

웹기반 정보시스템의 성능 및 수익성 극대화를 위한 고객시스템 튜닝 방법론

황성하*, 이강수*
*한남대학교 컴퓨터공학과

Customer System Tuning Methodology for Maximizing Capability and Profitability of a Web-based Information System

Sung-Ha Hwang*, Gang-Soo Lee*
*Department of Computer Engineering, Hannam University
e-mail : hsh0408@se.hannam.ac.kr

요 약

정보통신 인프라의 발전으로 상품에 대한 구매 의존도가 실제상점에서 웹 쇼핑몰로 변화하고 있다. 이러한 결과로 웹 쇼핑몰들이 기하급수적으로 증가하고 있다는 것을 알 수 있다. 따라서, 고객이 원하는 상품을 신속하고 안전하며 편리하게 구매할 수 있도록 서비스의 질적 향상을 위한 시스템 개발의 필요성이 대두되고 있다. 이를 위해선 고객들이 원하는 바를 충족시키고 만족시킬 수 있는 소비자(이하, 고객) 행동패턴의 획득과 폼 배치가 필요하다. 본 논문에서는 HCI 이론과 고객 행동론을 이용한 고객행동패턴 획득, 상품 레이아웃 이론을 이용한 웹 재구성, 웹서버의 워크로드 특성 및 클러스터링들을 조사, 활용 및 개량하여 고객시스템 튜닝 방법론(폼 배치 및 구조 설계방법 제시, 고객 행동패턴 획득 시스템 개발)을 제시하였다.

1. 서론

최근에 인터넷과 웹을 기반으로 하는 웹기반 정보시스템(예:e-비즈니스, 전자상거래 시스템)들이 폭발적으로 늘어나고 있다. 예컨대, 수익성을 목적으로 하는 웹 쇼핑몰이 그러하다. 기존의 웹 쇼핑몰의 경우, 고객의 행동패턴에 대한 이해부족과 폼 배치의 미흡, 서버의 과부하로 인한 접속불량, 서비스 시간초과와 지연으로 인해 웹 쇼핑몰이 재구실을 못하고 있는 실정이다. 고객 시스템은 고객 중심 설계와 실제 상점의 매장배치 및 비즈니스 플로우 구조에 해당하므로, 고객행동패턴과 폼(form)의 배치(layout)와 웹서버의 워크로드/클러스터링에 대한 이해가 필요하다.

일반적으로 고객들은 각 웹페이지를 꼼꼼히 살펴보고 세심하게 정리된 텍스트를 모두 읽어본 다음, 제일 먼저 관심을 끄는 링크나 찾고자 하는 항목과 비슷한 이름의 링크를 클릭하면서 대부분의 시간을 보낸다. 그렇기 때문에 고객들은 그들이 방문한 웹 쇼핑몰에 대한 전체 내용을 대부분 보지 못한 채 넘어가는 경우가 발생한다. 따라서, 고객행동패턴을 획득하여 고객에 대한 서비스 질을 높이고, 빠른 시간에 원하는 바를 쉽게 찾을 수 있도록 폼을 재구성할 필요가 있다. 예컨대, 이러한 폼 배치는 쇼핑몰 경영자에게는 수익성을, 고객에게는 서비스 시간을 절약하는 이점이 있다. 또 한가지 생각해 보아야 할 것이 기존의 웹기반 정보시스템들은 다수의 고객이 쇼핑몰에 접속할 때, 고객 서비스 요청에 대한 응답속도가 느려지는 경우가 발생했다. 아무리 최고의 고객 서비스를 기반으로 구축된 웹 쇼핑몰이라 할지라도 다수의 고객이 서비스를 요청할 때, 응답속도가 느려진다면 다시는 그 쇼핑몰에 접속하기를 꺼려할 것이다.

이를 위해 웹 서버에 대한 워크로드 분석과 클러스터링을 통한 측정으로 위와 같은 문제를 해결할 수 있다. 이러한 배경에서, 본 연구에서는 고객행동패턴과 폼 배치, 워크로드와 클러스터링을 분석, 측정, 평가, 보완하여 보다 향상된 웹기반 정보시스템을 개발하기 위해 "웹기반 정보시스템의 성능 및 수익성 극대화를 위한 고객시스템 튜닝 방법론"을 제시하고자 한다.

