

# 웹 로그분석을 이용한 실시간 온라인 마케팅 시스템 설계 및 개발

## Development of Realtime Online Marketing System using web log analytics

오재훈(Jaehoon Oh)\*, 김재훈(Jaehoon Kim)\*\*, 김종우(Jongwoo Kim)\*\*\*  
jaehoon@nethru.co.kr, jaehoon1980@nethru.co.kr, kjw@hanyang.ac.kr

### 초 록

e-Business 시장의 폭발적 성장으로 기존의 오프라인 기업들이 e-Business 시장으로 진입하게 되었으며, 이는 e-Business 시장의 과도한 경쟁을 야기시켰다. e-Business의 경쟁은 새로운 기술 및 서비스를 요구하게 되었으며, 특히 웹 사이트 방문자의 행동정보를 실시간으로 분석하고 개인화된 정보를 제공할 수 있는 시스템에 대한 요구가 증대되고 있다.

이에 본 연구에서는 사이트 방문자의 행동을 실시간으로 수집·분석하여 모니터링 할 수 있는 실시간 온라인 마케팅 시스템 ROMS(Realtime Online Marketing System)를 구현하였다. ROMS는 웹 로그 분석을 이용한 실시간 상담 기술, 사용자 프로파일, 컨텍스트(Context) 분석 기술 및 이벤트 탐지·처리 기술을 활용한다. 이러한 ROMS는 방문자 행동 수집, 방문자 행동 분석, 방문자 프로파일, 방문자 모니터링, 실시간 상담 및 마케팅 등의 기능을 제공한다.

### 1. 서론

인터넷과 IT 기술의 발전은 e-Business 시장의 성장과 함께 e-Business 시장 참여자의 확대로 이어졌다. 이는 e-Business 시장의 과도한 경쟁을 야기시켰으며, 시장의 경쟁에서 우위를 점하기 위하여 온라인 기업의 마케팅 투자 및 노력이 증대되고 있다. 인터넷 마케팅 협회(2009)에 따르면 전

체 마케팅 비용 7조 9천억 원 중 온라인 마케팅 비용이 1조 2978억 원으로 전체 광고시장의 16.4% 를 차지하고 있다[11].

그러나 최근의 고객들은 온라인 커뮤니케이션 기술의 발달로 다양한 경로를 통하여 상품 정보를 파악하고 타인의 소비 경험을 공유함으로써 제품에 대한 정보와 지식을 수집하고 있다[12]. 특히 가격 비교 사이트를 이용하여 다양한 판매 채널의 판매 가격

---

본 연구는 서울시 산학연 협력사업(JP100080)의 연구비 지원으로 이루어 졌음

- \* (주)넷스루 연구소 소장
- \*\* (주)넷스루 연구소 연구원
- \*\*\* 한양대학교 경영학부 교수

을 한 눈에 파악할 수 있기 때문에 기업들의 경쟁이 더욱 심화되어 가는 실정이다. 뿐만 아니라 온라인 상에서의 고객행동이 복잡해져 감에 따라 동적으로 변하는 고객의 행동정보를 파악하기가 점점 어려워지고 있다.

그러나 기존의 온라인 마케팅 방법들은 이러한 현상들에 대하여 충분히 인식하지 않고 있으며, 이에 대한 대응이 미흡한 실정이다. 실제 이메일 마케팅의 부정적 인식 및 검색광고 시장의 치열한 입찰 경쟁, 고객들의 가격 비교 쇼핑에 따른 마진 감소 등으로 기존의 온라인 마케팅의 투자 대비 효과는 기대만큼의 성과를 제공하지 못하는 실정이다.

이에 온라인 기업들은 새로운 온라인 마케팅 방법을 요구하고 있다. 특히 고객의 행동을 실시간으로 분석하고 고객의 니즈를 즉시 파악하여 마케팅을 수행할 수 있는 시스템에 대한 요구가 증대되고 있다.

