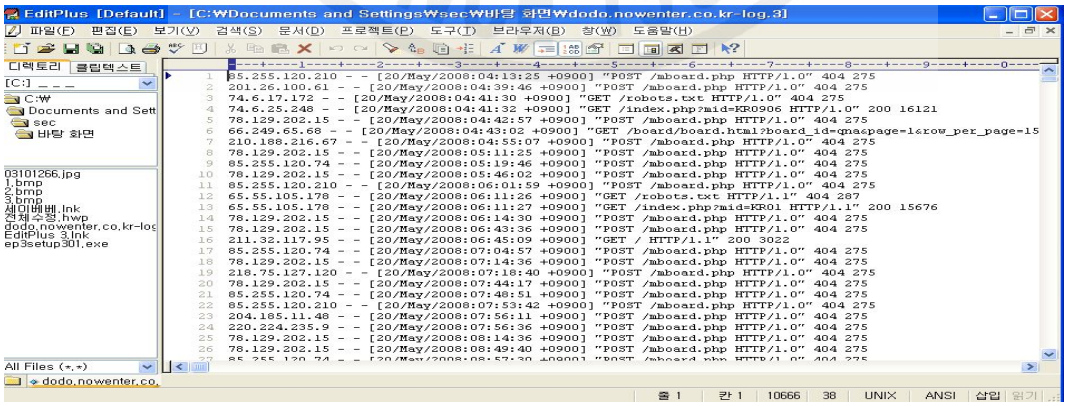


Fig. 2. 로그파일

Table. 5는 Fig. 2.의 로그 파일을 통해 들어온 로그 데이터의 레이아웃이다.

Table. 4. 구성

구분	설명
85.255.120.210	IP 주소
[23/May/2008:04:13:25]	날짜, 시간
POST	데이터 전송방식
mboard.php	경로
HTTP 1.0	프로토콜 버전

2. 시나리오 환경설정

1) 구성

로그파일은 웹 서버를 통해 이루어지는 대부분의 작업들을 기록하기 때문에 로그 분석을 위한 가장 중요한 데이터이다. 웹 서버에 저장된 순수 로그 파일로는 방문자의 IP, 접속시간, 접속 호스트, 전송상태 등과 같은 일반적인 정보만 제공할 뿐이어서 웹 사이트의 효율성에 영향을 미치는 방문자의 이동경로를 이해하기는 어렵다.

“쓰레기가 들어가면 쓰레기가 나온다(Garbage In Out)” 라는 오래된 컴퓨터 관련 격언처럼 로그분석이 쓰레기가 되지 않게 해주는 데이터 정제 작업이 이번 분석에서 가장 중요한 부분이다.

순수 로그파일로만으로는 방문자 중심의 데이터를 얻어낼 수 없기 때문에 원시 로그파일의 선별과 정제 과정을 거쳐 양질의 데이터를 추출해 내야한다. 데이터 정제의 수준이 로그분석의 수준과 그대로 연결되기 때문이다. 데이터 정제 과정에서 양질의 데이터를 얻기 위해서는 데이터를 단순히 걸러 내는 작업, 즉 확장자가 gif, jpg 등으로 끝나는 그림파일, 이외의 분석에 불필요한 파일들을 제거하는데, 이는 실제 방문자의 이동경로분석과는 무관하기 때문이다. 두 번째는 의미 있는 파일이라고 생각되는 데이터를 분석 가능한 형태로 변환하는 것이다.

본 연구에서는 온 페더 사를 대상으로 2008년 05월20일부터 05월28일까지 5일간의 로그 파일을 가지고 성별, 성향, 회원 수, 비회원수, 재접속 수, 페이지 빈도

수를 분석하였다.

온 페더 사의 웹사이트에 접속하는 방문자가 분석대상으로 선정하였다.

방문자 신분 확인 과정에서는 IP 분석법을 활용하거나 쿠키를 활용해 보완하지만, 쿠키 활용에 있어서 웹 서버에서의 별도 작업이 필요하다는 점과 개인보호정책 이슈와 관련된 문제로 인해 본 연구에서는 IP 만으로 방문자를 처리했다.