본 논문은 2장에서 웹기반 정보시스템에 관련된 연구동향과 문제점을 검토하며, 3장에서는 최적의 고객시스템 개발을 위한 고객시스템 튜닝 방법론을 제시하고, 4장에서는 고객시스템 튜닝 방법론에 대한 사례연구를 기술한다. 마지막으로 5장에서는 결과를 정리하고 앞으로의 연구방향을 끝으로 결론을 맺는다.

2. 관련연구

최근 인터넷과 웹을 기반으로 하는 웹기반 정보시스템(예: e-비즈니스, 전자상거래 시스템)에 대한 워크로드 분석, 성능평가 및 튜닝에 관한 방법론인 Capability Planning Methodology(CPM)에 대한 연구가 조지메이슨 대학의 Minasce 교수 연구팀에 의해 활발히 진행되고 있다[1]. 이 분야와 관련된 연구내용을 간략히 살펴보면 다음과 같다.

● 성능평가 및 워크로드 분석 모델

큐잉이론에 기반을 둔 시스템 성능평가 모델 및 방법[2]들이 1970년대에 연구되었으며, 이 결과들은 현재에도 그대로 사용되고 있다.

● 클라이언트/서버 성능평가

1990년대에는 LAN이나 WAN으로 통한 클라이언트/서버 시스템에 대한 성능평가가 활발하며[3], 기존의 큐잉 모델들을 이용하고 있다. 특히 Menache[4,5]는 메인프레임과 클라이언트/서버 시스템의 성능 특성의 차이를 제시하고 있으며 워크로드 및 성능평가를 통한 클라이언트/서버 시스템의 "CPM모델"을 제시하고 있다.

(*) 본 연구는 한국과학재단 목적기초연구(과제번호 : 2001-1-51300-002-2) 지원으로 수행되었음

● 웹 서버 성능평가

1990년 중반이후 웹서버의 수적인 증가와 웹의 기능확장(예, HTML, XML, CGI, 쿠키기능 등)에 따라, 기존의 정보시스템은 점차 웹기반으로 이전되고 있으며 웹서버에 대한 워크로드 분석, 성능평가 및 튜닝에 관한 연구가 활발히 진행되고 있다. 특히, 조지메이슨 대학의 Menasce[6] 교수팀은 웹 서버의 워크로드 특성은 기존의 시스템과는 다름을 보였고, 이에 따른 성능평가 방법들을 제시하고 있으며, Kant와 Won[7]은 서버 용량계획 방법을 제시하였고, Kilela[8]는 웹서버의 서비스시스템(예; Unix, HTTP, TCP, IP, 브라우저, 웹서버프로그래밍, CGI, Java, 데이터베이스, HTML 등)들의 특성들을 제시하여 최적의 성능을 갖는 튜닝의 필요성을 추상적(큐잉이론이나 실제 자료를 사용하지 않음)으로 제시하였다.

● 웹기반 정보시스템의 성능평가

1990년대 말부터 Menasce 교수팀에 의해 기존의 성능평가 이론 및 소비자 행동 모델 등을 이용하여 e-비즈니스 모델(B2B, B2C, G2B 등), e-비즈니스 하부구조, e-비즈니스를 위한 CPM, 웹기반 정보시스템의 성능평가와 워크로드 분석을 큐잉이론 및 정량적으로 평가하는 연구가 진행되고 있다. 이는 웹기반 정보시스템에 대한 연계적인 접근이며 공학적인 접근방법이라 할 수 있다[1].

