

Web 로그분석과 e-CRM

㈜ 이씨마이너 연구소장 이성백

1. 서론

(1) eCRM과 시장 환경의 변화

eCRM은 온라인 상에서 수집한 고객 데이터를 저장 및 분석하여 가치 있는 고객을 선별하고 회사의 역량을 가치 있는 고객을 획득, 유지하는 일에 투자하는 프로세스를 말합니다. 기존의 CRM과 목적은 같으나 채널이 오프라인에서 온라인으로 진화하였다는 점이 다릅니다.

채널이 온라인으로 진화하였다는 점이 의미하는 바는 무엇일까요?

먼저 시장을 생각해 보기로 합니다. 과거에는 상품을 중심으로 시장 점유율의 확대를 회사 실적의 향상으로 간주하였습니다. 더 질 좋고 우수한 상품을 만들어서 경쟁사보다 더 많이 팔면 된다는 것이 기본 전략입니다. 어떻게 하면 좋은 상품을 잘 만들어서 많이 판매할 것인가가 기본적으로 업체 입장에서의 이슈였고 고객의 관심 및 요구는 이후의 문제였습니다. 사실 고객의 관심 및 요구가 중요하지 않았기 때문에 부차적인 문제였기 보다는 이에 대한 조사 및 분석 비용이 너무 높았기 때문에 우선 순위에서 밀렸던 것이었습니다.

그러나 eCRM이 도입된 환경에서는 모든 것이 근본적으로 달라졌습니다. 기업과 고객의 채널이 온라인 상으로 형성되면서 고객은 선택의 폭이 넓어졌고 반대로 기업은 고객에 대한 지식을 획득할 수 있는 기반이 마련되었습니다.

고객 정보를 수집할 수 있는 기반이 마련되었기 때문에 모든 마케팅 전략도 근본적으로 바뀔 수 밖에 없게 되었습니다. 과거의 주먹구구식의 상품 전략 및 마케팅 전략은 운이 좋지 않고서는 더 이상 시장에서 좋은 반응을 기대하기 어렵게 되었습니다.

대신 고객의 수요를 정확하게 읽고 이러한 데이터를 바탕으로 과학적이고 분석적인 마케팅 전략을 수립하는 업체만이 시장에서 살아 남을 수 있게 된 것입니다.

(2) eCRM과 로그 분석

이러한 고객 정보 수집의 출발은 고객들이 온라인 상에서 행위를 하고 남긴 흔적인 '로그'에서 출발합니다. 현재의 온라인 환경이 주로 Web으로 이루어졌기 때문에 여기서의 로그도 주로 '웹 로그'를 의미하게 됩니다. 주로 방문객이 어떤 사이트를 돌아 다니며 어떤 페이지를 주로 읽는가가 분석 대상이 됩니다. 이러한 클릭스트림 외에도 게시판에 글을 쓰거나 채팅을 한다거나 이 메일을 쓰는 등 온라인 상에서의 고객 행위도 다양하며 이러한 모든 행위의 흔적이 고객을 이해하는 귀중한 기초 자료가 됩니다.

2. 로그분석의 개념

사용자가 웹사이트를 이용하면 이에 대한 기록이 로그라는 형태로 흔적이 남게 됩니다. 로그분석이란 이 데이터를 기반으로 위에서 말한 다양한 정보를 추출해 내는 것이라 할 수 있습니다.

그런데 로그분석은 사용자에 따라 단지 로그 정보를 분석하는 것에 한정시키기도, 로그 정보를 기반으로 한 보다 다양한 정보를 분석하는 확장된 개념으로 확대시키기도 합니다.

분류	내용
<p>좁은 의미의 로그 분석</p>	<p>좁은 의미의 로그분석은 분석 대상을 Web log에 한정 짓고 이에 대한 분석을 실시하는 것을 말합니다. 기본적인 페이지뷰부터 방문자의 클릭스트림 및 방문자 수, 참조 페이지 등을 Parameter 값으로 합니다. 현재 주요 로그 분석 툴들이 분석하는 값들입니다.</p> <p>최근에는 방문자 관계 관리(Visitor Relation Management)가 중요해지면서 방문자를 중심으로 로그 데이터를 추출 및 관리하는 경향이 대두되고 있습니다. 이는 다시 Login한 사람과 Login하지 않은 사람에 대해 접근 방법이 다른데, 전자의 경우는 사용자가 확인되었으므로 바로 사용자의 Web Usage 현황이 파악 가능한 반면에 후자의 경우는 쿠키를 사용하여 간접적으로 사용자 확인을 할 수 있습니다. 물론 쿠키는 브라우저 동일성만 파악해 주기 때문에 같은 PC를 다른 사람이 사용하는 경우에는 동일한 사람으로 간주된다는 문제점은 있습니다.</p> <p>이렇게 확인이 된 사용자의 Web Usage 현황을 파악하여 ID 기반으로 Log data를 관리하게 되면 실질적으로 우리 사이트에 주로 방문하는 사람은 누구이며 이들이 주로 관심을 가지고 보는 페이지는 무엇인지 파악할 수 있게 됩니다. 개별 방문자와 Relation을 맺고 one to one 마케팅을 수행해 나갈 수 있는 기반이 됩니다.</p>
<p>넓은 의미의 로그 분석</p>	<p>이에 반해, 넓은 의미의 로그 분석은 단지 로그 데이터뿐 아니라, 웹사이트에서 보유하고 있는 고객 등록 정보(Master DB), 구매 정보(Transaction DB) 등을 모두 대상으로 하는 분석을 말합니다.</p> <p>특히 거래 정보를 기반으로 하는 고객 분석은 데이터 충실도 측면에서 순수 로그 분석보다 우월하다는 특징이 있습니다. 또한 Web Log DB가 Master DB와 연동되는 경우 인구 통계학적으로 Web Usage 현황의 특징을 파악해 볼 수도 있으며 Case Based Reasoning 기법 등을 활용한 데이터 마이닝 작업을 통해 컨텐츠 및 상품 등을 추천할 수 있습니다.</p>

3. 로그 분석의 활용

로그파일은 사이트를 방문한 사람들의 방문 데이터를 서버가 인식하여 차곡차곡 방문자 기록을 정리한 하나의 데이터입니다. 이러한 데이터는 우리가 인식할 수 있는 문자형태로 되어있는 것이 아닌 암호처럼 거의 대부분이 숫자로 되어 있습니다. 때문에 자칫하면 로그 분석을 시스템적 관점에서 접근하는 경우가 많습니다. 가령 전송 데이터의 바이트가 어느 정도인지, Page view 및 Hit의 관계는 어떠한지 등의 경우가 그러합니다.

그러나 로그 데이터를 분석하는 목적은 결국 방문객들이 웹 사이트를 방문하면서 남긴 흔적들을 분석함으로써 고객들이 관심을 가지는 내용을 파악하고 고객들의 반응을 예측하여 인터넷 비즈니스에 전략적으로 활용할 수 있도록 하자는 것입니다. 우리가 로그 분석 툴을 도입하여 로그 분석 작업을 할 때 이와 같은 목적을 분명히 하지 않으면 분석 작업은 분석 작업으로만 그칠 수 있기 때문입니다.

그러나 분석 목적을 명확하게 세우고 로그 분석 결과를 활용할 때, 로그 분석의 의의를 찾을 수 있습니다. 로그 데이터의 전략적 활용 방안을 소개하면 다음과 같습니다.

첫 번째로 고객들의 성향을 파악하여 고객들이 원하는 맞춤 서비스를 제공할 수 있다는 것입니다. 고객의 다양한 요구를 예측하여 새로운 사이트를 개발하거나 새로운 시장 기회를 창출하게 됩니다

로그 분석에서 고객의 성향은 암묵적으로 파악하게 됩니다. 고객들이 주로 즐겨 찾는 콘텐츠는 무엇인가? 일정한 고객 집단별로 즐겨 찾는 정보는 무엇인가? 고객들이 주로 입력한 키워드는 무엇인가? 방문 횟수가 높은 고객은 누구인가? 등의 질문에 대해 로그 분석을 통해 답변할 수 있게 됩니다.