다음은 원시 로그파일을 Clementine을 이용하여 방문자 ID를 기준으로 로그데이터를 분석하였다.[13]

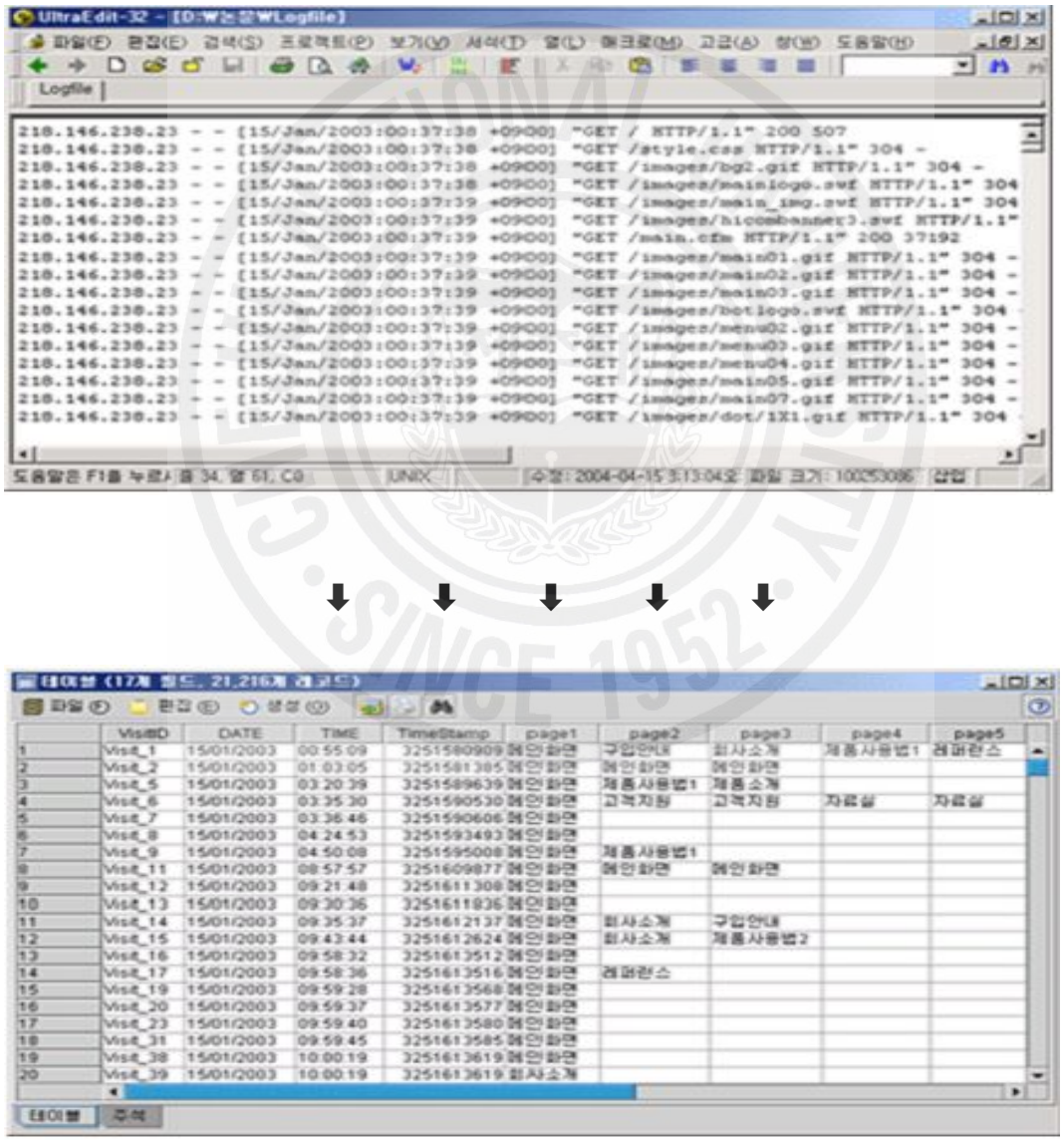
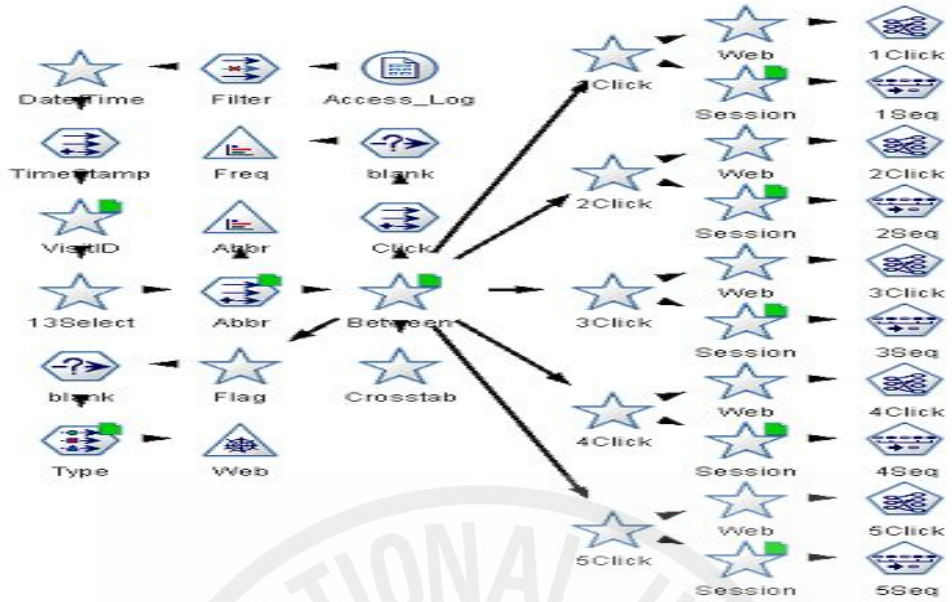