그러나 기존의 성능평가 연구[1~4]와 웹 용량계획, 웹서버 튜닝방법[5~9]은 현재의 웹서버 특성 중 성능요소만을 고려하였으며, 기존의 off-line 비즈니스 프로세스를 on-line(즉, 웹기반) 비즈니스 프로세스로 최적화하는 튜닝(또는 리엔지니어링) 방법이 부족하다. 또한 고객시스템 튜닝을 위해 기본적으로 제시되는 고객행동패턴에 대한 연구가 부족하다. 고객시스템(즉, 웹기반 정보시스템에서의 사용자 입력 및 출력 시스템)의 구조는 실제 상점의 매장 배치 및 구조에 해당하므로, 폼(form) 배치(layout)는 매우 중요하다. HCI 분야에서도 이 문제를 다루지만 웹 쇼핑몰에서의 고객행동패턴 획득과 폼 배치 방법에 대한 연구는 부족하다. 또한 웹 서버 구조와 특성을 이용하여 고객 시스템 구조(폼 배치)를 설계하고, SW 프로세서 구조를 설계하고 S/W 프로세서 구조를 H/W로 매핑하는 체계적인 방법이 부족하다.

3. 고객 시스템 튜닝 방법론

고객 시스템 튜닝 방법론이란 컴퓨터의 주된 사용자가 극소수의 전문가로부터 일반사용자로 증가함에 따라 컴퓨터 시스템 설계시 고객의 행동패턴을 고려한 고객 중심 설계와 이에 상응하는 폼 배치 및 구성으로 효율성 높은 시스템을 개발하기 위한 방법론이다.

고객 시스템 튜닝을 위한 방법은 <그림 1>과 같다. 먼저 1단계에서 고객행동자료를 획득하기 위해 웹 서버에 저장되어 있는 로그데이터를 분석한다. 로그 데이터는 고객이 웹 사이트를 이용하면 이에 대한 기록이 로그라는 형태로 기록되는 것으로 접속자 수와 방문자 수, 방문자들의 분류, 웹 사이트의 페이지별, 시간별 등의 정보를 알 수 있다. 이를 분석하여 고정 데이터베이스로 구성된 고객정보 테이블에 저장한다. 2단계는 1단계에서 분석한 고객행동패턴 결과를 바탕으로 일반 질의(SQL)를 이용하거나 차트, 테이블 방식 즉, 디자인적 정보를 분석하는 OLAP, 고객에 대한 숨겨진 행동패턴을 예측하기 위한 데이터 마이닝을 통해 경영자의 이익을 위해 고객의 취향과 원하는 바를 충족시킬 수 있는 고객 중심 마케팅 솔루션을 위한 고객관계관리(CRM)를 생성한다. 생성된 자료에 HCI 이론[10] 및 고객 행동론[11]을 포함하여 고객들의 성향을 분석하고 웹 사이트내 페이지 방문횟수를 통해 어느 페이지에 주로 관심을 나타내는지 아닌지를 알아내어 웹 재구성에 유용하게 반영할 수 있다. 고객관계관리(CRM)는 기업의 수

익성을 위한 전략이며, 전략 수행을 위해 필요한 고객 마케팅 방식이자, 고객관리의 컨셉이자, 솔루션의 하나이다.[12]

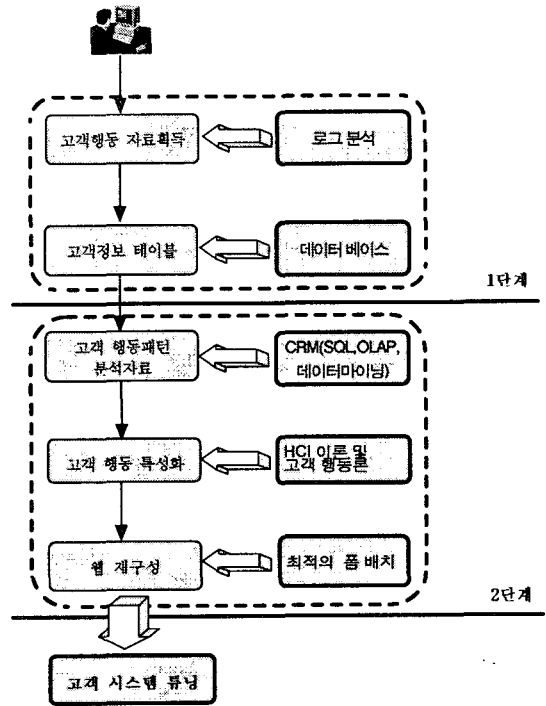


그림 1. 고객 시스템 튜닝 방법론

<그림 2>는 고객 행동 패턴을 획득하기 위한 1단계의 과정을 보여 준다. 웹 서버의 로그데이터에 기록된 각 고객에 대한 고객행동패턴을 고정크기로 구성된 버퍼에 저장하면 세션시간에 입력된 데이터를 저장할 고객행동 로그 테이블(단기 DB)이 생성되며 이를 분석하면 일정