본 연구에서는 웹 로그분석을 이용한 실시간 온라인마케팅 시스템 ROMS(Realtime Online Marketing System)을 설계 및 구현하였다. ROMS는 온라인 사이트 방문자의 행동을 실시간으로 수집하기 위한 방문자 행동 수집 기술, 고객의 행동 정보를 이용한 Context 분석기술, Context 기반의 사용자 프로파일 기술 및 탐지·처리 기술을 활용한다. ROMS는 웹 사이트의 방문자를 모니터링하고 시스템 사용자가 사이트 방문자에게 실시간으로 마케팅을 실행할 수 있는 시스템이다.

## 2. 관련연구

### 2.1 웹 로그 분석 기술

#### 2.1.1 방문자 행동 수집 기술

웹 로그(web log)란 웹 사이트 방문자들이 제품이나 서비스를 구매하는 과정을 통해 발생하는 데이터이다. 정보획득이나 구매를 목적으로 인터넷 사이트를 방문하는 방문자들은 로그의 형태로 사이트 내에 흔적을 남기는데 이러한 데이터를 기반으로 해서 다양한 정보를 추출해 내는 것이 웹 로그 분석이다[16].

이러한 웹 로그분석을 위하여 웹 사이트 방문자 행동을 수집하는 기술은 크게 웹 서버 로그파일 방식(Log File), 페이지 태깅방식(Page Tagging), 패킷 스니핑(Packet Sniffing) 방식의 세가지로 분류할 수 있다 [10].

웹 서버 로그파일 방식은 웹 서버에 남은 로그파일을 분석하는 방식으로 일정기간 기록된 데이터를 분석하는 배치형 분석에 많이 사용된다. 서버 수가 적고 서버 관리자와 분석가가 일치하거나 동일한 조직에 속한 경우 많이 사용된다.

패킷 스니핑 방식은 네트워크를 통과하는 패킷을 분석하여 HTTP 패킷 만을 추출하여 로그파일에 기록하는 방식이며, 분석 방법은 로그 파일 방식과 동일하다.

페이지 태깅 방식은 방문자가 페이지를 조회 하거나 특정한 행동을 할 때마다 페이지 정보를 수집하는 방식이다. 페이지 태깅 방식에는 하드 태깅(Hard Tagging)방식과 소프트 태깅(Soft Tagging)방식이 있다. 하드 태깅 방식은 웹 브라우저에서 분석대상 정보를 가공하여 서버로 전송하는 방식이다. 소프트 태깅 방식은 웹 브라우저는 기본 정보만을 전송하고 전송된 정보를 서버에서 분석하여 중요 정보를 추출하는 방식이다. ROMS 는 두가지 방식을 모두 지원하는 Hybrid 방식으로 방문자 행동 정보를 수집한다.

### 2.1.2 방문자 행동 분석 기술

초기의 웹 로그분석은 웹 사이트 상에서 고객의 행동을 분석하고, 이를 웹 사이트 디자인에 반영하기 위하여 사용되었다[2].

그러나 e-Business 성장과 함께 온라인 상에서의 고객행동이 복잡해져 감에 따라 웹 로그분석은 고객의 행동 패턴을 분석하는데 사용되고 있다.

Fu(1999)는 웹 사용자의 접속 패턴을 클러스터링하고 일정한 규칙을 제공하는 방법을 개발하였고, Igor(2000)는 확률 기반의 군집 프레임워크를 제안하고 고객들의 웹 사이트 방문 형태를 다양하게 분류하였다[2][18][19].

이러한 웹 로그분석은 웹 사이트 방문자들이 많이 본 페이지를 방문페이지로 변경하거나, 웹 사이트에서의 방문경로를 파악하여 상품을 배치하는 등 고객의 행동을 파악하고 마케팅을 수행하기 위한 방법으로 이용되고 있다.

이러한 웹 로그분석은 고객의 콘텐츠 선호에 따른 웹 사이트의 효율적 설계 및 관리를 지원하고 있다. 또한 시장분석과 프로모션에 투입되는 마케팅 비용을 감소시키며, 고객의 성향과 행동패턴을 분석할 수 있다[3].

## 2.2 온라인 마케팅 기술

### 2.2.1 실시간 상담 지원 기술

온라인상의 상담시스템은 '웹 협업시스템' 영역에 속한다. '웹 협업시스템'이란 컴퓨터를 통한 사람들 간의 정보의 공유, 수정, 협업적 정보생산 작업등을 포함한다[6]. 온라인 상담시스템은 용도 및 목적에 따라 의료, 정신상담, 학습 등 다양한 기능 및 목적으로 활용되고 있다[1][5][7].