Fig. 3. Clementine을 이용한 로그분석



Fi.g 4. Clementine 스트림

노 드	설 명
Access_Log	본 연구의 분석을 위해 가변형식인 W사이트의 로그파일을 불러온다.
Filter	데이터 필드명을 변경하고, 스트림에서 사용하지 않는 필요 없는 필드들을 제거한다.
Date/Time	한 개의 필드에 설정된 날짜와 시간을 두 개의 필드 (날짜 필드와 시간 필드)로 변환한다.
TimeStamp	시간필드를 정수로 변환한다. 예 : time_in_secs(TIME) + (date_in_days(DATE) * 86400)
VisitID	각 IP 별 방문자를 정수로 변환한 후 visit_를 삽입한다. 예 : 203.234.83.2->visit_1, 157.62.145.2->visit_2, ...
13Select	분석에 필요 없는 파일들(jpg, gif...)을 제외하고, 분석에 필요한 파일들을 선택한다.
Abbr	필드 데이터를 간단한 문자로 보기 좋게 변환한다. 예 : pageA = 메인화면, pageB = 회사소개, ...
Between	방문자별로 페이지 이동을 순차적으로 각각의 필드에 변수로 저장한다.
Flag	연관성과 순차성 분석을 위해 방문자별 이동 페이지를 이분형으로 변환한다.
Session	방문자의 세션에 관한 흐름을 위해 세션별 아이디를 적합한 데이터로 변환한다.

Fig. 5. 로그분석을 위한 각 노드의 설명

위의 그림은 이번 분석과정을 담은 스트림이다.

Ⅳ. 분석결과

1. 성별 및 성향

통계 시스템인 SPSS WIN 10.0을 이용하여 2008년 05월 20일부터 05월 28일까지 8일간을 대상으로 분석하였다.

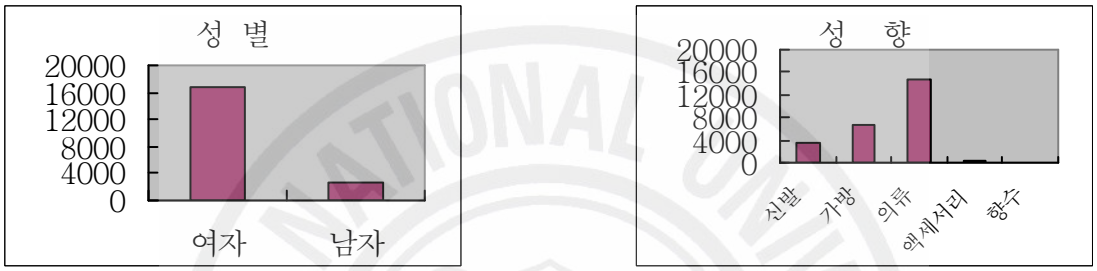


Fig. 6. 성별 및 성향

쇼핑몰을 이용하는 방문자 패턴이 남자 보다 여자의 비율이 현저히 높게 나타났으며 의류전문 쇼핑몰 특성상 의류부분에서 성향이 높게 분석 되었다.