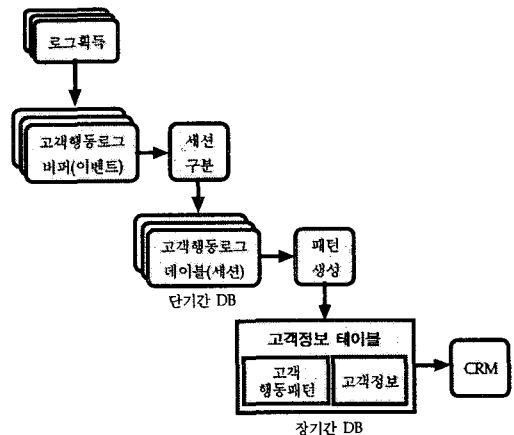


그림 2. 고객 행동 패턴 획득과정

패턴이 생성된다. 분석된 결과 데이터인 고객행동패턴 파일과 각각의 고객에 대한 상세 정보가 들어있는 고객정보 파일을 종합한 결과를 장기간 저장하고 업데이트 할 고객정보 테이블이 생성한다. 고객정보 테이블

에 고객관계관리(CRM) 즉, SQL, OLAP나 데이터 마이닝을 이용하여 고객 마케팅 전략에 효과를 높일 수 있다.

<표 1>은 고객에 대한 행동패턴을 획득하기 위해 데이터가 저장될 파일구조를 나타낸다. 파일에 대한 구성요소를 구분함으로써, 일반 질의(SQL)를 통해 여러 가지 고객 정보를 획득할 수 있다. 즉, <표 2>와 같이 데이터베이스 스키마 구성에 의해 파일에 해당하는 데이터만 제대로 갖춰져 있다면 일반 질의(SQL)를 이용하여 고객행동패턴에 대한 새로운 방향의 예측으로 최적의 고객시스템을 구축하는데 필요한 정보를 충분히 얻을 수 있다.

표 1. 파일구조

파일	구성요소
고객파일	고객 n에 대한 각각의 레코드
고객레코드	고객 ID, 고객정보, 고객행동패턴, 고객세션, 고객행동참조
고객행동참조	로그인 TS, [페이지], 로그아웃 TS
고객정보	고객주소, 고객성별, 고객 연령, 직업
고객행동패턴	i×j 매트릭스

[]:순열, TS:타임스탬프

표 2. 데이터베이스 스키마

변이	저장기간, 길이	단기간, 고정길이 DB	장기간, 가변길이 DB
	변이 DB	고객행동로그	행동패턴
불변이 DB			페이지 배치, 고객정보

<그림 3>은 <그림 2>에서 세션구분에 대한 내용으로 로그데이터를 획득하기 위해 세션구분의 기준을 정한 것이다. 세션은 고객의 연속된 행동의 단위를 말하며, 세션 구분은 페이지 사이의 시간간격에 일정시간(예, 30분)이 넘는 경우 세션의 끝으로 파악한다. 이와 같이 특정시간(예, 30분)을 기준으로 세션의 단위를 정한다면 그 만큼 분석과정을 쉽게 일반화 시킬 수 있겠지만 그 만큼 분석의 정확도는 떨어진다. 왜냐하면 웹 사이트의 성격에 따라서 체류시간 등이 천차 만별이기 때문이다. 예컨대, 실생활화표는 정식절차를 수행한 고객의 세션을 말하며 점선화표는 타임아웃(time out)에 의해 생성된 고객의 세션이다.