이러한 온라인 상담시스템은 e-Business

시장의 성장과 함께 고객 마케팅을 위한 도구로 진화해 왔다. 실제 국내외 다양한 업체들이 실시간 마케팅 도구로써, 실시간 상담기능을 제공하는 시스템을 개발하여 제공하고 있다[22][23][24][25][26][27].

온라인 상담시스템은 문자채팅, 음성 및 화상 대화, 파일전송 및 로컬정보, 고객 연결 승인 및 강제 종료 등의 기능을 제공하고 있다.

### 2.2.2 방문자 프로파일 기술

방문자 프로파일은 방문자의 기본정보 및 부가정보 등을 담는다. 이러한 방문자 프로파일은 방문자를 구분하고 방문자에게 맞는 서비스를 제공하기 위하여 이용된다[13].

방문자 프로파일은 성별, 나이, 연령, 직업 등의 정적 프로파일과, 방문자의 상황이나 피드백된 서비스 내용 등으로 얻을 수 있는 동적 프로파일로 구분할 수 있다[4]. 방문자 프로파일은 방문자들이 관심 있는 부분이나 실제로 접하고 있는 내용을 근거로 방문자에게 적합한 서비스를 제공할 수 있다[8].

IT 기술의 발전은 개인의 정보를 쉽게 파악할 수 있게 되었으며, 개인화된 서비스 제공을 위한 방문자 프로파일에 대한 연구역시 증가하고 있다.

Fang Liu et al(2004)는 방문자가 과거에 방문했던 범주에 속하는 웹 페이지들에 우선순위를 부여하여, 개인화된 검색을 구현하였다. 이러한 개인화된 검색의 구현을 위하여 방문자 프로파일이 사용되었다[9][17].

Jaime Teevan et al(2005)가 사용한 시스템은 웹 페이지, 전자 메일, 클라이언트 컴퓨터에 저장된 문서등 방문자에 의해 생성, 복사되거나 방문자가 본 모든 정보에 대해

인덱스를 생성한다. 이 모든 정보가 방문자 프로파일을 만드는 데 사용되었다[20].

### 2.2.3 컨텍스트 분석 기술

컨텍스트(Context)는 “개체의 상황을 특성화 하는데 사용할 수 있는 어떤 정보이고, 여기서 개체는 사람, 장소, 물리적 또는 컴퓨팅 객체이다.”로 정의할 수 있다[15]. 이러한 컨텍스트는 사용자가 처한 환경에서 사용자의 현재 위치, 행동, 환경 및 작업 등 사용자에게 대한 정보와 그 정보들의 지속적인 변화를 말한다. 컨텍스트는 활용분야에 따라 사용자의 ID, 위치인식 등 다양하게 활용 가능하다[8][14].

특히 온라인 상에서의 컨텍스트 분석은 온라인 마케팅을 위한 기술로 점점 부각되어 왔다. 그러나 현재 실시간으로 고객의 관심 상품이나 콘텐츠를 Catch하는 간편적인 기술은 많은 웹 사이트에서 응용되고 있으나, 방문 고객의 특성과 관심사를 분석하고 요약하는 기술에 관한 연구는 미비한 실정이다.

## 3. 실시간 온라인 마케팅 시스템 요구사항

### 3.1 시스템 요구사항

본 연구에서 제안하는 ROMS의 요구사항을 분석하기 위하여 유스케이스(Use Case) 분석 방법을 이용하였다. ROMS의 액터(Actor)로는 사이트 방문자, 시스템 사용자로 구분하여 액터별 요구기능을 도출하였다. 도출된 요구사항은 회원관리, 사이트 관리, 사용자 관리, 마케팅, 모니터링의 5분류로 구분하였다. 5가지 요구사항에는 다양한 기능들을 포함하며, 그 중 핵심 기능을 <표

1>과 같이 정의하였다.