2. 회원, 비회원 수

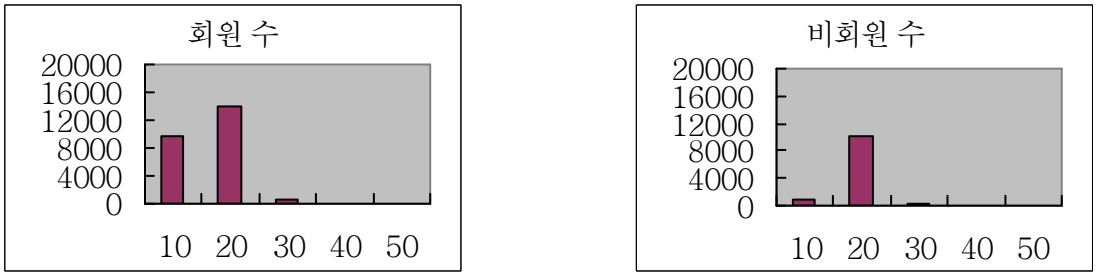


Fig. 7. 회원, 비회원 수

일반 회원인 경우는 20대의 비율이 높게 분석 되었으며, 한편 비회원의 경우는 쿠키에 정보가 기록되어 쿠키 정보를 토대로 로그파일을 분석할 수 있었다.

3. 페이지 빈도 수

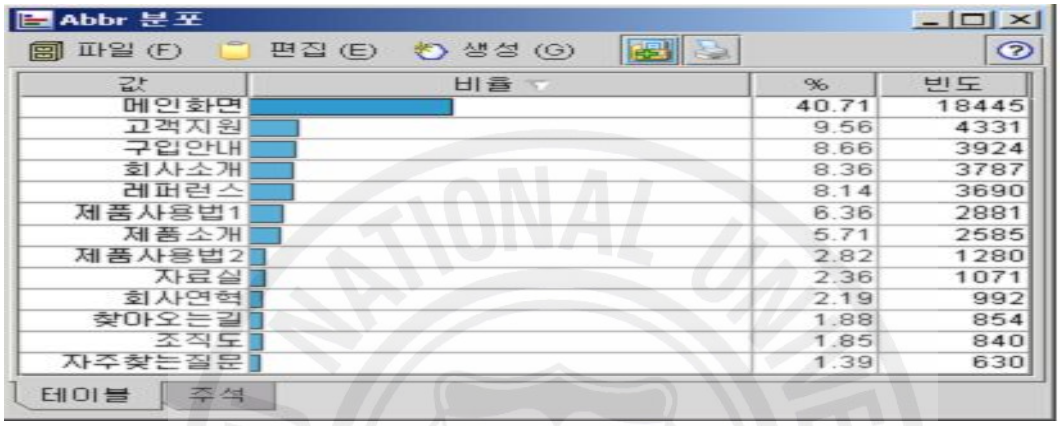


Fig. 8. 페이지 빈도 수

가장 많이 방문한 페이지로는 메인화면으로 분석되었다. 메인화면의 경우 사이트 정보탐색을 하기 위한 시작페이지이기 때문에 많을 수 밖에 없다. 이외에도 웹 사이트의 제품에 관한 기본적인 정보를 얻고자 고객지원과 구입안내, 회사소개 등의 페이지에 접속이 많다. 이러한 경우 제품에 대한 정보 부분을 명확히 설정하지 못했다는 것을 알 수 있다. 그렇기 때문에 제품에 대한 정보와 철저한 고객지원을 보강하여 방문자를 유도할 수 있도록 해야 한다. 특히, 고객지원 페이지에 많은 사람이 방문하므로 고객지원에 관하여 정기적인 정보를 제공하여 충성고객을 확보할 수 있다.

4. 재접속 수

1회 방문과 오늘 최초 방문의 데이터를 Target로 정의할 때 데이터 변환 시 같은 시간대를 의미하며, 클릭 여부를 가지고 Target의 속성을 분류할 수 있다.

