

## 효율적 웹 개인화를 위한 웹 사용자들의 경향분석

- Trend Analysis of Users for The Effective Web Personalization -

임영문\*

Leem Young Moon

김홍기\*

Kim Hong Gi

### Abstract

온라인 기업의 경우 쉽게 가질 수 있는 고객데이터는 고객이 자사의 홈페이지에 접속하여 남기고 간 흔적(Web Log)이나, 고객이 직접 제공하는 데이터 일 것이다. 현재 많은 온라인 기업이 가장 기본적인 분석으로 웹 로그 분석을 시행하고 있으나, 그 양이 너무 많아 수시로 처리하는 데는 문제가 있기도 하지만, 가장 쉽고, 기본적인 분석임은 피할 수 없는 사실이다. 기존의 웹로그 분석에 관한 연구들이 웹로그 분석을 통하여 사용자패턴 분석에 그친대 비하여 본 논문은 무선 페이지를 위한 콘텐츠 추출 및 기존의 데이터중심의 마케팅 전략에서 벗어나 동양적인 정서를 가미한 마케팅 전략도 함께 제시한 것에 그 의의가 있다.

## 1. 서론

### 1.1. 연구배경 및 필요성

상품과 고객이 만나는 채널은 인터넷의 발전과 더불어 점점 더 다양해지고 있으며, 고객이 상품을 비교할 수 있는 채널 또한 다양해지고 있다. 따라서, 기업은 기존에 활용하였던 고객과의 접근 방법을 사용해서는 다른 기업과의 차별성 및 우월성을 나타낼 수 없게 되었다. 또한, 고객은 더욱더 다양한 지식을 바탕으로 기업과 대하기 때문에 허술하고 눈에 보이는 전략으로는 고객의 만족 및 감동을 이끌어 낼 수 없게 되었다. 온라인 기업의 경우 쉽게 가질 수 있는 고객데이터는 고객이 자사의 홈페이지에 접속하여 남기고 간 흔적(Web Log)이나, 고객이 직접 제공하는 데이터 일 것이다. 현재 많은 온라인 기업이 가장 기본적인 분석으로 웹로그 분석을 시행하고 있으며, 그 양이 너무 많아 수시로 처리하는 데는 문제가 있기도 하지만, 가장 쉽고, 기본적인 분석임은 피할 수 없는 사실이다. 무선인터넷은 핸드폰이나 PDA(Personal Digital Assistants) 등 무선 단말기를 이용해서 정보를 주고받을 수 있는 차세대 인터넷의 형태중 하나이다. 단말기 기술의 발달과 무선통신기술의 발전과 더불어 무선인터넷 또한 기업의 중요한 고객 채널로 부각되고 있다.

\* 강릉대학교 산업공학과

무선인터넷을 이용한 고객관리는 고객에게 더욱더 세분화된 서비스를 제공할 수 있으며, 고급화된 서비스를 제공함으로써 고객의 만족과 구매욕구를 증진시킬 수 있다. 하지만, 개인화된 서비스를 제공하기 위해서는 고객에 대한 패턴 연구 및 세분화 작업이 먼저 이루어져야 한다. 이러한 작업을 위한 다양한 연구중 한 분야가 웹로그를 이용한 사용자의 패턴분석일 것이다. 기존의 웹로그 분석에 관한 연구들은 웹로그분석을 통하여 사용자패턴 분석에 그친데 비하여 본 논문은 무선 페이지를 위한 콘텐츠 추출 및 기존의 데이터중심의 마케팅 전략에서 벗어나 동양적인 정서를 가미한 마케팅 전략을 제시하는데 그 의의가 있다.

## 1.2. 연구의 방법

본 논문에서는 웹로그 데이터를 가지고 연관규칙을 적용하여 웹 사용자들의 패턴을 연구하고, 효율적 웹의 재구성을 제시하며, 무선인터넷 페이지 구축시 적합한 콘텐츠를 선정하여 그 예제 페이지를 구축하고자 한다. 본 논문에서 사용되는 데이터 및 연구방법을 보면, 첫째, 샘플 데이터는 강릉에 소재한 치과병원의 웹 페이지 데이터를 사용한다. 현재 웹 페이지를 통한 홍보 및 병원 홈페이지로서의 그 기능을 제대로 하고 있지 못한 실정이다. 둘째, 샘플 데이터는 데이터의 신뢰성을 고려하여 한달 동안의 웹로그를 사용하여 분석을 실시한다. 셋째, 분석 절차는 웹로그 데이터로부터 필요 없다고 생각되는 데이터들을 제거한다. 1차 정제된 데이터들을 가지고 연관규칙을 적용한다. 데이터베이스는 MS-SQL Server 7.0을 사용한다. 넷째, 통계적 분석 도구로서 Excel, SPSS를 사용한다. 다섯째, 웹마이닝 된 결과를 가지고 웹페이지의 리모델링을 제시한다. 또한, 무선 콘텐츠를 추출하여 예제 무선 페이지를 구축한다. 무선 페이지 구축시 이용하는 도구로는 UP.SDK 4.0이며, WML(Wireless Markup Language)과 ASP를 사용한다. 웹서버는 Windows2000 Server를 사용한다.