가령 고객들이 각종 마케팅이라는 키워드를 사이트에서 자주 찾는다면 고객들은 ‘마케팅’에 관한 자료에 관심이 많다는 것을 알 수 있습니다. 마케팅 관련 정보를 사이트 메인 화면에 Link를 걸어 놓거나, e-mail로 고객에 보내 줌으로써 고객에 대한 사이트의 관심을 증진시키고 신규 회원을 더 확보할 수 있게 됩니다.

두 번째로 사이트의 정보 설계를 경제적으로 할 수 있게 됩니다. 로그 분석을 통해 잘 찾지 않는 콘텐츠는 서브 화면으로 돌려 놓거나 폐기하는 방법을 통해 무분별하게 콘텐츠를 확장해서 방문객을 혼란스럽게 하는 일을 사전에 방지할 수도 있습니다. 특히 사이트 디자인에서는 항해 설계가 중요하게 되는데 로그 분석을 통해서 사이트 리모델링에 참고할 수 있게 됩니다.

세 번째로 특정 타겟에 대한 집중적인 마케팅 및 광고전략이 가능하게 됩니다. 마케팅 및 광고전략은 기존의 매스마케팅이나 매스 광고처럼 무차별적으로 광고집행을 하는 것이 아닌 방문자의 방문경로 및 페이지뷰, 임프레션이 많은 타겟 페이지에 집중적으로 타겟에 맞는 광고를 집행하여 효과를 보는 것을 말합니다. 또한 방문자들이 많이 몰리는 시간대를 분석하여 그 시간대에 집중적으로 광고를 노출시킨다면 광고주의 입장에서 최상의 효과를 노릴 수 있습니다. 최근 배너 광고의 효과에 대한 회의론이 제기되고 있는 시점에서 고객 분석에 근거하여 적절하게 타겟 광고를 실현할 수 있다면 광고주 입장에서는 경제적으로 광고 비용을 사용할 수 있어 좋게 되고, 고객 입장에서는 광고가 스팸이 아닌 자신에게 필요한 정보 제공의 형태로 수용될 수 있어 좋습니다. (Opt in Marketing, Permission Marketing)

네 번째로 최적의 환경에서 사용자들이 사이트를 탐색하고 방문하도록 서버 및 회선 등의 기술적 자원을 계획하거나 수립할 수 있습니다. 처음에 웹호스팅으로 사이트를 구축하였으나 사용자들이 점차

늘어나면서 사용자들의 데이터베이스를 처리하는데 문제가 발생하는 경우 서버 호스팅을 하거나 자체 서버를 가져 오게 될 수 있습니다. 서버가 사용자가 갑자기 늘어나 요구를 감당할 수 없기 전에 Traffic 분석을 통해서 사전에 서버를 늘리는 방법 등을 통해 대응을 할 수 있습니다.

그 이외에 깨진 링크나 삭제된 페이지 등을 관리할 수 있기 때문에 방문객들이 404 Not Found에 접하지 않도록 합니다. 사이트의 페이지가 수백 페이지에 이를 경우 링크 관리는 로그 분석을 통해 하는 것이 보다 효율적인 방법입니다.

4. 로그 분석의 한계

로그 분석은 웹 사이트 운영에 대한 블랙 박스(Black Box)와 같이 웹 사이트를 통해 이루어지는 모든 정보를 정확하게 알 수 있는 개념은 아닙니다. 로그 데이터 그 자체만으로 분석할 수 있는 정보는 그다지 많지 않은 것이 사실입니다. 또한 로그 데이터 그 자체로는 금융권의 데이터와는 달리 부정확하다는 특징을 가지고 있고 이 점을 분명히 인식하고 있어야 합니다.

예를 들어 로그 데이터에 저장되는 클라이언트 IP를 살펴 봅시다. 클라이언트 IP는 사용자가 인터넷을 시작한 위치 정보에 해당 됩니다. 클라이언트 IP정보를 통해 웹 사이트에 접속하는 사용자들의 정보를 파악할 수 있습니다. 하지만 사용자가 유동 IP를 사용할 경우, 또는 Proxy Server를 이용할 경우 등에서는 정확한 클라이언트 IP를 파악할 수 없습니다. 또 캐쉬를 통할 경우 사용자는 페이지를 봤지만 로그에는 그 데이터가 저장되지 않습니다. 그리고 뒤로 가기(BACK) 통해 이동할 경우도 그 행위들은 로그에 기록되지 않습니다. (물론 네트워크 단계에서 Packet Sniffing 방식으로 추적이 가능하며 혹은 Server Plug in 방식으로도 Back Button의 클릭 현황을 감지할 수 있습니다.)

쿠키의 경우도 브라우저 동일성은 파악하지만 1 PC에서 2인 이상이 사용할 경우에는 정확하게 사용자를 확인할 수는 없습니다. 기타 클라이언트의 쿠키 거부라든가 개인 정보 보호를 위한 쿠키 삭제 프로그램에 대해서는 역시 대응 방안이 없습니다.

이러한 한계를 극복하기 위한 가장 좋은 방법은 온라인 상의 고객 정보 수집을 단순 웹 로그에 국한하지 않고 다양한 채널을 통해 고객 정보를 수집하여 통합하는 것입니다. 가령 콜센터, 게시판, 이메일, 핸드폰, 채팅 등 고객과의 온라인 상의 접점은 결코 웹에 국한되지 않습니다.

고객 정보 수집이라는 목적 하에 다양한 채널을 통해서 고객 정보를 수집 통합한다면 로그 분석이 가지고 있는 한계를 극복할 수 있습니다.

5. 로그 데이터 일반

(1) 특정페이지에 관련되는 정보가 로그파일에 저장

일반적으로 특정 웹 페이지를 보기 위한 사용자의 요구로, 웹 서버는 해당 웹 페이지와 관련된 여러 파일등에 접근하게 됩니다. 따라서 사용자가 요청하는 특정 웹 페이지 뿐만 아니라 해당 웹 페이지와 관련된 이미지 파일, 이미지 데이터, Include 파일 등에 대한 정보가 로그 파일에 저장되는 것입니다.

로그 파일에는 웹 서버가 수행한 작업들에 대한 풍부한 정보가 들어 있습니다. 이들은 단순히 특정

작업 요청과 성공 여부에 대한 것 뿐만 아니라 실패 했을 경우 그 해결책에 대한 정보도 들어 있습니다.

그리고 로그 파일에 대한 활용이 여기에서 그치는 것이 아니라 그 분석 능력에 따라 웹 사이트 성공에 필요한 많은 정보를 얻을 수 있습니다.

사용자가 웹 사이트에 접속한 후 모든 작업들은 웹 서버를 통해 이루어 집니다.. 따라서 로그 데이터를 제대로 이해하기 위해서는 사용중인 웹 서버의 특성과 관리 등에 대해서 살펴 볼 필요가 있습니다.

(2) Web server 운영 현황

현재 세계적으로 운영 중인 웹 서버는 Apache가 대략 60% 정도 사용되고 있으며 그 다음 Microsoft-IIS 서버의 사용 비중이 21% 정도가 됩니다. 두 서버가 Web Server 시장에서 대략 80% 이상을 차지하고 있다고 볼 수 있습니다. 따라서 여기서는 Apache와 IIS를 중심으로 살펴 보도록 하겠습니다.

(3) 로그 데이터 포맷

로그 파일에 저장되는 항목들은 웹 서버에 따라 조금씩 다릅니다. 가령 Microsoft IIS는 로그 파일을 `\\Winnt\System32\Logfiles\W3SVC`에 저장하며, Netscape, Apache 와 같은 NCSA 계열의 웹서버에서는 로그파일을 Document Directory와 같은 장소의 Logs라는 Directory에 `access_log`라는 이름으로 기록하는 것이 일반적입니다

초기에 웹 서버에 저장하는 로그 데이터 항목들은 웹 서버마다 달라서 로그 분석에 어려움이 많았습니다. 항목들이 다르기 때문에 한 로그 분석 프로그램이 모든 로그 파일을 분석할 수 없습니다. 그래서 등장한 것이 Common Log Format(CLF)입니다.. 이 경우 공통적인 형식으로 로그 데이터를 저장하기 때문에 웹 서버가 다르다 하더라도 로그 데이터 형식은 비슷하게 됩니다. 그리고 그 항목들은 대부분의 상황에 필요한 항목들을 제공합니다. 하지만 CLF 이외에도 해당 웹 서버가 고유하게 제공하는 형식도 있습니다.