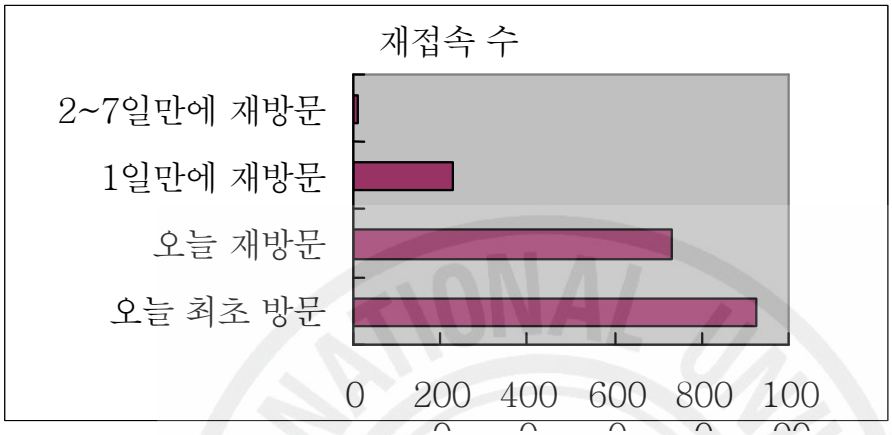


Fig. 9. 재접속 수

Ⅴ. 결 론

인터넷 쇼핑몰의 사용자가 쉽고 편리하게 인터넷 쇼핑몰을 사용할 수 있도록 하기 위한 사용편성 증대가 중요한 문제로 대두되고 있으며 실제 사용자의 웹서핑과 검색행동의 촉진을 위한 설계가 절실히 요구되고 있다. 그러나 인터넷 쇼핑몰의 사용편의성도모를 위한 기존연구방법들이 많이 적용되어 왔지만 이러한 방법론들은 쇼핑몰을 사용하는 사용자의 활동과정중의 패턴 즉, 탐색, 검색에 따르는 과정을 반영하지 못하므로 인하여 소비자의 사용패턴과 인터넷 쇼핑몰의 설계불일치가 지속되고 있다.

본 연구에서는 이러한 고객의 인터넷 쇼핑몰 사용패턴을 분석하는 모델링 방법론을 제시하고 데이터마이닝의 여러 기법을 관련 연구로 소개하였으며,

본 연구에서는 의류 전문 온라인전문몰인 http://on-pedder.com사를 대상으로 로그분석을 통한 단순한 수치를 분석하는 것 이외에 전반적인 인터넷 비즈니스 전략 구상 및 고객대응을 위해 방문자의 클릭횟수 별 이동경로 분석을 누구나 쉽게 할 수 있도록 Clementine과 SPSS WIN 10.0 를 사용하여 기본적인 데이터 탐색을 보여주었다.

이러한 결과가 쇼핑몰 설계에 주는 시사점은 다음 3단계와 같다.

1단계, 소비자가 처해있는 기술적 환경 경우에 맞추어 설계 되어야 한다.

2단계, 사용자들이 전문몰에 대해 기대하는 바는 충분한 정보검색이다. 따라서 정보검색이 구매로 연결될 수 있게 하는 페이먼트 시스템의 정립이 중요한다.

3단계, 재방문 시 환영메시지를 송부할 수 있는 시스템구축이 필요하다.

하지만 본 연구에서는 제시된 모델링 방법론을 실제 인터넷 쇼핑몰 설계에 적용하지는 못했다. 이것은 본 연구의 한계이며. 향후 연구에서는 제시된 모델링 방법론을 가지고 기존 쇼핑몰의 소비자적합적인 재구축과정이 실제로 행해져야 할 것이다.

참고문헌

[1] 강현철 외 4인, SAS Enterprise Miner4.0을 이용한 데이터마이닝, 서울 : 자유아카데미, 2001

[2] 김광용 외 1인, 인터넷 설문조사를 활용한 사이버 쇼핑몰 디자인에 관한 연구,경영정보학연구, 1999

[3] 김성태, 인터넷 쇼핑몰 로그 데이터를 활용한 마이닝 분석 연구, 경희대, 2004

[4] 김재형, 인터넷비즈니스 기반의 고객관리(CRM)을 위한 웹 로그분석에 관한 연구, 한양대, 2000

[5] 김주민, 웹페이지 방문자의 성향 분석을 위한 웹 로그 서버 시스템의 설계 및 구현, 광운대 정보통신대학원, 2004

[6] 김형택 외 1인, 효과적인 인터넷 마케팅을 위한 웹 로그 분석, 비비컴, 2001, P.18, PP28-31, PP38-39, PP260-261