## 2. CRM의 정의

과거의 마케팅은 기업이 상품을 만들어 특별한 고객 정보 분석이 없이 상품에 대한 홍보 및 광고가 주류를 이루었다. 하지만 현대에는 이러한 마케팅 활동에 의해 기업의 수익을 향상시키는 데 한계를 느끼면서 이러한 상황에서 고객에 대한 이해를 바탕으로 한 마케팅 활동이 등장하게 되었다. 이러한 고객에 대한 이해를 바탕으로 영업, 마케팅, 고객 서비스 프로세스를 재구성하는 일련의 활동을 고객 관계 관리, 즉 CRM(Customer Relationship Management)이라 정의할 수 있다. CRM은 다양한 채널을 통한 고객과의 커뮤니케이션으로부터 수집된 정보를 기반으로 고객과의 관계를 유지 발전하는 과정이다. 즉 신규 고객 획득, 우수 고객 유치, 고객가치 증진, 잠재 고객 활성화 및 평생 고객 확보를 목표로 고객 분석을 통해 고객을 이해하고 이를 통해 고객과의 지속적인 관계를 유지함으로써 고객 가치를 극대화하기 위한 일련의 과정이라 할 수 있다. 정보 기술의 발달로 개별 고객 단위로 고객의 가치 측정이 가

능해졌으며 어떻게 충성도 높은 고객과 관계를 구축하고 이를 지속적으로 유지 할 것인가에 대한 가장 효율적인 방법을 찾을 수 있는 기법을 알 수 있게 되었다.

이러한 기법을 통하여 기업 내-외부의 데이터를 전체 조직의 관점에서 통합 시스템을 구축하고 이러한 기반 하에서 기업에 필수적인 고객 정보를 데이터 마이닝(Data Mining)을 통하여 분석을 하여 기업의 신속한 의사결정을 할 수 있도록 정보를 제공하는 역할을 한다. CRM은 궁극적으로 고객 로열티 향상을 통한 고객의 평생 가치(LTV : Life Time Value) 증대를 목적으로 한다[2].

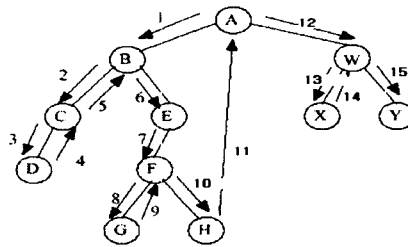
### 3. 무선 인터넷과 CRM

대고객 채널은 CRM을 구성하고 있는 핵심 요소 가운데 하나이다. 고객으로부터 정보를 얻어 새로운 수익 기회를 포착해 내거나 새로운 정보를 고객에게 알려 수익성 있는 행위를 유도하는 것은 CRM의 가장 중요한 수단중의 하나이다. 고객 채널은 고객을 직접 만나는 대면 미팅을 비롯하여 서신, 전화, 이메일, 인터넷, 통신, 무선전화 등 전산 및 통신 기반 시설이 발달함에 따라 그 영역이 급속도로 확대되고 있다. 특히 무선 데이터 통신 수단을 이용하는 고객 채널은 새롭게 관심 받고 있는 영역으로서 이미 폭넓은 적용분야를 갖고 있다[3]. 무선 인터넷 기술을 CRM에 적용시키는 방법은 크게 두 가지로 구분할 수 있다. 첫째, 고객 흔적 분석 중심의 무선 인터넷 CRM이다. 이 방법은 eCRM 영역에서 주로 사용하고 있는 방식으로 무선 인터넷을 기존의 유선 인터넷과 동일한 개념으로 이해, 무선데이터 망을 타고 들어온 고객이 남기는 웹로그와 페이지 방문 이력 등을 이용한다. 즉 무선 인터넷을 또 하나의 인터넷 고객 채널로 이용하고 사이트에 접속하여 고객이 남기는 흔적을 고객 분석의 기초 자료로 삼는 접근 방법이다. 둘째, 고객 채널 활용 중심의 무선인터넷 CRM이다. 이 방법은 고객과 연결하고 고객을 응대하기 위한 수단으로서 무선 데이터 통신 기능을 적용한다. 즉, 고객에게 서비스를 하기 위해 출장 서비스에 PDA를 사용하여 서비스 센터와 데이터를 송수신하거나 고객이 이동전화를 활용하여 직접 원하는 사이트에 접속하여 정보를 얻거나 서비스를 의뢰하는 경우가 이에 해당한다. 앞의 두 가지 형태로 크게 대별되는 무선 인터넷의 CRM으로의 활용은 최근 들어 적용 사례가 늘어나고 있는 추세이며, 이동통신 망의 대역폭이 증가함에 따라 서비스의 다양성과 유용성 역시 늘어날 것이다.