현재의 대부분의 웹 로그는 CERN과 NCSA 에서 HTTP 프로토콜로 규정한 Common Log Format 를 따릅니다.. 이 표준을 따르는 로그 항목에는 사용자의 IP 주소, 데이터 전송을 위한 프로토콜, 에러 코드, 전송된 데이터 길이 등이 포함됩니다.

6. 로그 데이터 형식 - Apache

Apache 는 기본적으로 CLF 형식으로 저장합니다. 필요한 경우 LogFormat directive를 이용하여 다른 형식으로 저장할 수 있습니다.

(1) Common Log Format

Common Log Format (CLF) 형식은 스페이스로 구분되면 다음과 같은 형식으로 기록됩니다.

(만약 값은 가지지 않는 항목들은 ' - ' 로 기록됩니다.)

host	ident	authuser	date	request	status	bytes
------	-------	----------	------	---------	--------	-------

Host	클라이언트의 도메인 이나, IP 주소
ident	IdentityCheck 가 활성화되어 있고 클라이언트 측에서 identd를 실행하고 있다면 클라이언트 측으로부터 제공되는 정보가 저장된다.
autouser	만약 해당 사이트가 기본적인 HTTP 인증을 요구한다면 이 내용이 사용자 ID가 된다.
date	날짜. 포맷은 아래와 같다. [day/month/ year : hour : minute : second : zone]
Request	(" ")로 묶여 있으며 클라이언트 측에서 요구한 정보이다
Status	클라이언트로 보내지는 상태 코드
Byte	헤더 정보를 제외하고서 전송된 데이터의 크기

(2) Error Log

로그 항목에서 Error Log는 중요한 항목 중 하나입니다. 에러에 대한 정보를 기록하지 않을 경우 무엇이 잘못되었는지, 어디에서 잘못 되었는지 알 수 없게 됩니다.

Error Log는 서버가 에러를 만날 때마다 그 정보를 기록할 파일을 지정합니다.. 만약 파일명이 “ / ” 로 시작되지 않으면 서버측의 루트 디렉토리에 상대적인 경로로 인식됩니다.

(3) Referrer Log

방문자가 어디를 통해서 사이트에 방문했는지에 대한 정보는 매우 유용한 정보가 될 것이다. 이 정보는 특히 다른 사이트에 배너 광고를 했을 때 그 배너 광고의 효과가 어느 정도인지 파악하고 광고 정책을 세우는 데 핵심적인 자료가 될 것입니다.

Refer Log 정보는 해당 웹 페이지를 요청하기 전에 어디에 있었는지 대한 정보이다. 외부 뿐만 아니라 내부적으로 네비게이션을 파악하는 자료로도 쓰일 수 있습니다. 그러나 Exit Page에서 어떤 사이트로 방문객이 갔는지에 대한 정보는 주지는 못 합니다.

7. 로그 데이터 형식 - IIS

(1) 기본적인 파일 형식

IIS는 기본적으로 네 가지 로그 형식을 지원하게 됩니다. 마이크로소프트 IIS 로그 파일 형식(Microsoft IIS Log File Format), NCSA 공용 로그 파일 형식(NCSA Common Log File Format), W3C 확장 로그 파일 형식(W3C Extended Log File Format), ODBC 로깅(ODBC logging) 등이 있습니다.

이 중 Microsoft IIS Log Format 과 W3C Extended Log File Format 는 마이크로 소프트에서 정의한 것이고 둘 다 ASCII 형식을 지원합니다. ODBC Logging 는 데이터 베이스에 직접 로그 데이터를

저장하는 방식입니다.

파일 이름들은 로그 데이터 형식에 따라 다르게 기록 되는데 이름을 가지고 어떤 형식 설정인지 대략 파악할 수 있습니다. 예를 들어 ex20000327.log 의 경우 W3C 확장 로그 형식이고 로그 파일의 생성 시점은 하루 단위인 것을 알 수 있습니다. 파일 형식을 대략 살펴 보면 맨 앞 부분(예 in,nc,ex...)은 로그 형식을 나타내고 그 뒤의 숫자는 로그가 기록된 날짜 시간을 나타냅니다.

*in : 마이크로소프트 IIS 로그 형식
*nc : NCSA 공용 로그 파일 형식
*ex : W3C 확장 로그 파일 형식
*yymmddhh : 시간별 로그
*yymmdd : 일별 로그
*yymmww : 주별 로그
*yymm : 월별 로그 (yy는 년,mm는 달, dd 일,hh 시,ww 주를 의미)

(2) 로그 파일 생성 시점과 형식

로그 파일이 생성되는 시점은 매시간,매일,매주,매월 등으로 필요에 따라 설정할 수 있습니다. 기준이 되는 것은 로그 데이터 양이 됩니다.. 로그 데이터가 많을 경우(트래픽이 많다는 의미) 로그 파일 크기가 너무 커지게 되기 때문에 관리하기 불편해질 수 있습니다

ODBC Logging를 제외하고(이 경우 데이터에 직접 저장하기 때문) 모두 ASCII 형태로 저장되기 때문에 다른 파일 형식으로 쉽게 변환해서 사용할 수 있습니다.

(3) ODBC Logging

로그 파일에 저장할 경우 다양한 형태의 분석 툴을 이용할 수 있는 장점이 있지만 그래도 ODBC Logging 형식을 통해 직접 데이터 베이스에 저장 하기를 원할 수 있습니다. ODBC Logging을 이용하면 웹 서버에 이루어 지는 모든 요청 때마다 데이터에 저장하기 때문에 파일에 저장하는 것보다 서버에 많은 부담을 주게 된다. 일반적인 경우는 권장하지 않습니다. 저장되는 데이터는 ASCII 파일을 사용하는 것과 동일합니다.

참고적으로 SQL을 모르는 사용자들을 위해 필요한 컬럼 들로 구성된 INETLOG 테이블을 만들 수 있는 파일이 있습니다. %WinDir%\System32\logtemp.sql이 그 파일인데 파일에 들어 있는 내용을 그대로 쿼리 툴에서 그대로 실행 시키면 됩니다.

(4) W3C 확장 로그 파일

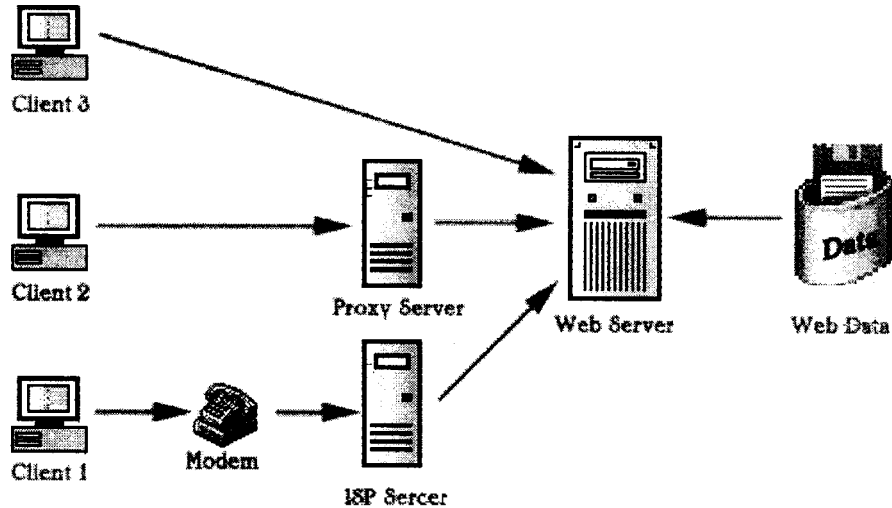
IIS 로그 파일 형식 중 사용 정도에 따라 가장 효율적으로 사용할 수 있는 W3C 확장 로그 파일 형식에 대해서 알아 보기로 합니다..특히 이 로그 포맷은 ID matching 기술을 적용하기 위해서 필요한 형식입니다.