<표1> 시스템 요구사항

요구사항 분류		사용자	
		사이트 방문자	사이트 관리자
회원관리	회원가입/탈퇴		○
	회원 정보변경		○
사이트 관리	사이트 추가/삭제		○
	사이트 정보변경		○
사용자 관리	사용자 추가/삭제		○
	사용자 권한변경		○
마케팅	상담	○	○
	메시지 전달	○	○
	메시지 조건 설정		○
모니터링	방문자 정보		○
	방문자 행동 정보		○
	방문자 필터링		○
	실시간 분석		○

사이트 방문자는 ROMS에 등록된 웹 사이트에 접속한 방문자로서, 사이트 관리자에게 상담을 요청하거나, 사이트 관리자가 전송한 마케팅 메시지를 전달받을 수 있어야 한다.

사이트 관리자는 ROMS에 접속하여 서비스를 이용하는 사용자로서, 회원관리, 사이트 관리, 사용자 관리 및 방문자 모니터링을 통한 마케팅을 수행할 수 있어야 한다. 먼저 회원관리에서는 회원가입, 탈퇴, 정보변경 기능을 정의하였다. 사이트 관리에서는 시스템에서 제공하는 서비스를 이용할 대상 사이트를 추가, 삭제, 정보변경 기능을 정의하였다. 사용자 관리는 서비스를 이용하기 위하여 추가된 사이트를 관리 할 사용자를 추가, 삭제 할 수 있으며, 사용자의 이용 권한을 변경할 수 있도록 정의하였다.

그리고 마케팅 기능으로 사이트 관리자는 사이트방문자에게 상담메시지와 마케팅 메시지를 전송할 수 있다. 또한 사이트 관리자가 설정한 조건에 따라 자동으로 마케팅 메시지를 전송할 수 있다.

### 3.2 고객 행동정보 수집 요구사항

ROMS가 수집하는 고객 행동 정보는 유입 정보, 인구통계학적 정보, 행동 정보, 클라이언트 정보로 구분할 수 있다. 각 항목에 대한 자세한 내용은 <표 2>와 같다.

<표2> 고객 행동정보 수집 요구사항

요구사항 분류	세부 항목
유입 정보	방문경로[직접/참조/검색], 참조 페이지, 랜딩 페이지, 검색 엔진, 검색어, 광고 방문 정보[CPC/CPT/배너/ 이메일]등
인구통계학 정보	나이, 성별, 지역 등
행동 정보	조회 페이지, 조회 상품, 조회 카테고리, 내부 검색어, 장바구니 상품, 구매 상품 등
클라이언트 정보	브라우저, 운영체제, 해상도, IP, 지역 등

유입정보의 경우 사이트 방문자가 사이트 방문 시 방문경로를 파악한다. 방문 경로의 경우 크게 사이트 URL 및 즐겨 찾기를 통하여 방문한 직접방문, 특정 페이지의 링크를 통하여 접속한 참조방문, 검색을 통하여 접속한 검색방문으로 구분하여 파악하여야 한다. 또한 특정 유료광고를 통하여 사이트 방문 시 해당 광고 정보를 파악한다.

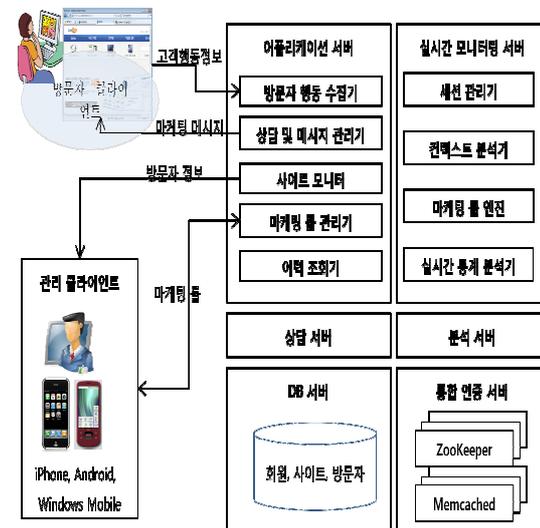
인구통계학 정보로는 사이트 방문자의 나이, 성별, 지역 등의 정보를 파악하며, 클라이언트 정보로는 사이트 방문자의 브라우저, 운영체제, 해상도, IP, 접속 지역등을 파악한다.