### 4. 순회 패턴

서류(documents)와 객체들(objects)이 상호 접근을 허용하도록 서로 연결되어 있는 환경에서 제공하는 분산된 정보에 접근하는 패턴을 탐사하는 새로운 데이터마이닝의 기법이 연구되고 있다. 그런 정보를 제공하는 환경으로는 world wide web 과 on-line service로 CompuServe, America Online, Hitel 등이 있다. 사용자는 관심 있는

정보를 찾을 때 제공되는 하이퍼링크와 같은 것을 통해서 한 객체에서 다른 객체로 이동한다. 이런 환경 하에서 사용자의 접근 패턴(access patterns)을 이해하면 시스템 설계를 개선시킬 수도 있고 더 좋은 마케팅 결정으로 이끌어 낼 수도 있다. 시스템 설계면에서는 크게 연관된 객체들 사이에는 효율적 접근을 제공하고 그 페이지에 더 좋은 제작 설계를 제공할 수 있다. 마케팅 결정면에서는 적절한 위치에 광고를 함으로써 더 좋은 소비자/사용자의 분류와 행위 특성 분석을 제공할 수도 있다. 그런 환경 하에서 사용자의 접근 패턴을 포착하는 것을 순회 패턴(traversal patterns)탐사 [1]라 일컫는다.



[그림 1] 접근 패턴의 예

순회 패턴을 찾는 과정은 다음과 같다. 사용자의 로그 파일에서 한 사용자의 접근 패턴의 예로서 [그림 1] 에서와 같이 먼저 순회 노드를 정리하면 {A, B, C, D, C, B, E, F, G, F, H, A, W, X, W, Y}로 이루어진 것을 알 수 있다. 이 접근 패턴에서 순방향 참조 집합은 {ABCD, ABEFG, ABEFH, AWX, AWY}로 된다. 각 사용자에 대한 순방향 참조 집합을 구하고, 모든 사용자들에 대한 순방향 참조 집합에서 순차 패턴에서와 같이 최소 지지도를 만족하는 빈발 참조 시퀀스를 발견한다. 마지막으로 빈발 참조 시퀀스에서 최대 참조 시퀀스[1]를 찾아내는 것이 순회 패턴 탐사를 하는 것이다.

## 5. 웹 로그 화일에 대한 패턴 탐사 및 분석

### 5.1. 웹 로그 화일의 특징

웹 액세스 로그 파일 (Web access log file)은 인터넷을 사용하는 사용자가 개설된 웹의 홈페이지를 액세스했을 때부터 찾아보는 모든 페이지의 이름과 시간을 기록하는 파일이다. 웹 서버는 모든 개별적인 액세스에 관한 URL과 시간 정보를 기록으로 남긴다[4]. 본 논문에서 사용된 데이터는 강릉대학교 치과병원의 웹 액세스 로그 데이터이고, 서버의 OS(Operating System)는 Linux이며, 2001년 8월 12일부터 2001년 9월 13일까지 축적된 것이다. 원형의 데이터 형식은 [표 1]과 같다. 전체 트랜잭션의 수는 365,933개 였다. 사용자의 도메인 이름 (또는 IP 주소), 사용한 시간 (일 /월 /년 /시 :분 :초), 사용한 메소드(GET 또는 POST), 그리고 요청한 화일의 이름(URL 주

소), HTTP(Hyper Text Transfer Protocol)버전, 사용 상태 코드(성공 또는 에러 코드), 전송된 바이트 수 등의 순서로 기록된다. 즉 [표 1]의 첫 번째 레코드는 210.95.147.93 이라는 IP 주소를 가진 도메인에서 2001년 9월 9일 07시 16분 00초에 dental.kangnun.ac.kr/title.html 이라는 문서를 요청했으며 그 때의 HTTP버전은 1.1 이고 성공적으로 수행되어 8920 bytes 가 전송된 것이다 .

[표 1] 전처리 전의 원시 웹 액세스 로그데이터

210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/title.html	HTTP/1.1	200	8920
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/main.php3	HTTP/1.1	200	15525
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/images/intro/top1.jpg	HTTP/1.1	200	7785
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/images/intro/top2.jpg	HTTP/1.1	200	16889
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/images/intro/top3.jpg	HTTP/1.1	200	9869
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/images/bchju.gif	HTTP/1.1	200	33994
210.95.147.93	[09/Sep/2001:07:16:00 +0900]	*GET	/images/bbojon.gif	HTTP/1.1	200	17241

웹 서버에서 웹 액세스 로그 파일을 저장 관리하기 위해서는 상당히 큰 어려움이 존재한다. 그 이유는 홈페이지를 통해서 액세스하는 모든 IP 주소, 시간, 탐색하는 페이지 등의 기록을 유지하기 위해 필요로 되는 메모리의 대량 사용 때문이다. 그러므로, 대부분의 웹 서버에서 로그 파일의 관리를 위해 일정한 시점이 지나거나 일정한 크기 이상이 되면 그 파일을 삭제하거나, 아주 기초적인 통계 처리를 통하여 기본적인 정보만 추출하고 삭제하는 경우가 보통이다.

## 5.2. 웹로그 파일의 전처리

### 5.2.1. 불필요한 이미지파일의 삭제

사용자가 웹페이지를 액세스 했을 때 [표 1]에서 볼 수 있듯이 사용자 IP 210.95.147.93은 top1.jpg, top2.jpg등을 요구하는데, 이러한 이미지들은 패턴 탐사에 부적절한 결과를 가져올 수 있으므로 1차 정제과정에서 제거되었다. 또한, 분석에 사용된 로그 데이터에는 바이러스로 인한 로그가 남아 있었는데 그것 또한 삭제되었다. [표 2]는 이러한 결과를 보여준다. 처리 후 트랜잭션의 수는 62,971개였다.