W3C 확장 로그 파일 형식의 특징은 사이트 성격에 따라 로그에 저장할 항목들을 선택할 수 있다는 것입니다.

항목	설명
Date	활동이 발생한 날짜
Time	활동이 발생한 시간 (시간대별 사이트 접속 사용자 현황 파악 및 사이트에 머무른 시간 등의 분석)
Client IP	서버를 액세스한 클라이언트의 IP 주소 웹 서버에 접근한 사용자의 IP 주소. IP 주소 자체는 아무 의미도 없지만 Session ID와 연계된 Table을 만들 경우 사용자 확인의 좋은 수단이 됨.
Service Name	클라이언트 컴퓨터에서 실행되었던 인터넷 서비스
Server Name	로그 항목이 생성된 서버 이름
Server IP	로그 항목이 생성된 서버의 IP 주소 웹 서버가 몇 대의 서버로 이루어져 있을 경우 구별 정보로 쓸 수 있음
Server Port	클라이언트가 연결된 포트 번호
Method	클라이언트가 시도한 작업(예: GET 명령)
Stem	액세스한 리소스 (예: HTML 페이지, CGI 프로그램 또는 스크립트) 사용자가 요청한 페이지에 대한 모든 내용들이다. 실제 웹 사이트를 이루고 있는 파일 이름들이다. 예를 들어 default.asp 를 열었을 때 default.asp를 이루고 있는 default.asp, 이미지 파일(.gif, .jpg등), Include 파일(.css, .js 등), 자발 애플릿 (.cg, .CGI 파일 등의 이름이 모두 기록된다.
Query	클라이언트가 시도한 쿼리. 즉, 클라이언트가 일치하는 대상을 찾으려 한 하나 이상의 문자열 이 부분은 해당 파일에서 필요한 파라미터가 기록되는 부분이다. 예를 들어 검색하기 위한 search.asp 가 있다고 하자. Search.asp 는 Keyword에 해당하는 파라미터가 필요하다. Search.asp?keyword=internet 일 경우 URL Stem 항목에는 Search.asp가 URI Query에는 keyword=internet가 기록 된다. 대부분의 경우 간과 하기 쉽지만 URI Stem을 제대로 파악하기 위해 꼭 필요한 항목이다.
HTTP status	HTTP 관점에서 본 작업 상태
Bytes Send	서버가 보낸 바이트 수
Bytes Received	서버가 받은 바이트 수
Protocol Version	클라이언트가 사용한 프로토콜(HTTP, FTP) 버전. HTTP의 경우 HTTP 1.0 또는 HTTP 1.1입니다.
User Agent	클라이언트에 사용된 브라우저
Cookie	보내거나 받은 쿠키의 내용
Refer	사용자의 방향을 현재 사이트로 재지정한 사이트 로그에 기록된 파일을 요청하기 바로 전 위치에 대한 정보다. 외부에 사이트에 접근한 경우는 사이트 오기 바로 전 사이트에 대한 URL이 적힐 것이고, 사이트에서 돌아 다닐 경우 사이트의 바로 전 내용이 기록되게 된다. 이 정보는 특히 배너 광고 효과 등을 파악하는데 아주 유용하게 사용될 수 있다.

8. 로그 분석과 운용 환경

웹 로그 분석과 관련된 운용 환경은 어떠한 까요? 다음의 그림을 살펴보기로 합니다.



[그림 1 : 로그 데이터의 소스들]

그림에서 보듯이 웹 로그 분석을 위한 데이터들은 Proxy 서버, 웹 로그 파일, 그리고 운영체 데이터에서 얻을 수 있습니다.

(1) Proxy Server

위 그림에서 모뎀을 통하여 ISP Server를 이용하거나 기업내의 Proxy Server를 이용하는 경우(Client1, Client2)가 이에 해당 됩니다. ISP Server도 Proxy Server 기능을 합니다. Proxy Server는 방화벽으로 더 많이 알려져 있지만 인터넷 사용자들에게는 속도를 빠르게 해주는 캐싱을 해주는 서버로 더 많이 알려져 있습니다.

Proxy Server는 Local과 Remote Web Server(사용하기 원하는 웹사이트) 중간에 위치하면서 Local에서 요구하는 문서를 Proxy서버가 대신해서 Remote Server에 요청한 후 Local로 넘겨 줍니다. 동시에 Proxy Server에 그 문서가 저장됩니다.

다시 같은 내용을 요청하거나 다른 사용자가 같은 내용을 요구할 경우에는 해당 사이트에 요청하는 것이 아니라 Proxy Server가 자신에게 저장된 내용을 사용자에게 전달하게 됩니다. 따라서 많은 사용자가 사용할 경우에는 빠른 웹서핑을 할 수 있게 해 줍니다. Proxy Server에 없는 내용이나 저장 기간이 지난 경우 등은 해당 사이트에서 다시 가져다가 Proxy Server에 저장합니다.

Proxy Server를 이용할 경우 웹 서버에 요구하는 작업등의 기록은 웹 서버 로그 파일이 아닌 Proxy Server에 기록되게 됩니다. 즉 Proxy Server에 저장된 내용을 사용자들이 요청할 경우에는 웹 서버로 접근하지 않고 Proxy Server에 저장된 내용을 바로 보여 주기 때문에 해당 웹 서버에는 전혀 영향을 주지 않게 됩니다.

하지만 Proxy Server에 저장된 내용을 통해 서핑 하다가 Proxy Server에 저장되지 않은 내용을

요청할 경우에는 웹 서버에 해당 페이지를 요청하게 됩니다. 이 경우 처음 접속해서 진행하는 일련의 네비게이션 패턴을 보여주는 것이 아니라 중간부터 시작하는 네비게이션 패턴을 보여 주게 됩니다.

또 IP를 통한 세션 처리를 할 경우에는 여러 사용자가 방문해도 한 사용자로 처리될 가능성이 있습니다. 따라서 이 경우는 쿠키 등 다른 항목들의 도움을 받아 보다 정교한 분석을 해야 합니다.

가능하다면 Proxy Server 로그 데이터를 분석 자료로 사용하면 좋겠지만 Proxy Server의 정보는 일반적으로 얻을 수 있는 데이터는 아니기 때문에 아주 특별한 경우를 제외하고는 Proxy Server 로그 데이터를 분석 데이터로 활용하는 경우는 드뭅니다.

(2) 웹 서버 로그 파일

웹 로그 분석에서 가장 중요한 데이터이고 대부분의 웹 로그 분석 내용들은 웹 로그 파일을 대상으로 한 것입니다.

(3) 운영계 데이터

운영계 데이터는 실제 웹 사이트에서 운영 중인 회원 데이터나 콘텐츠 데이터를 말합니다.

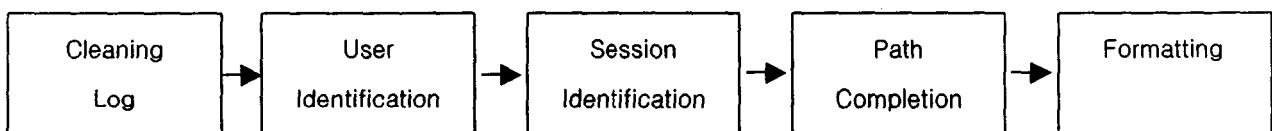
웹 서버에 저장되는 로그 파일에는 사용자가 요구한 파일 이름이 기록됩니다. 하지만 파일 이름만을 가지고 사용자가 정확히 무엇을 요구했는지는 알 수 없습니다. 따라서 로그 파일에 저장된 내용과 운영 계 데이터 연결 시켜 보다 정확한 분석을 할 수 있습니다.

웹 로그 파일 데이터와 운영계 데이터를 연결 시키기 위해서는 로그 파일의 내용을 DBMS에 저장하는 과정을 거쳐야 합니다. 이 과정은 웹 로그 분석을 위한 DW(Data Warehouse)을 구축하는 과정으로 이해할 수 있습니다.