마지막 행동 정보로는 사이트 방문자가 사이트에 방문하여 조회한 페이지, 조회 상품 및 카테고리, 내부 검색어, 장바구니 상품, 구매 상품 등의 행동을 수집한다.

## 4. 실시간 온라인 마케팅 시스템의 구현

### 4.1 시스템 구성

ROMS 는 웹 서비스를 기반으로 구축되며, <그림 1>과 같이 방문자 클라이언트, 관리 클라이언트, 어플리케이션 서버, 실시간 모니터링 서버, 상담 서버, 분석 서버, DB 서버, 통합 연동 서버로 구성된다.



<그림 1> ROMS 시스템의 구성

어플리케이션 서버는 외부 액터와 상호 작용을 담당하며 WAS(Web Application Server) 위에서 Web Application으로 동작

한다. 어플리케이션 서버는 방문자 행동 수집기, 상담 및 마케팅 메시지 관리기, 사이트 모니터, 마케팅 룰 관리기, 이력 조회기 등으로 구성된다.

실시간 모니터링 서버는 수집된 방문자 행동을 분석하여 방문자의 컨텍스트를 추출하고 방문자 프로파일을 작성한다. 모니터링 서버는 세션 관리기, 컨텍스트 분석기, 마케팅 룰 생성기, 마케팅 룰 엔진, 실시간 통계 분석 등의 모듈로 구성된다.

방문자 클라이언트는 사이트에 방문자가 접속하면 방문정보를 어플리케이션 서버의 방문자 행동 정보 수집기로 전달하고, 방문자 행동 정보 수집기는 행동 이력을 로그로 기록한 후에 모니터링 서버로 전송한다. 방문자가 상담을 요청하면 상담 및 메시지 관리기로 전달된다.

상담 서버는 방문자와 사이트 관리자 사이의 상담을 관리하며, 분석 서버는 배치 분석 보고서를 담당한다. 통합 인증 서버는 어플리케이션 서버상의 어플리케이션간의 통합 인증 기능과 캐쉬 기능을 제공한다. 통합 인증서버의 성능을 향상시키기 위하여 오픈 소스인 memcached의 메모리 캐쉬 기능을 사용하며, ZooKeeper를 이용하여 분산 시스템을 구성함으로써 고 가용성을 달성할 수 있도록 구성하였다[21][28].

#### 4.2.2 상담 클라이언트

상담 기능에는 상담 메시지 전송, 상담 내용 조회, 마케팅 메시지 조회 등의 기능이 있다. 상담에는 두 가지 종류로 구분된다. 방문자가 사이트 관리자에게 요청하는 상담과 사이트 관리자가 방문자에게 요청하는 상담(proactive chat)이다.

Proactive 상담을 지원하기 위하여 방문자 클라이언트가 주기적으로 상담 서버에 상담

내용을 요청하도록 하였다. 또한 방문자 클라이언트는 상담중의 상담 상태에 따라 요청 주기를 조절할 수 있도록 구성하였다.

#### 4.3 모니터링 서버

모니터링 서버는 세션 관리기, 컨텍스트 분석기, 실시간 마케팅 엔진, 실시간 통계 분석기로 구성된다.

세션관리기는 방문자의 행동 정보를 세션 단위로 구분한다. 방문자 클라이언트는 상담 서버에 주기적으로 상담 정보를 요청하는데 이 요청을 Ping 메시지로 변환하여 모니터링 서버에 전달함으로써 방문자의 세션을 유지하도록 하였다. 방문자가 사이트를 떠나면 Ping 메시지 전송이 중지되며 일정 기간이 경과하면 세션관리기가 방문자의 세션을 종료한다.

컨텍스트 분석기는 수집된 행동 정보를 컨텍스트 모델에 따라 표현한다. 컨텍스트 분석기는 방문자 컨텍스트에 사용할 변수를 추출하여 방문자 세션에 설정한다. 마케팅 엔진은 방문자의 컨텍스트가 정의된 마케팅 메시지의 전송 조건과 일치하는 경우에 마케팅 메시지를 방문자에게 전달한다. 방문자에게 마케팅 메시지를 전달하기 위하여 방문자 클라이언트는 마케팅 메시지가 있는지 주기적으로 검사하도록 하였다. 실시간 통계 분석기는 방문자의 행동을 분석하여 사이트의 통계를 제공한다.