[표 2] 이미지가 정제된 로그데이터

IP	Date	Trace	Type	Size
209.247.40.108	09/Sep/2001:04:05:48	/intro/layout_f4.html	200	5844
209.247.40.105	09/Sep/2001:04:08:36	/intro/layout_f5.html	200	5846
209.247.40.108	09/Sep/2001:04:19:01	/intro/layout_f6.html	200	5846
211.90.77.210	09/Sep/2001:05:48:52	HTTP/1.0		
210.95.147.93	09/Sep/2001:07:16:00	/	200	595
210.95.147.93	09/Sep/2001:07:16:00	/title.html	200	8920
210.95.147.93	09/Sep/2001:07:16:00	/main.php3	200	15525
210.95.147.93	09/Sep/2001:07:16:20	/intro/layout0.html	200	6376

### 5.2.2. URL주소에 대한 매핑 및 동일한 프레임 인식

사용자가 접근한 URL에 대해 부호화를 시키고, 동일한 프레임에서 로드된 파일들은 한번의 요구로 인식시켰다. 동일한 프레임에서 로드된 파일들은 [표 1]에서 보면,

[표 3] 동일한 프레임 제거

210.100.153.58	24/Aug/2001:15:07:46	A
210.100.153.58	24/Aug/2001:15:07:56	D
210.100.153.58	24/Aug/2001:15:07:57	B
210.100.153.58	24/Aug/2001:15:08:04	E
210.100.153.58	24/Aug/2001:15:08:12	C
210.100.153.58	24/Aug/2001:15:08:27	A

한 사용자(210.95.147.93)가 동일한 시간에 /, /title.html, /main.php3 파일을 액세스 했는데 이 3가지 페이지들은 메인페이지 파일들로서 동일한 프레임에서 로드된 파일들로 사용자가 3번의 파일을 클릭 했다고 볼 수 없다. [표 3]은 동일한 프레임에서 로드된 파일들이 제거된 모습을 보여주고 있다. [표 4]는 동일 프레임인 파일들을 정제하지 않았을 경우에 나오는 결과를 나타내고 있는데, 3개로 분리된 메인 페이지가 서로 다른 페이지의 요청으로 인식되어 결과가 나오는 것을 볼 수 있다.

[표 4] 동일 프레임으로 인식하지 않은 결과

1. http://dental.kangnung.ackr/ 2. http://dental.kangnung.ackr/ title.html	15.21%
1. http://dental.kangnung.ackr/ title.html 2. http://dental.kangnung.ackr/	5.61%
1. http://dental.kangnung.ackr/ 2. http://dental.kangnung.ackr/ title.html 3. http://dental.kangnung.ackr/ main.php3	4.12%

[표 5]와 [표 6]은 URL에 대해 부호화 한 것을 보여주고 있다. [표 3]은 이러한 파일들을 정제한 것을 보여준다. 정제후 전체 트랜잭션의 수는 61,847개였다.

[표 5] 페이지 부호화

부호화 전	부호화 후
/title.html, /main.php3%, /	A
/cyber/board.php%	B
/board/board.php%	C

[표 6] URL의 부호화

209.247.40.105	09/Sep/2001:03:52:33	z2	200	5844
209.247.40.108	09/Sep/2001:04:05:48	z2	200	5844
209.247.40.105	09/Sep/2001:04:08:36	z2	200	5846
209.247.40.108	09/Sep/2001:04:19:01	z2	200	5846
210.95.147.93	09/Sep/2001:07:16:00	A	200	595
210.95.147.93	09/Sep/2001:07:16:00	A	200	8920
210.95.147.93	09/Sep/2001:07:16:00	A	200	15525
210.95.147.93	09/Sep/2001:07:16:20	z2	200	6376
210.95.147.93	09/Sep/2001:07:16:23	Q	200	12684

[표 5]에서 보면 /title.html, /main.php3, / 은 A로 부호화되었는데, 이는 위에서도 언급했듯이 한 페이지가 3개의 프레임으로 나뉘어져 있기 때문이다. 게시판형식의 페이지들은 읽기, 쓰기, 수정 등의 항목들이 있는데 본 논문에서는 이러한 것들도 하나

로 인식하였다. /board/board.php% 와 같은 것들이 게시판형식의 페이지이다.

### 5.2.3. 수평 데이터베이스

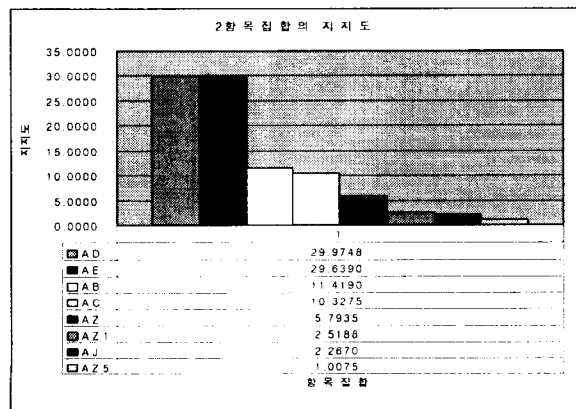
[표 7]은 최대순방향참조를 구하기 위해 수직으로 되어 있던 데이터베이스에서 트랜잭션아이디를 주키(Primary Key)로 삼고 시간 순으로 정렬을 하였다. 트랜잭션의 수는 1192개였다.