9. ID matching 기법

일반적인 Raw data 수준의 Log 만으로는 사용자 관점에서 바로 유용한 정보가 되지 못 합니다. 일정한 정제 과정을 거친 후 산출된 데이터를 DB에 적재하게 됩니다. 이러한 로그 데이터를 DB에 적재하기 전에 데이터를 전처리 하는 과정은 Filtering Engine에 의해 수행됩니다.

여기서는 ID matching 기법을 이용한 데이터 처리 과정을 살펴 보기로 합니다.



[그림 2 : 로그 데이터 전처리 과정]

(1) Cleaning Log

일반적으로 하루에 천만 페이지 뷰를 기록하는 웹 서버가 하루에 쏟아내는 로그 데이터 양은 20기가 정도 됩니다. 그러나 이렇게 많은 데이터를 모두 스토리지로 가져갈 필요 없이 필요 없는 데이터를 제거하는 과정을 Cleaning Log라고 합니다.

보통 확장자가 gif나 jpeg, jpg등으로 끝나는 그림 파일 들의 로그가 주로 제거되는데 이러한 이미지 요청 결과는 실제로 방문자의 Click Stream과는 무관하다고 보아도 되기 때문입니다.

보통 1/10에서 1/40 정도로 분량이 축소됩니다.




(2) User Identification

보통 사용자 확인의 기준으로 IP가 사용되고 있으나 ID matching 기법에서는 IP보다는 실제 등록 고객의 ID를 기준으로 사용자 확인을 하고 있습니다.

그러나 Login인 된 사용자만을 분석 대상으로 하는 경우 login을 하지 않은 방문자는 분석 대상에서 제외된다는 문제점이 있습니다.

여기에 대해서는 IP 분석법을 활용하거나 쿠키를 활용해서 보완하는 방법이 있습니다. 그런데 쿠키를 활용하는 경우는 웹 서버에 약간의 작업이 필요합니다. 또한 개인 보호 정책 이슈 문제, 2인이 IPC를 사용하는 경우 등의 문제도 있습니다.

한편 방문객이 로그인 하는 순간 사이트의 진입 시점부터의 경로까지 역추적하여 방문객의 클릭스트림을 완성하게 되는데 이 때 IP, Agent(OS, Browser) 등이 역추적 수단으로 활용됩니다.

		
211.40.179.171	Mozilla/3.04(Win95, 1)	A,B,F,O,G, A,D
211.40.179.171	Mozilla/3.01(X11,1,IRIX6.2,IP22)	A,B,C,J
211.40.179.171	Mozilla/3.04(Win95, 1)	L,R

[그림 3 : 두 번째 IP는 브라우저가 달라 다른 사용자로 분류됨]

(3) Session Identification

Session이라 함은 방문객이 사이트에 진입하여 페이지를 거친 후 최종적으로 사이트를 나가게 되면서 이 과정을 하나의 Session 이라고 하게 됩니다.

그런데 방문객이 특정 페이지에서 다른 작업을 하게 되거나 다른 브라우저를 생성하게 되는 경우 실제로는 그 페이지를 보고 있지 않으나 로그 분석 측면에서는 그 페이지를 보고 있는 것으로 간주됩니다.

즉 이미 세션은 종료되었다고 보아야 하는데 종료 시점이 로그 분석에 명확하게 나타나지 않기 때문입니다.

보통 이 경우는 30분 (실제 실험 결과는 25.5분) 정도 경과하였을 때 Session이 종료했다고 간주하게 됩니다. 30분 이후에는 또 다른 클릭스트림이 이어지더라도 이것은 별도의 세션으로 간주하게 되는 것입니다.

이렇게 세션을 초과 시간을 기준으로 분리하는 과정을 Session Identification 이라고 합니다.

211.40.179.171	Mozilla/3.04(Win95, 1)	A,B,F,O,G	Session is split by an excess of time
211.40.179.171	Mozilla/3.04(Win95, 1)	A,D	
211.40.179.171	Mozilla/3.01(X11,1,IRIX6.2,IP22)	A,B,C,J	
211.40.179.171	Mozilla/3.04(Win95, 1)	L,R	

[그림 4 : 시간 초과로 세션이 분리 (현재 그림에서는 Time이 생략됨)]

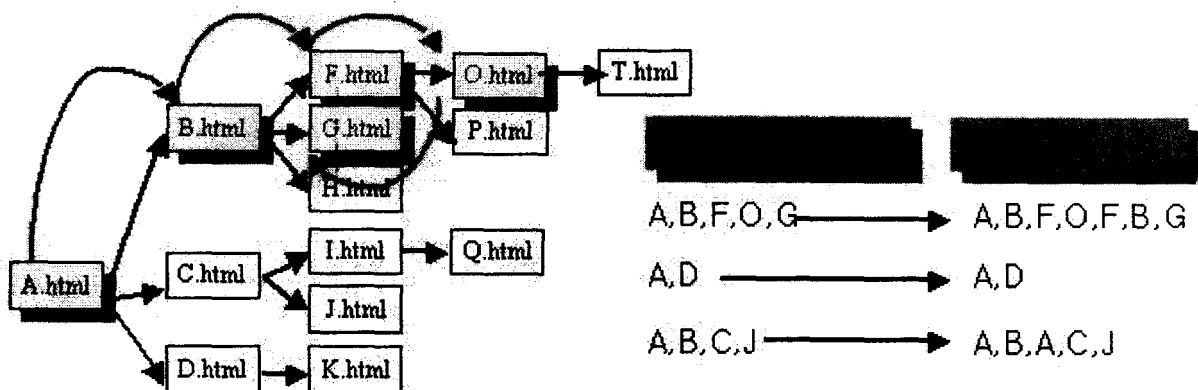
(4) Path Completion

방문객이 페이지를 서핑하다가 백버튼을 누르는 경우가 있습니다. 또 페이지 요청을 한 후 취소 버튼을 누르는 경우도 있으며 페이지 전송이 완전하게 종료되지 않은 경우도 있습니다.

이에 대해서는 Web Log만 가지고는 명시적으로 나타나지 않습니다. 가령 A 페이지에서 B 페이지로 이동한 후 백 버튼을 눌러 A 페이지로 돌아 갔다고 하여도 웹 로그 상에는 돌아간 A 페이지는 나타나지 않습니다. 브라우저에서 보는 A 페이지는 브라우저가 캐쉬한 페이지이기 때문입니다.

이런 경우는 네트워크 단계에서 패킷 등을 잡아서 분석할 수도 있고 서버 플러그 인을 기반으로 분석할 수도 있습니다. 그러나 논리적으로 분석할 수도 있습니다. 가령 위의 사례에서 백 버튼을 눌러 돌아간 A페이지에서 C페이지로 돌아 갔다고 하면 방문자가 실제로 취한 경로는 A->B->A->C이지만 웹 로그 상에는 A->B->C로 나타나게 됩니다.

그런데 B에서 C로 가는 링크가 존재하지 않는다면 B에서 A로 갔다가 다시 C로 이동하였음을 논리적으로 추론해 볼 수는 있습니다.



[그림 5 : 경로를 논리적으로 재구성하는 사례]

(5) Formatting

Web log data를 web mining을 위해 특정한 포맷으로 전환하는 과정입니다. Web Mining 기법이 다양하기 때문에 마이닝 기법에 적합한 로그 데이터 포맷도 각각 존재하게 됩니다.

ID matching 기법은 기본적으로 사용자 ID를 중심으로 데이터를 관리하게 되는데 앞서도 말했듯이 이는 ID matching에 근거한 데이터 관리 포맷이라 생각할 수 있습니다.

ID matching 자체가 데이터 마이닝을 고려한 로그 데이터 관리 포맷으로서 Web DW 구축을 위해 Legacy DB와도 쉽게 연동이 됩니다.

10. 의사결정지원을 위한 로그분석 프로세스

아무리 1차 필터링을 거친 후라고 하더라도 일일 천만 페이지 뷰를 기록하는 사이트에서 2기가 바이트의 로그 데이터가 산출됩니다. 그러나 전략적 의사결정을 지원하기 위해서 이렇게 많은 로그 데이터를 모두 가져갈 필요가 없습니다. 로그 데이터를 효율적으로 관리하기 위해 다음과 같은 분석 과정을 거치게 됩니다.