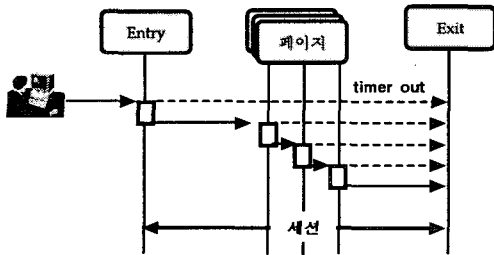


그림 3. 세션구분

4. 사례 연구

본 논문에서는 웹 정보 시스템 중에서도 고객에게 가장 큰 비중을 차지하고 있는 웹 쇼핑몰을 연구대상으로 했다. 위에서 언급한 고객 시스템 튜닝 방법론을 쇼핑몰에 적용할 경우 고객이 무슨 제품에 주로 관심을 보였는지 알 수 있고, 나아가 특정 제품을 장바구니에까지 넣었으나 구매하지 않은 경우도 파악하여, 향후 고객이 다시 방문했을 때 실제 구매를 일으킬 수 있도록 유도할 수 있으므로, 모든 종류의 웹기반 정보 시스템 튜닝시에 이 연구결과를 그대로 활용할 수 있다.

4.1 폼 배치 및 구조의 설계

<그림 4>는 백화점 쇼핑몰에서의 페이지 배치와 구조를 나타낸다. 예컨대, 고객이 처해 있을 수 있는 페이지 사이에 고객이 백화점 쇼핑몰 사이트에서 방문해 있는 동안 고객 행동 패턴을 나타낸다. 각 페이지를 분석하고

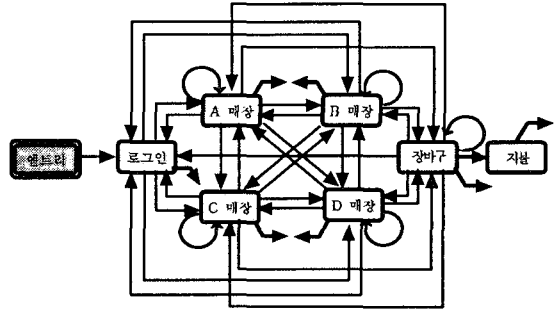


그림 4. 백화점 쇼핑몰에 대한 폼 배치 및 구조

<표 3>은 <그림 4>의 각 페이지에 대한 배치를 매트릭스로 표현한 것이다. 예컨대, 엔트리 열과 나가기 행은 기본적으로 어떤 상태에서도 엔트리 페이지로 거꾸로 네비게이션하지 않고 나가기 페이지 또한 다른 페이지로 네비게이션하지 않으므로 0으로 한다. 여기에서 고객 네비게이션의 최소화를 위한 배치구조의 최적화(즉, 폼 배치의 최적튜닝)가 필요하다.

표 3. 페이지 배치 매트릭스

	0	1	2	3	4	5	6	7	8
엔트리(0)	0	1	0	0	0	0	0	0	0
로그인(1)	0	0	1	1	1	1	1	0	1
A매장(2)	0	1	1	1	1	1	1	0	1
B매장(3)	0	1	1	1	1	1	1	0	1
C매장(4)	0	1	1	1	1	1	1	0	1
D매장(5)	0	1	1	1	1	1	1	0	1
장바구니(6)	0	1	1	1	1	1	1	1	1
저장(7)	0	0	0	0	0	0	0	0	1
나가기(8)	0	0	0	0	0	0	0	0	0

4.2 고객 행동패턴 획득

고객이 1회 방문하는 동안, 고객은 모든 상태를 다 방문할 수는 없을 것이며, 동일 고객의 여러 차례에 걸친 사이트 방문이나 서로 다른 고객의 여러 차례에 걸친 사이트 방문은 상태 방문 빈도 면에서 다를 수 있다. 따라서 사이트 방문 시간에 고객 행동의 특징을 규정하기 위해서는 상태 방문 빈도를 파악해야 한다. 상태 방문 빈도는 고객의 매장내 방문패턴(즉, 행동패턴)을 의미하며 고객마다 다르다. 이를 획득하기 위해서는 특정 고객에 대한 지속적인 모니터링이 필요하다. 고객별 행동패턴은 마케팅에 활용 할 수 있으며, 전체평균 행동패턴은 웹 서버의 성능튜닝을 위해 사용한다.