[표 7]수평 데이터베이스

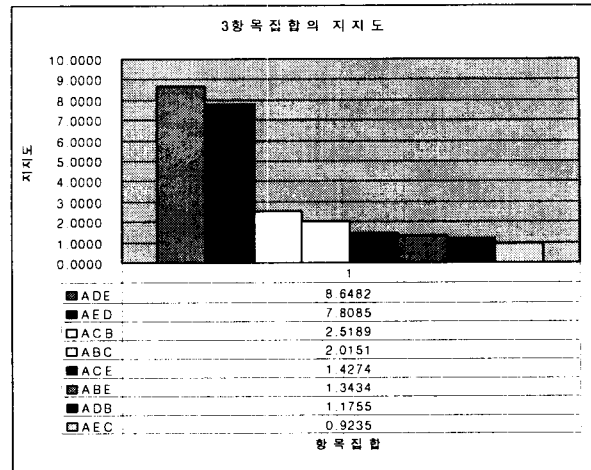
TID	UID	Trace
2	128.134.0.124	AD
4	128.134.230.130	AE
7	128.134.253.132	AD
1366	211.114.226.137	A
2561	211.114.226.184	AEDB
5302	211.114.226.65	AE
5331	211.114.226.65	ACB
5362	211.114.226.67	AJ

### 5.2.4. 실험결과

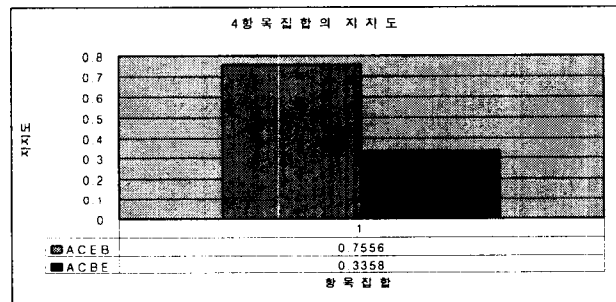
[표 7]에서 구한 수평 데이터베이스를 가지고 연관규칙을 적용 시켜서 순회패턴 탐사를 실시한 결과(지지도 1%이상) [그림 2], [그림 3], [그림 4]와 같은 결과를 얻을 수 있었다. [그림 2]에서 보면 AD, AE, AB 순으로 지지도가 높았다. 즉, 메인페이지에서 치학정보로 가는 지지도가 29.97%, 메인페이지에서 자료실로 가는 지지도가 29.63%, 메인페이지에서 사이버진료실로 가는 지지도가 11.41%로 나타났다. 3항목 집합을 구한 결과는 [그림 3]과 같았다. ADE(지지도-8.64%), AED(지지도-7.80%), ACB(지지도-2.51%), ABC(지지도-2.01%) 순으로 나타났다. 즉, 메인화면-->치학정보-->자료실, 메인화면-->자료실-->치학정보, 메인화면-->게시판-->사이버진료실, 메인화면-->사이버진료실-->게시판 의 순으로 지지도가 높게 나타났다.



[그림 2] 2항목 집합의 지지도



[그림 3] 3항목 집합의 지지도



[그림 4] 4항목 집합의 지지도

[그림 4]에서 볼 수 있듯이 최소지지도 1%를 만족하는 4항목 집합은 존재하지는 않았지만, 최상위를 나타내는 항목집합은 ACED, ACBE로 나타났다.

### 5.3. WebTrends를 통한 분석

위에서 언급된 분석 내용들은 SQL 언어를 이용하여 순회패턴 알고리즘을 적용하여 분석한 것들이다. 이 절의 분석은 웹로그 분석 툴인 WebTrends를 사용하여 분석한 내용들이다. WebTrends를 통한 분석을 실시한 목적은 CRM을 적용함에 있어서 보다 다양한 정보를 얻기 위해서이다.

#### 5.3.1. Link 분석

Link 분석은 웹사이트의 상호 유기적인 연결 관계를 알아보는 분석으로 링크가 잘못 되어있거나 사용하지 않는 웹 페이지 같은 것을 알아보는 분석으로 전체적인 사이트



의 연결성 및 현재의 사이트 상태를 알아보는데 효과적이다. WebTrends를 사용하여 Link 분석을 실시한 결과 18개의 HTML페이지와 기타 동적인 페이지들이 있었으며, 손상된 페이지나 사이트 링크(총 396개의 링크)가 손상된 것은 하나도 존재하지 않았다. 하지만, 홈페이지 제작당시 잘못된 태그가 사용된 것으로 나타나는 페이지가 몇 개 보이거나 큰 문제를 일으킬만한 것은 아닌 것으로 판단된다. 하지만, 웹 페이지 중에서 이미지의 용량이 큰 것을 사용하여 전송 속도가 늦어지는 경우는 있었다. [표 8]은 그 예를 보여주고 있다.