첫 번째 분석 단계로 사이트 분석 단계가 있습니다. 1차적으로 필터링을 통해 로그 데이터를 1/10에서 1/40으로 줄인 다음 이 로그 데이터를 바탕으로 기존의 로그 틀 들에서도 분석되고 있는 Traffic 및 Referrer, 그리고 System 관련 통계치 들만을 산출하게 됩니다. 가령 우리 사이트의 페이지 뷰는 얼마인가? 방문객 수는 얼마인가? 주요 참조 도메인은 어떻게 되는가에 대한 분석을 수행하게 됩니다.

두 번째 분석 단계로 방문객 분석 단계가 있습니다. 즉 이 과정에서는 방문자들이 주로 취하는 네비게이션 경로는 무엇인가? 또 주로 자주 찾는 페이지는 무엇인가 등의 정보를 찾게 됩니다.

주의할 것은 여기서는 특별히 방문자의 신원 확인이 필요 없다는 것입니다. 여기서 관심을 가지는 것은 일반적인 사용자들의 Web Usage '현황' 이기 때문에 가령 홍길동이라는 사람이 컴퓨터에 관심이 많다는 정보는 특별히 필요하지는 않습니다.

세 번째 분석 단계로 회원 분석 단계가 있습니다. Login을 한 사용자들을 대상으로 개별 사용자의 Click stream, 선호 페이지, 방문 횟수 및 사이트 상주 시간들을 별도로 관리하게 됩니다.

가령 이 회원이 영화 관련 정보를 자주 본다든 사실을 알 수 있고, 또 최근 방문 횟수가 증가하고 있다는 사실들을 파악하여 이 회원이 현재 영화에 대한 관심이 높다는 사실을 알 수 있는 것입니다.

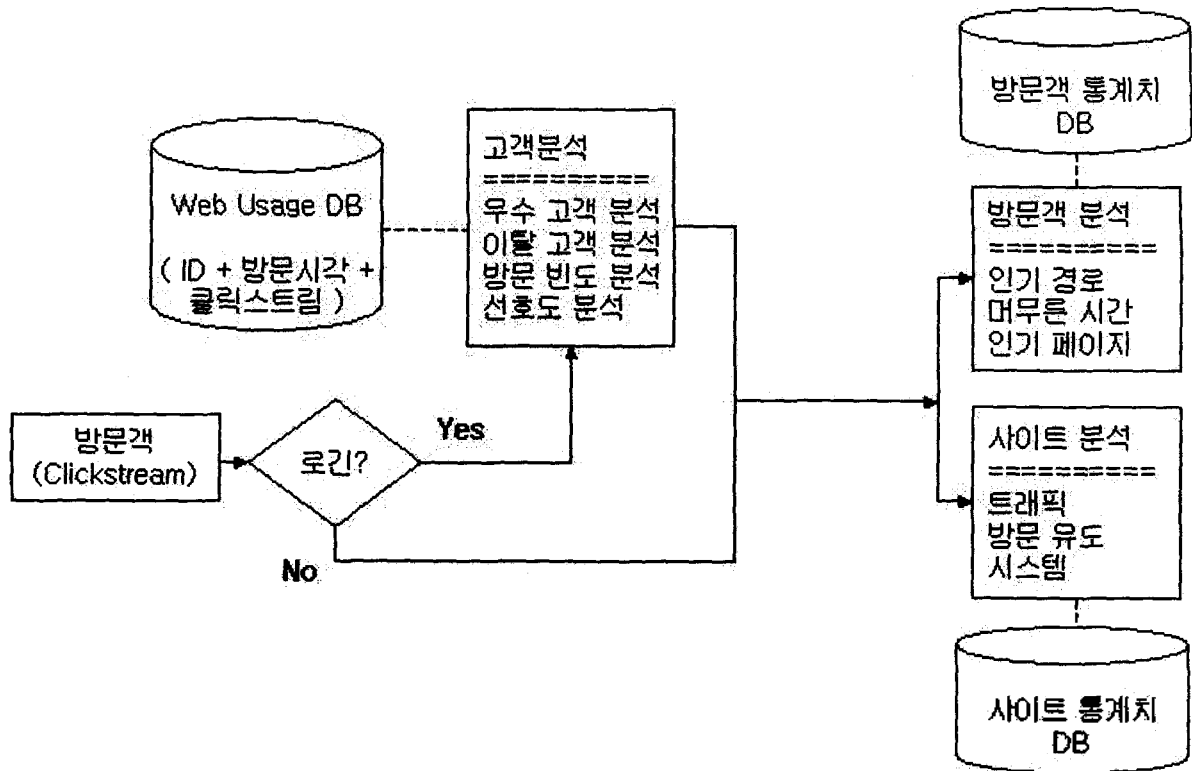
이렇게 ID 중심으로 분석된 웹 행태는 운영계 데이터와 연계되면 실질적으로 업체에 수익이 높은 고객의 행태 패턴을 분석할 수 있는 기반이 마련됩니다.

[그림 6]은 이러한 로그 분석 프로세스를 흐름도로 보여 주고 있습니다.

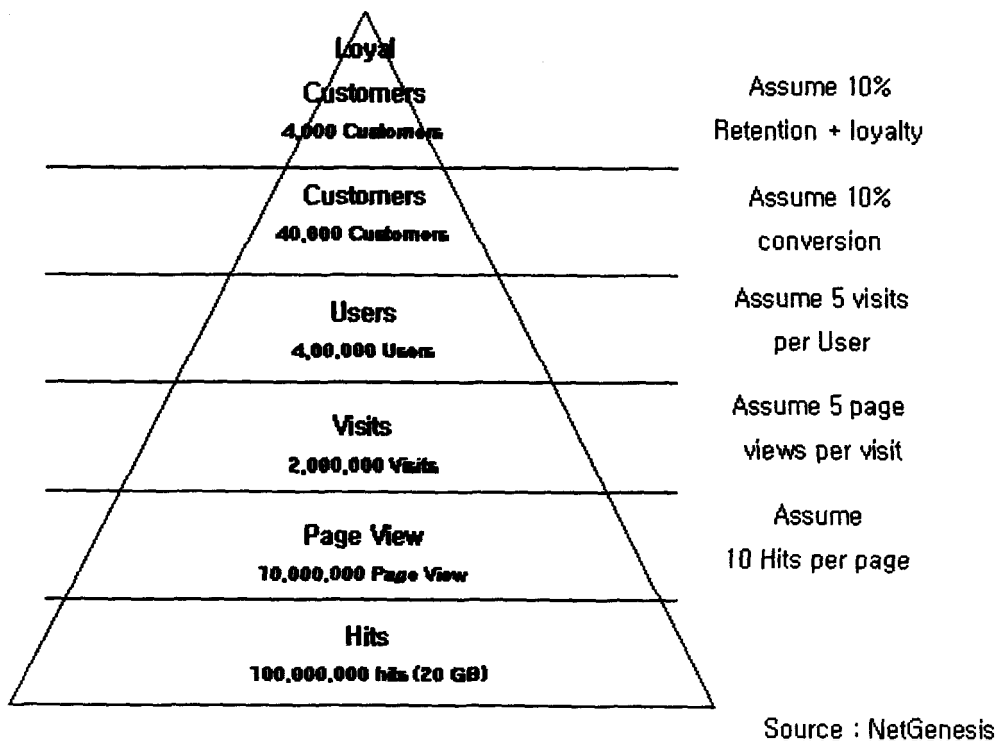
방문객과 사이트 분석을 1차적으로 한 후 오긴 단계를 거친 회원에 대해서는 ID Matching 형태로 로그 데이터를 따로 관리합니다. 그리고 DBMS로의 적재까지 지원합니다.

[그림 7]은 이렇게 로그 데이터를 log in 후의 고객들의 웹 정보만을 관리할 경우, 로그 데이터와 관련한 스토리지 문제를 ROI 관점에서도 큰 무리 없이 해결할 수 있음을 보여 주고 있습니다.

특히 Commerce 내역과 연계할 경우 로그 데이터를 보다 전략적으로 활용할 수 있습니다. 피라미드 상단의 4000여명의 회원의 웹 행태를 집중적으로 관리함으로써 전략적인 1:1 Relation Marketing을 수행할 수 있습니다. 그러나 기본적인 데이터 관리 전략 없이 무분별하게 불필요한 로그 데이터를 모두 관리하면 시스템 및 리소스만 낭비하는 결과를 가져 올 수 있습니다.