우선, 고객에 대한 고객행동패턴을 획득하기 위해선 로그데이터를 분석해야 한다. 분석된 로그파일을 고객행동 벡터를 이용하여 고객행동패턴 매트릭스를 만든다. 고객행동 벡터는 가변길이면서 순서정보를 포함하며 고객행동패턴 매트릭스는 고정길이면서 순서정보를 가지지 않으며 방문횟수를 저장한다. 웹 쇼핑몰에 접속한 고객에서 세션식별에 의해 처리된 100명의 로그데이터를 저장할 이벤트 버퍼를 분석하여 고객행동 로그 테이블을 만든다. 예컨대, 고객 1의 행동벡터=<0, 1,

2, 3, 3, 4, 5, 7, 8>, 고객 2의 행동벡터=<0, 1, 3, 4, 5, 5, 3, 8>, 고객 3의 행동벡터=<0, 1, 2, 4, 8>, 고객 4의 행동벡터=<0, 1, 3, 3, 6, 4, 6, 7, 8>, 고객 5의 행동벡터=<0, 1, 8>이다. 고객행동 벡터를 가변으로 정한 이유는 고객이 다수의 전이횟수를 사용할 수 있다는 가정하에 정한 것이다. 이런 방식으로 고객 100명에 대한 행동벡터로 구성되었다고 가정한다. 이와 같은 고객의 행동벡터를 통해 접속자 수와 페이지 방문고객의 수를 파악할 수 있으며 고객이 한 페이지를 방문해서 다른 페이지로 이동한 전이횟수를 알 수 있다.

<표 4>은 고객 n의 행동벡터를 분석한 고객행동패턴 매트릭스를 나타낸다. 이 매트릭스 결과를 통해 고객에 대한 페이지 방문횟수와 방문물을 통해 고객이 자주 방문한 페이지를 알 수 있으며, 고객 ID를 통해 고객주소와 고객성별, 고객연령, 직업에 따른 고객행동패턴을 획득할 수 있다. 나아가 획득한 고객행동패턴을 이용하여 워크로드 특성과 클러스터링을 분석할 수 있다. 고객행동패턴 매트릭스는 결과를 종합하여 장기간 저장되는 고객정보 테이블에 기록된다. 고객정보 테이블을 일반 질의(SQL)를 사용하면, 고객에 대한 요일별, 월별, 연도별 고객의 행동패턴 등의 고객정보를 획득함으로써, 고객에 대한 마케팅 전략에 이용할 수 있다.

표 4. 고객행동패턴 매트릭스

페이지 고객 n	0	1	2	3	4	5	6	7	8
고객 1	1	1	1	2	1	1		1	1
고객 2	1	1		2	1	3			1
고객 3	1	1	1		1				1
고객 4	1	1		2	1		2	1	1
고객 5	1	1							1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

<그림 5>는 웹 쇼핑물에서 발생하는 워크로드(단위 작업 레코드 구성)에 대한 예로 점들(CPU, I/O횟수)은 트랜잭션을 나타낸다. 클러스터링 분석은 대량의 워크로드에 대한 유사한 워크로드의 평균값을 클러스터로 그룹화 시킨다. 각 점들의 CPU 시간과 I/O 횟수의 평균 트랜잭션을 계산하여 유사한 트랜잭션을 클러스터 C1, C2, C3처럼 그룹핑한다. 이와 같은 클러스터링 분석은 웹 서버에 대한 워크로드 성능을 높이는데 사용된다. 즉, 클러스터링 분석을 함으로써 클러스터 중심점으로부터 각 점의 거리를 최소화 하고 하나의 클러스터와 다른