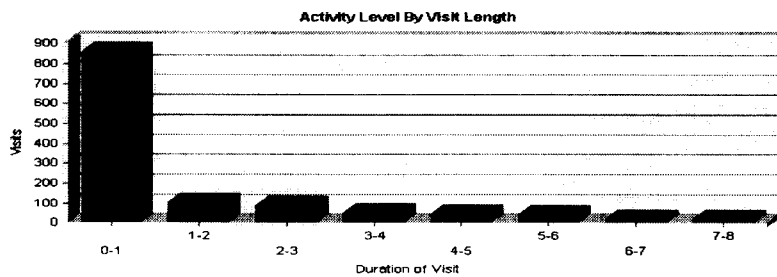
[표 8] 전송이 느린 페이지

Biggest (Slowest) Pages				
Title & URL of each Page	Size of Page & Its Elements	Download time		
		Modem (28.8)	ISDN (128Kbit)	T1 (1.5Mbit)
☒ 병원소식	256K (이미지 용량: 2K)	85.33s	16.00s	1.30s
☒ 병원 배치도	96K (이미지 용량: 6K)	32.00s	6.00s	0.49s
☒ 홈 타이틀	87K (이미지 용량: 8K)	29.00s	5.44s	0.44s
☒ 교통 안내	50K (이미지 용량: 16K)	16.67s	3.13s	0.25s
☒ 전화번호 안내	44K (이미지 용량: 44K)	14.67s	2.75s	0.22s
☒ 병원 현황	40K (이미지 용량: 6K)	13.33s	2.60s	0.20s

링크 분석을 실시한 결과 전체적으로 큰 문제는 없는 것으로 사료된다. 그러나, 웹 제작시 이미지의 크기를 너무 크게 만들어서 제작한 부분은 수정이 요구된다.

### 5.3.2. 사용자들의 체류시간 분석

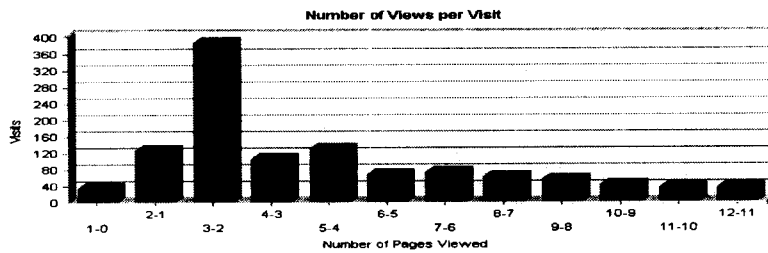
사용자들의 체류시간을 분석하는 목적은 사용자들이 치과병원 웹사이트에 접속하여 정보를 획득하고 나간 시간을 알아봄으로써, 사용자가 실제로 정보를 얻고자 했는지 아니면, 그냥 한번 웹사이트를 둘러보고 나갔는지를 알아볼 수 있는 중요한 정보를 제공해 주기 때문이다.



[그림 5] 사용자들의 체류시간 분석

치과병원 웹사이트의 1주일간 웹로그 데이터를 이용하여 접속자들의 체류시간을 알아본 결과는 [그림 5]와 같이 나타났다. [그림 5]에서 볼 수 있듯이 1분 미만으로 접속한 사용자들이 대부분이며, 3-4분 이상 장시간 접속한 사용자들은 거의 없는 것으로

로 나타났다. 따라서, 대부분의 사용자들은 접속한 후 자신이 작성한 글에 대한 답변만 보고 사이트를 빠져나갔거나, 그냥 한번 접속하여 여기 저기 둘러보고 나간 사용자들이 대부분임을 알 수 있다. 전자에서 언급한 내용이 맞는지 후자가 맞는 추측인지는 사용자가 평균적으로 열어본 페이지를 분석해보면 알 수 있는데, [그림 6]은 그 결과를 보여주고 있다.



[그림 6] 사용자가 열어본 페이지 수

[그림 6]에서 보면 3 또는 2페이지를 열어본 사용자들이 가장 많은 것으로 나타났는데, 이 부분은 WebTrends분석이 잘못 나온 것으로 보인다. 왜냐하면, WebTrends는 프레임이 나눠져 있는 페이지를 하나의 페이지로 인식하지 않고 각각 다른 페이지로 보고 있기 때문이다. 따라서, 5.2.4절에서 살펴본 순회패턴 실험 결과 값과 비교분석을 해 본다면, 실제로는 1 또는 2페이지를 열어본 사용자가 대부분일 것으로 사료되며, 4페이지 이상 열어본 사용자는 거의 없는 것으로 판단된다. 하지만, 본 연구에서도 게시판과 같은 항목에서는 읽기, 쓰기 등과 같은 페이지들을 모두 하나의 페이지로 인식하고 실험을 하였기 때문에 메인 페이지를 접속하고 게시판이나 사이버진료실, 치학정보를 접속한 사용자들이 대부분일 것으로 추측된다. 따라서, 대부분의 사용자들은 자신이 작성한 게시물에 대한 답변을 보고 나가거나, 자신이 원하는 정보만을 취득하고 나간 것으로 판단된다.