[그림 6 : 로그 분석 과정]



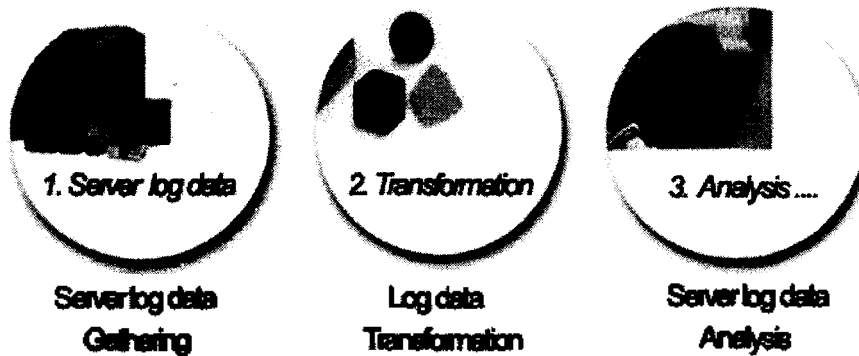
[그림 7 : 로그 데이터 피라미드]

11. 로그분석과 e-CRM

(1) 개념

로그 파일은 사이트를 방문한 사람들의 방문 데이터를 서버가 인식하여 누가 언제 무엇을 요청했고 가져갔는지, 웹 서버에 얼마나 많은 사람이 왔는지, 어디에서 왔으며, 무엇을 제일 좋아하는지, 무엇을 제일 싫어하는지, 가장 오래 보는 또는 가장 많이 보는 페이지는 무엇인지 등등의 방문자 기록을 정리한 하나의 데이터입니다.

이러한 로그파일을 하나하나 분석하여 현 시장 환경 및 고객들의 반응을 예측하며, 인터넷 비즈니스에 전략적으로 활용할 수 있게 하는 것이 바로 e-CRM의 시작입니다.



(2) 고객 DB 구축

로그분석을 수행하면 홈페이지 방문자의 사용현황, 접속현황, 방문현황, 에러현황, 데이터량 현황 등과 같은 일반현황에 대하여 월별, 요일별, 일별, 시간대별 분석 정보를 제공할 수 있습니다. 그러나 이러한 일반적인 정보들만으로는 CRM을 수행하기에는 불충분합니다. 로그분석을 통해서 CRM 관점에서 관심을 가지는 부분은 '도대체 누가 우리 사이트에 오느냐' 라는 근본적인 질문에 답하는 것이며, 이것과 관련된 기반 기술이 ID Matching 기법입니다. 이것을 사용하면 개개인의 행위 분석에 초점을 맞출 수 있고 모든 Web Log 데이터도 개인 중심으로 추출되고 적재될 수 있습니다.

사용자 확인이 된 방문자의 방문 횟수, 사이트 상주 시간, 특별히 자주 찾은 사이트 페이지, 주요 방문 시간대 등의 데이터를 관리하게 됩니다. 예를 들어 ID가 apple인 사람이 우리 사이트에 26번 방문하였고 최근 사이트를 부쩍 자주 찾는다는 사실을 알고 있으며 주로 저녁 시간대에 방문한다는 사실을 알 수 있는 것입니다.

이러한 개별 방문자들의 웹 방문 상황을 기록 & 관리하게 되면 우리 사이트를 특별히 자주 찾고 관심을 가지고 있는 상위 고객 5%의 목록을 추출할 수 있게 되는 것입니다. 즉 우리 사이트의 가치를 높여 줄 수 있는 중요 고객이라는 뜻입니다.

중요 고객 목록은 분석 기간 전체를 통해서 추출할 수도 있고, 최근 기간에 방문한 고객들만 분석에서 추출할 수도 있습니다.

이렇게 추출된 고객들에 대해서는 자동 E Mail 발송 프로그램과 같은 Marketing Application 등을 활용하여 별도로 캠페인 관리 등을 할 수도 있습니다. 이외에 아래 박스에 소개된 방법 등을 활용하여 Web 상에서 VIP 서비스를 구현해 볼 수도 있습니다.

ID를 중심으로 방문자 정보를 추출 및 적재할 수 있으면 사이트에 빈번하게 오는 방문자 상위 1% 리스트를 구할 수 있으며 고객이라는 key로부터 고객 Master DB 및 Transaction DB와 쉽게 연동이 가능하기 때문에 거래 내역 기준에 따라 고객을 분류할 수도 있습니다.

그렇다면 로그 분석을 통해 추출해낸 상위 우수 고객들에게 어떠한 응대를 할 것이냐는 문제가 나오게 됩니다.

이러한 고객들에게 응대할 수 있는 방법으로 다음과 같은 것이 있습니다.

이용 요금 최소화 : 예를 들어 만약 당신의 사이트가 통상 페이지뷰 당 5센트를 부과한다고 하면, 당신은 적어도 100 페이지뷰를 기록하는 고객에게는 4센트라는 할인율을 적용하고 1,000 페이지뷰를 기록하는 고객에게는 3센트를 적용하는 등 더 많은 페이지뷰를 기록하는 고객에게 더 낮은 요금을 물게 하여 100만 페이지뷰 이상을 기록하는 고객에게는 0.01센트까지도 낮추어 줄 수 있습니다.

더 나은 서버 제공 : 웹 상의 서버는 서버에 과부하가 걸려있을 때라도 고객이 기다릴 만한 시간 내에 답을 줄 수 있도록 가치 있는 고객으로부터 오는 요청을 처리하는 데 우선권을 부여할 수 있습니다. 웹 상에서는 종종 아무런 사전 징후도 없이 같은 사이트에 "수많은 고객이 순식간에 들이닥치는"(예를 들면 갓 나온 소프트웨어를 다운로드하려) 경우를 볼 수 있습니다. 이때 대부분의 접속 권한을 단골 고객들에게만 제공할 수 있도록 합니다.

문제는 방법인데 단지 마일리지나 캐쉬백등의 금전적 보상 방식은 고객으로 하여금 그만큼의 돈을 기계적으로 더 쓰게 하는 방법에 불과하기 때문에 저는 개인적으로 좋아하지 않는 방식입니다.

대신 Web 상의 Loyalty에 대해서는 이 역시 Web 상에서 응대를 하는 방식이 바람직하다고 봅니다.

첫째, 단순히 마일리지 등의 방식의 보상은 Service Provider 입장에서는 고스란히 빚더미로 오게 되지만 Vip Service on the Web(이하 VSW)는 초기 투자 비용은 다소 높을지라도 추가 및 유지 비용은 대단히 낮습니다.

둘째, 금전적 보상은 일회적, 단발적 보상으로 그치기 때문에 진정한 Relation management라 보기 힘들지만 VSW은 지속적인 고객 관리가 가능합니다.

그렇다면 VSW는 어떤 형태로 구현을 할 것인가?

개인화도 한가지 방법이라고 생각합니다. 단 개인의 흔적 정보를 다양한 채널을 통해 전사적으로 수집해야 한다는 것이 저의 생각입니다.

특별한 접속 권한 부여 : 예를 들어 가치 있는 고객들은 일반에게 공개되기 전에 인기 있는 컬럼을 읽을 수 있게 할 수 있습니다. 또는 음악 사이트에서 특별판 사운드 트랙을 다운로드 할 수 있게 합니다. 또 많은 접속 수를 기록하는 고객에게만 공개되며 초심자들이 올리는 글로 꼭 차는 일이 없는 토론 그룹을 만들어줄 수도 있습니다. 직접 마케팅에는 두려움, 욕심, 그리고 갈망이라는 세 가지 판매 요소가 있다는 것을 기억한다면, 특별한 권한을 부여한다는 것은 고객의 충성도를 유지하는 데에 매우 효과적인 방법입니다.

(3) 기존 로그분석과의 차이점

기존의 웹 로그 분석 툴이 제공하는 리포트는 주로 시스템 관점에 의거한 것입니다. 기본적인 Traffic 정보 및 링크 분석, 에이전트 분석 등에 초점을 둡니다.