[7] 남궁 영, 웹 로그분석을 통한 이러닝 효율화 방안에 관한 연구, 단국대, 2005

[8] 명노해 외 2인, 웹마이닝을 통한 웹사이트의 사용편의성에 관한 연구, 대한인8공학회 춘계학술대회 논문집, 2001

[9] 박종명, 인터넷 쇼핑몰에 대한 소비자의 신뢰와 만족이 구매의도에 미치는 영향],한양대 대학원, PP.3-6, 2002

[10] 박희석 외 1인, 웹사이트 메뉴 Depth를 줄이는 방식간의 비교 분석, 대한인간공학회지, vol 19, n03, 2000

[11] 변대호 외 2인 AHP를 이용한 사이버몰의 평가, 대한경영학회지, 24, June 10. 10. 전자상거래의 역기능 개선을 위한 주요 실패요인 분석, 경영정보과학연구, 제8권 제1호, 1998, 2000

[12] 엄용환, 웹 로그분석에 대한 고찰, 성결대학교 논문집 제30권, PP.351-361, 2001

[13] 이상준, 웹 로그분석 사례연구 : 이동경로분석 관점에서, 단국대, 2004

[14] 이학식 외 1인, SPSS10.0매뉴얼, 서울 : 법문사, 2001

[15] 이충근,“인터넷 상거래 사이트의 유용성에 관한연구”,1998

[16] 장남식 외 2인, 데이터마이닝, 서울 : 대청미디어, 1999

[17] 최국렬 외 8인, 데이터마이닝 이론과 실습, 서울 : 청구문화사, 2001

[18] 최재혁 외 1인, 웹 사용편의성 평가, 대한인간공학회 추계학술대회 논문집, 2000

[19] 최종후 외 3인, Answer Tree를 이용한 데이터마이닝의사결정나무분석, 서울 : 고려정보산업, 1998

[20] 향희철 외 3인, 웹 사용성 평가 checklist개발, 대한인간공학회, 추계학술대회, 논문집, 1998

[21] 한국전산원 http://www.nca.or.kr

[22] 허준 외 1인, 클레멘타인을 이용한 데이터마이닝 : 입문편, 서울 고려정보산업, 1999

[23] Bloch, M., Pigneur, Y., and Segev, A. On the road of Electronic Commerce, March 1996, (http://www.sterm.nyu.edu/~mbloch/docs/roadtoec/ec.htm).

[24] Large, Tedd & Hartley, Information seeking in the online age principles and practice, London:Bowker-Saur, 1999

[25] Tmax Soft, WebtoB Administration Guide(Ver3.1), Tmax Soft, 2002

감사의 글

힘차게 시작했던 2년 반 동안의 대학원 생활이 벌써 하루하루 흘러 어느덧 아쉬운 끝으로 다가 왔습니다. 2년 반 전만 해도 대학원이라는 생활이 참으로 힘들거라는 두려움으로 시작했지만 대학원 과정을 같이한 동기들과 선배님과 후배님 덕분에 생활 할 수 있었습니다.

2년 반 동안 지도해 주신 곽호영 교수님, 이상준 교수님, 김장형 교수님, 안기중 교수님, 변상용 교수님, 송왕철 교수님, 김도현 교수님, 변영철 교수님, 논문을 쓰며 포기라는 단어가 나오려 할 때 마다 격려와 용기를 주신 한경복 선배님, 권훈 선배님, 정은경 선생님, 혜선이, 연구실 선·후배님들께 감사의 마음을 전합니다.

또한, 나의 사랑하는 가족이 없었다면 이 논문은 완성되지 못했을 것입니다.

매일 새벽까지 논문 쓰느라 지친 나를 위해 미성이를 돌봐주시고 반찬거리도 챙겨주신 친정아버지, 어머니 정말 감사하고, 사랑하고, 존경합니다.

동필오빠, 윤희언니, 남영형부, 동준형부, 춘미언니, 유미언니, 영미, 영찬, 영재, 주연, 대균, 서희, 뱃속에 있는 서희동생, 시아버님, 어머님, 민정고모, 수범 큰아버지께 감사하고, 사랑합니다.

출근할 때 아침식사를 챙겨주지 못해도 잔소리 한번 안하는 나의 사랑 경범오빠, 맛있는 간식 제대로 못해줘도 무럭무럭 잘 자라준 미성, 제대로 된 태교 못해줘도 잘 크고 있는 혜성(태명)이에게 정말 감사하고, 미안하고, 존경하고, 사랑합니다.

앞으로는 아내로서, 엄마로서 최선을 다하고, 사랑하는 나의 가족들을 잘 보살피겠습니다.

이 논문을 나의 보물 경범오빠, 미성, 혜성이에게 바칩니다.

2008년 6월