## 6. 웹로그 파일의 분석을 통한 CRM접근

### 6.1. 무선인터넷 콘텐츠 추출

5장의 분석결과를 보면 대부분의 사용자들은 2개에서 3개의 페이지를 요청했으며, 주로 치학정보, 사이버진료실, 자료실을 클릭 했다는 것을 알 수 있다. 실제 홈페이지를 살펴보면 사이버진료실에 게시된 글의 건수가 가장 많은 것으로 보이는데, 이는 실제로 병원홈페이지에 접속해서 정보를 얻어 가는 사람보다 아직은 한번쯤 둘러보는 사용자들이 많기 때문에 이러한 결과가 나타나는 것으로 사료된다. 특이한 점은 치학정보 사이트를 요구한 사람이 많다는 점이다. 이것은 현재 심하게 아프지 않거나 전혀 아프지 않은 사람들이 다양한 정보를 얻고자 하는 것으로 보인다. 따라서, 치과병원에서의 무선 콘텐츠는 치학정보, 사이버진료실, 자료실, 진료안내가 그 주된

내용으로 구성되는 것이 바람직할 것으로 생각되며, 대부분이 게시판 형식이므로 구현 과정에서 현재 쓰이고 있는 무선 인터넷의 환경을 고려하여야 할 것이다.

### 6.2. 무선 인터넷 서버의 로그 파일

치과병원의 경우 현재 무선인터넷 환경이 갖추어지지 않아 본 논문에서는 무선 인터넷 로그 파일을 분석하지는 않았다. 하지만, 추후 무선 인터넷 환경이 이루어진다면 무선 인터넷 로그 파일도 반듯이 분석을 해야 한다. 유선 인터넷과는 다르게 무선 인터넷은 핸드폰 단말기나 PDA(Personal Digital Assistant)등과 같은 무선 인터넷 단말기와 유선 인터넷환경에서 시뮬레이터로 접속을 하게 된다. 유선 인터넷 로그 파일의 분석과 마찬가지로 무선 인터넷 로그 파일도 사용자 패턴 분석 및 접속 환경 분석은 꼭 필요한 부분이 될 것이다. 특히, 무선 인터넷 환경의 표준화가 이루어지지 않은 현 시점에서는 접속환경 분석은 상당히 중요한 부분이 된다. 왜냐하면, 기간망 사업자에 따라 WML(Wireless Markup Language), HDML(Handheld Devices Markup Language)등 사용 언어가 틀리기 때문이다. 따라서, 접속자의 환경을 조사하여 무선 인터넷 서버 환경을 재조정 할 수 있기 때문이다. [표 9]는 무선 인터넷 로그 파일을 보여주는 것으로서 로그 분석 방법도 유선 인터넷 분석 방법과 유사할 것으로 사료된다.

[표 9] 무선 인터넷 로그파일

Date	Time	Client IP	Server IP	Port	Method	Request File	Type	Client Browser
2001-11-05	13:05:15	211.114.224.70	211.114.224.70	80	GET	/wap/search.wml	404	UPG1+UP/4.0.5c
2001-11-05	13:05:19	211.114.224.70	211.114.224.70	80	GET	/wap/tutorial.wml	404	UPG1+UP/4.0.5c
2001-11-05	13:05:21	211.114.224.70	211.114.224.70	80	GET	/wap/mail.wml	404	UPG1+UP/4.0.5c
2001-11-05	13:06:49	211.114.224.70	211.114.224.70	80	GET	/wap/login.asp	404	UPG1+UP/4.0.5c
2001-11-05	13:06:56	211.114.224.70	211.114.224.70	80	GET	/wap/index.asp	200	UPG1+UP/4.0.5c

### 6.3. 마케팅 전략

치과병원의 홈페이지 특성상 다른 전자상거래를 하는 페이지에서 펼치는 것과 같은 마케팅전략을 사용하게 되면, 병원의 이미지만 나쁘게 보이게 될 것이다. 병원이라는 특이성을 감안해서 신뢰감과 믿음을 줄 수 있어야 한다. 따라서, 현재 사이버진료실을 찾는 사용자들에 대한 체계적인 관리가 필요하다. 왜냐하면, 사이버진료실을 찾는 사용자들이 실제 병원의 고객이 될 가능성이 높기 때문이다.

#### 6.3.1. 회원제 실시 방안

현재 치과대학병원은 회원제를 실시하지 않고 있다. 하지만, 점진적으로 회원제를 추진해야 될 것이다. 회원제를 실시해야만 홈페이지를 접속한 사용자들에 대한 효과적인 관리가 이루어질 수 있기 때문이다. 하지만, 무작정 게시판에 글을 쓰는 사람들에

게 사용자 등록을 요구한다면 사용자들은 다른 사이트로 이동을 하게 될 것이다. 초창기에는 회원제를 실시하되 필요로 하는 사람들에게만 실시하면 된다. 즉, 회원에 가입하게 될 경우 지속적으로 이메일이나 핸드폰 문자 메시지로 치환정보를 보내주면 등록된 회원을 유지시킬 수 있을 것이다.

### 6.3.2. 가족관계시스템의 적용

기존의 CRM기법들은 서구적인 기법들로서, 고객을 단순히 마케팅의 대상으로만 여겨 왔다. 하지만, 이러한 방법은 다소 유효적이고 친족관계에 집착하는 사람들에게는 마케팅활동 자체를 단순히 물건을 팔기 위한 하나의 수단으로 생각하게 할 것이다. 즉, 고객과의 유대관계와 신뢰감, 공감대 형성이 이루어지지 않게 됨을 의미한다. 기업이 이러한 고객과의 유대관계와 신뢰감, 공감대를 형성하고 활용 할 수 있다면 보다 효율적인 마케팅을 할 수 있을 것이다. 즉, 현재 치과병원을 내원하는 환자들을 자연스럽게 온라인 사용자들로 유치하면 된다. 실제 진료를 받는 사람이라면, 회원으로 가입하게 될 것이며, 자신이 병원에 가야하는 날짜나 오늘 진료 받은 것에 대한 정보를 이메일이나 핸드폰으로 연락을 받게 되면 고객은 신뢰감을 느낄 수 있을 것이다. 이러한 전략을 꾸준히 사용하면서 고객들에게 가족관계를 유도하여 한 집안 구성원 전체를 치과병원에서 관리해준다는 느낌을 받았을 때 시간이 지나감에 따라 한번 지나쳐 가는 고객이 아닌 가족전체가 이용하는 병원이 될 것이다.