즉 방문자가 급격하게 증가하고 있는 추세라고 감지되는 경우 서버 확장이라는 시스템 기반의 의사 결정을 내리는데 도움이 됩니다. 또 깨진 링크나 없는 페이지 분석 등을 통해서 사이트 페이지 관리를 할 수도 있습니다. 그 외에 방문자들이 주로 사용하는 브라우저 종류 및 버전 등을 파악하여 이에 최적화된 서버 환경을 구축하는 경우도 해당됩니다.

그러나 결국 이러한 데이터 값들은 Parameter 값으로서 '사이트의 방문 현황'이라는 통계치의 특성 값은 될 수 있으나 실질적으로 비즈니스 기반의 의사 결정에는 도움이 되지 않는 경우가 많습니다.

단적인 예로 우리 사이트에 가장 중요한 고객은 누구냐고 묻는 질문에 답변을 할 수 없습니다.

또 이러한 중요 고객들이 주로 관심을 가지고 보는 페이지는 무엇이며, 이탈 고객은 누구인가에 대한 질문 역시 비즈니스 관점에서는 중요함에도 불구하고 여전히 기존의 로그 툴에서는 할 수가 없는 것입니다.

그러나 ID matching 기술에 기반하여 로그 데이터를 추출 적재하는 경우는 조금 다릅니다.

먼저 누구인가(Who is he/she)에 대한 문제부터 시작합니다. 방문자가 Log in 을 거쳐 사용자 확인이 된 이후부터 방문자가 Web상에 남기는 행위 흔적들을 추적하는 측면에 관심을 둡니다.

로그 데이터는 ID를 중심으로 적재되고 관리되므로, 개별 방문자의 Web Usage 현황을 쉽게 파악할 수 있습니다. 또한 ID를 근거로 Web Usage DB가 기존의 Legacy 환경과 쉽게 연동이 될 수 있습니다. 즉 종합적인 Web enabled DW가 구축의 기반 요소가 될 수 있기 때문에 Log 데이터를 넘어서는 각종 비즈니스적 관점의 질문에 대해서도 답변을 할 수 있게 됩니다. 또한 정제 과정을 거쳐 데이터의 충실도를 확보하게 되면 실질적인 Web data mining도 가능하게 됩니다.

고객과 실질적인 Relationship을 형성할 수 있게 되는 첫걸음 바로 e-CRM의 시작은 ID Matching 기법을 사용한 로그분석으로부터입니다.

(4) e-CRM을 위한 분석항목

ID Matching 기반으로 Log data가 적재된 Web DB에 근거하여 e-CRM 관점에서 분석해야 할 항목들을 고려하여 보았습니다. 이러한 비즈니스 기반의 의사 결정 지원 수준도 Web를 통한 사용자의 활동 내역과 로그분석에 따라 바로 가능한 것들도 있고 기존 Legacy DB와 연동하여 가능한 것도 있습니다.

구분	질문 예시
<p>ID Matching 기반의 로그분석 tool이 분석할 수 있는 비즈니스 관점의 분석항목</p>	<ol style="list-style-type: none"> 1. 가장 즐겨 찾는 콘텐츠 카테고리는 무엇인가? 또한 그렇게 즐겨 찾는 콘텐츠 카테고리별로 사용자 그룹은 어떻게 되는가? 2. 콘텐츠에 대한 방문객의 선호를 통해 방문객의 소비 성향은 파악할 수 있는가? 방문객이 무엇을 특히 원하고 있는가? 3. 메인 페이지에서 서브 페이지로 넘어가는 과정에서 트래픽의 분포 현황은 어떠한가? 4. 캠페인 기간 동안의 방문 현황과 일반적인 기간 동안의 방문 현황을 비교할 수 있는가? 5. 방문횟수, 방문빈도, 최근 방문 상황을 통합한 사이트에 대한 고객의 Stickiness 는 파악할 수 있는가? 6. 방문객이 주로 방문하는 시간대는 어떠한가? 7. Agent(가격 비교)를 통해 들어 왔는가? 그냥 다른 경로로 들어 왔는가? 8. 메인 페이지의 콘텐츠 분량은 적절한가? 즉 향해 경로가 좁되 깊은가? 그냥 넓으면서도 얇기만 하는가? 9. 고객의 방문경로 추적을 통해 향후 이 고객이 재방문 할 가능성이 높다는 것을 추정해 볼 수는 없는가? 10. 고객의 방문이 주기적인가? 아니면 어쩌다 한번 오는 고객인가? 11. 메일링 리스트를 통한 고객의 방문률은 어떠한가? 잠재 및 이탈 고객에 대한 메일링 리스트의 효과는 어떠한가? 12. 방문객의 파일 다운로드 현황은 어떠한가? 13. 비개인화된 상태에서 추천된 콘텐츠를 메인 화면에 배치하는 경우 실질적으로 얼마나 트래픽을 보다 유도하는가?
<p>Legacy System과 연동된 상황에서 분석할 수 있는 항목들</p>	<ol style="list-style-type: none"> 1. 고객의 상품 구매 이력 및 문의 및 조회 이력을 관리할 수 있는가? 2. 배너 광고에 대한 고객의 반응을 어떠한가? 인구 통계학 및 소비 성향, 구매 이력에 따라 고객의 광고 반응 현황을 정리해 볼 수 있는가? 3. 고객들의 요구 사항을 체계적으로 정리할 수 있는 방법은 없는가? (게시판의 콘텐츠 마이닝, E Mail 인바운딩 문의 정리) 4. 고객 등록 정보를 바탕으로 고객의 소득 수준 및 소비 성향을 추측해 볼 수는 없는가? 5. 고객을 소비 성향에 따라 우량 고객, 보통 고객, 불량 고객 등으로 분류해 볼 수는 없는가? 6. 고객이 특정한 물건을 사고 싶다는 시점을 파악하여 메일로 적절하게 상품 안내를 할 방안은 없는가? (가령 가족의 생일 등을 파악, 소망 상자등) 7. 마일리지 관리를 고객이 하고 있는가? 8. 아직 물건을 사지 않은 고객의 행위 분석을 통해서 고객이 물건을 향후 살 가 능성이 있는지 없는지 알 수는 없는가? 9. 다른 사이트의 배너 광고를 통해서 오는 고객들의 실제 구매 현황은 어떠한 가? (광고 ROI)

12. 결론

작은 구멍가게, 슈퍼마켓으로부터 대기업까지 CRM 통한 고객만족이라는 필요조건을 충족시키기 위해서 끊임없는 노력을 기울여 왔으며, 이제 Web이라는 온라인이 고객 접점의 필수적인 채널이 됨으로써 e-CRM 또한 간과할 수 없는 중요한 영역이 되었습니다. 하지만 e-CRM을 잘 할 수 있는 IT 기반을 갖추는 것은 많은 요구사항을 충족시켜야 하며 투자비용 또한 적지 않습니다. 실질적인 예산이 있더라도 그것을 어떻게 효과적으로 사용하여 e-CRM 환경을 구축할 수 있는 지에 대한 대답 또한 쉽지 않습니다. 하지만 분명한 것은 Web이라는 채널에서 발생하고 있는 고객관련 유용한 정보들을 얼마만큼 효과적으로 추출하고 분석하느냐에 따라 e-CRM의 실패와 성공이 구분될 수 있다는 것입니다.

e-CRM의 효과적 구축을 위해서는 웹으로부터의 정보 추출, 유용한 고객 정보의 DB화, 적합한 마케팅 액션 및 반응에 관한 피드백 등의 요소들이 필수적입니다. 이들 모든 요소들을 한번에 모두 도입하기에는 ROI라는 관점에서 상당히 조심스러울 수 밖에 없습니다. 단계적으로 구축해나가는 것이 하나의 전략이 될 수 있으며 웹으로부터의 고객 정보 추출이 e-CRM을 위한 첫 단계가 될 것입니다. 즉 ID matching 기반의 Web 로그분석기법을 사용하여 고객 중심의 정보를 잘 수집하고 DB화 하는 것으로 e-CRM을 시작할 수 있을 것입니다.