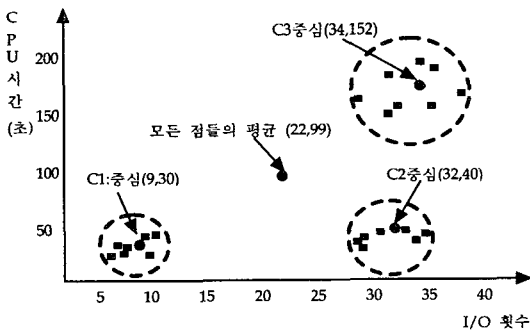


그림 5. 웹 쇼핑물의 트랜잭션에 대한 클러스터링

클러스터의 거리를 최대로 해야 한다. 예컨대, 각 클러스터간의 거리가 짧아지면 과도한 트랜잭션을 요하므로 웹 서버에 과부하가 발생할

수 있다. 이것의 측정결과로 워크로드의 특성을 파악할 수 있다. 워크로드의 특성으로는 첫째, TCP/IP의 특성인 비연결 방식이므로 작업(work)의 처리방식이 다르고 둘째, 고객은 8초 이내에 요구한 작업이 수행되기를 원하므로("8초규칙"이라함) 데드라인이 정해진 시스템이며 셋째, 작업은 예측 불가능하며 넷째, 웹 서버의 과부하로 인한 접속불능이나 대기는 웹 쇼핑몰에 큰 타격을 줄 수 있다. 즉, 다른 경쟁 쇼핑몰로 갈 것이며 향후에도 접속을 꺼릴 것이다.

5. 결론

본 연구는 인터넷을 기반으로 하는 정보통신 인프라가 발전함에 따라, 정보시스템의 패러다임이 변화하므로 인해 웹기반 정보시스템(고객 시스템)을 구축할 때, 기존의 성능평가 방법이나 자료들을 그대로 적용할 수 없다. 특히, 고객 중심 설계로 이루어진 고객 시스템은 더 그러하다. 고객 시스템이 지닌 가장 큰 특징 중의 하나는 바로 고객의 관심과 행동 양식을 파악한 가운데 전개 될 수 있다는 점이다. 특히, 웹 쇼핑몰의 경우 경쟁자의 수익성과 얼마만큼의 고객의 입장에서 설계되었는지는 고객평가에 따른다. 이러한 배경에서 고객행동패턴을 획득하기 위해 로그분석을 중심으로 측정된 통계자료와 HCI 이론, 고객 행동론 등의 이론을 적용하여 고객이 원하는 고객 중심 시스템을 구축하기 위해 고객 시스템 튜닝 방법론을 제시한다. 고객 시스템 튜닝 방법으로 고객행동패턴을 획득한 자료를 분석하여 최적의 고객시스템을 튜닝하는 기술을 연구 개발함으로써, 최대의 수익을 얻기 위해 전문 컨설턴트를 양성할 수 있으며 최적의 웹기반 정보시스템을 구축할 수 있다. 최적의 고객 시스템의 구현과 보안성에 관한 문제는 향후 연구과제로 남긴다.

6. 참고 문헌

- [1] D. Menasce, *Scaling for E-business: Technologies, Models, Performance, and Capacity Planning*, Prentice-Hall, 2000.
- [2] L. Kleinrock, *Queuing Systems*, Weley, 1975.
- [3] P. Harrison, N.Patel, *Performance Modeling of Communication Network and Computer Architecture*, Addison Wesley, 1993.
- [4] D. Menace, *Capacity Planning and Performance Modeling: from Mainframe to Client-server Systems*, Prentice-Hall, 1994.
- [5] D. Menasce and H. Gomaa, "A Method for Design and Performance Modeling of Client/Server System," *IEEE tran.on S.E.*, Vol 26-11, Nov. 2000.
- [6] D. Menasce, *Capacity Planning for Web Performance - metric, moelis, & method*, Prentice-Hall, 1998.
- [7] K. Kant and Y. Won, "Server Capacity Planning for Web Traffic Workload," *IEEE tran. on Knowledge and Data Engineering*, Vol. 11-5, Sep./Oct., 1999.
- [8] P. Killela, *Web performance tuning*, O'reiley, 1998.
- [9] F. Arlitt, et al., "Internet Web Servers : Workload Characterization and Performance Implications," *IEEE/ACM trans. Networking*, Vol.5, No.5, 1997.
- [10] J. Dix, E. Finlay, D. Abowd, Russell Beale, *Human-Computer Interaction*, Prentice-Hall, 1998.
- [11] J.E Engel, *Consumer Behavior(8th Edition)*, The Dryden Press, 1995.
- [12] B. Mobasher, et al., "Automatic Personalization Based on Web usage Mining," *Communication of the ACM*, Vol.43, No.8, 2000.