## 7. 결론 및 추후 연구과제

인터넷의 발전과 더불어 기업과 고객이 만나는 대고객 채널은 점점 더 다양해지고 있으며, 고객의 수준도 높아지고 있다. 이러한 시점에서 기업은 과거에 수행했던 마케팅 방법으로는 더 이상 새로운 고객유치는 물론 기존 고객들을 우리의 고객들로 유지할 수 없다. 이와 같은 관점에서 CRM이라는 개념이 등장하게 되었으며, CRM을 도입하고자 하는 기업은 점점 늘어나고 있는 추세이다. CRM도입을 위한 방법론은 여러 가지가 있으나, 그 중의 하나가 웹 로그 분석일 것이다. 현재까지 진행된 웹 로그 분석은 주로 일반적인 전자상거래를 하는 기업들을 중심으로 이루어져 왔다. 본 논문에서 사용된 예제 데이터는 치과병원 로그데이터로서 다른 데이터들과는 특이성을 가지고 있다. 병원이라는 기업은 다른 기업과는 달리 사회 봉사적 이미지가 강하기 때문에 6장에서도 언급했듯이 고객과의 유대관계 및 신뢰감, 공감대 형성이 중요하다. 따라서, 병원에 진료차 방문하는 고객들로 하여금 고객의 전화번호나 가족관계를 유도하여 문자 메시지 나 E-Mail을 통하여 지속적으로 구강정보나 진료예약 날짜 등을 제공해 준다면, 병원을 찾는 고객들은 병원과의 유대감을 가지게 될 것이고, 다음에 내원하는 횟수가 증가될 것으로 예상된다. 또한, mCRM(mobile CRM) 적용 시 마케팅에 중점을 두는 것보다는 정보전달에 중점을 둔 mCRM 구현이 병원의 장기적인 수익구조에 도움을 줄 것으로 사료된다. 본 논문의 의의는 병원이라는 특이

성을 가진 기업을 대상으로 웹 로그 분석을 실시하여 익명 사용자들의 패턴을 분석하였고, 개인화를 위한 무선 인터넷 콘텐츠를 추출하였다는 점이다. 또한, 데이터 중심의 마케팅이 아닌 유대관계 및 신뢰감을 바탕으로 한 마케팅에 대한 연구란 점에 의의를 두고 있다. 익명 사용자들의 웹 경향분석을 통하여 전체적인 고객들의 패턴을 알 수 있고, 그 패턴을 웹 콘텐츠의 중요도 구성에 적용 시킬 수 있다. 익명 사용자들을 대상으로 한 패턴분석은 전체적인 흐름밖에 알 수 없지만, 회원제를 하고 있는 웹사이트의 경우는 특정고객의 성향을 파악하는데 중요한 자료가 될 수 있으며, 익명 사용자들에 대한 웹 경향 분석은 회원제 적용 및 지속적 회원확보를 추구하는 마케팅을 위한 기저가 될것으로 판단된다. 앞으로 의료기관을 대상으로 하는 다양한 CRM적 분석과 그 적용에 관한 연구가 추후 연구과제로 남고, 본 논문에서 제시한 웹로그 분석 방법과 개인화를 위한 무선 인터넷 콘텐츠 추출 방안은 의료기관뿐 아니라 다른 분야에서도 적용될 수 있을 것이다.

## Acknowledgement

본 연구는 2001년 두뇌한국21 지원사업에 의해 지원되었음.

## 8. 참고문헌

- [1] Ming-Syan Chen, Jong Soo Park and Philip S. Yu, "Data Mining For Path Traversal Patterns in a Web Environment", In Proceedings of the 16th International Conference on Distributed Computing Systems, pp. 385-392, Hong Kong, May, 1996.
- [2] 김종승, 조진호, "CRM 최신 트렌드 및 도입전략", 비티엘글로벌, 2001.
- [3] 무선인터넷백서편찬위원회, "무선인터넷 백서 2001", (주)소프트뱅크미디어, 2001.
- [4] 박중수, "웹로그 파일에서 빈발 항목집합 탐사", 성신여자대학교 기초과학연구지, pp. 446-449, 1999.

## 저 자 소 개

### 임 영 문 :

- 연세대학교 학사, 석사학위 취득. 텍사스주립대학교 산업시스템공학 공학박사  
미국 ARRI연구소에서 선임연구원 및 연구교수, 현 강릉대학교 산업공학과 교수로 재직중.

### 김 흥 기 :

- 현재 강릉대학교 산업공학과 대학원 재학. 관심분야는 WAP & WEB Application