#### 4.4 관리 클라이언트

사이트 관리자가 사이트 모니터링을 위하여 사용하는 관리 클라이언트는 Flash 기반으로 구현하였으며, <그림 2>와 같이 구성하였다.



---

## 참고문헌

---

- [1] 고대근, 남윤희, “화상교육 시스템을 이용한 사이버 집단 상담의 효과 분석”, 한국콘텐츠 학회 논문지, 제7권 제4호, pp.213-223, 2007.
- [2] 고재문, 서준용, 김운식, “수주생산기업 B2B에서 e-CRM을 위한 웹 로그 분석”, 산업공학, 제18권 제2호, pp. 205-220, 2005.
- [3] 김정하, 김소은, “웹 로그 분석을 통한 여행사 웹 방문 경로에 관한 연구”, 관광 레저연구, 제16권 제3호, pp.209-226, 2004.
- [4] 김종원, “사용자 프로파일에 기초한 유즈넷 뉴스 그룹 자동 결정 방법”, 한국지지 및 지능 시스템학회 논문지, 제14권, 제2호, PP.142-149, 1999.
- [5] 변광준, 예홍진, 위규범, “멀티미디어 원격상담 도우미의 설계 및 구현”, 정보과학회 논문지, 제4권 제4호, pp. 418-426, 1998.
- [6] 서여진, 박우창, 송순오, “전자상거래용 대화형 상담 시스템 설계 및 구현”, 한국정보과학회, 2003년도 학술발표 논문집, 제30권 제1호, 2003.
- [7] 이병문, 조원희, 윤영미, “실시간 심음 전송을 위한 원격 의료상담시스템”, 한국콘텐츠 학회논문지, 제9권 제8호, pp.158-16, 2009.
- [8] 이세일, 이상용, “실시간 컨텍스트 정보의 정량화 단계를 개선한 령력적 필터링”, 한국퍼지 및 지능시스템 학회지, 제17권 제4호, pp.288-493, 2006.
- [9] 이재원, 김응모, “서비스 제공자 목록에 의한 사용자 프로파일 기반 웹 서비스 검색의 정확도 향상”, 한국정보과학회 가을 학술발표 논문집, 제33권 제2호, 2006.
- [10] 이종현, 임혜영, 황준, “차세대 정보 마이닝 동향”, 인터넷정보학회지, 제7권 제1호, pp53-66, 2006.
- [11] 인터넷 마케팅 협회, 2009년 인터넷 광고비 예측. 2009.
- [12] 정주훈, 이호배, “실시간 마케팅의 선행요인 및 성과 연구”, 한국마케팅 저널, 제9권 제4호, pp.1-41, 2008.
- [13] 최정환, 장현후, 최현우, 김구수, 엄영익, “홈 네트워크 환경에서 사용자 위치에 기반한 지능적인 서비스 알고리즘”, 한국정보처리학회 춘계학술 대회 논문집, 제14권 제1호, 2007.
- [14] A.K.Dey, "Context-Aware Computing: The CyberDesk Project", Proc. of the AAAI 1998 Spring Symposium on Intelligent Environments(AAAI Technical Report SS-98-02), pp.51-54, 1998.
- [15] Anind K. Dey and Gregory D. Abowd, “Towards a Better Understanding of Context and Context-Awareness”, Technical Report GIT-GVU-99-22, Geogia Institute of Technology, College of computing, 1999.
- [16] Cooley, R., Mobasher, B., & Srivastava, J.. “Data preparation for mining world wide web browsing patterns”, Knowledge and Information System, Vol.1, No.1, pp.123-132, 1999.
- [17] Fang Liu, Clement Yu, Weiyi

Meng, "Personalized Web Search for Improving Retrieval Effectiveness", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 1, pp.28-40, 2004

[18] Fu, Y., K. Sandhu and M.Y. Shih, "Clustering of Web Users Based on Access Patterns", Proceedings of the Workshop on Web Usage Analysis and User Profiling(WEBKDD'99), San Diego, CA, August, 1999.

[19] Igor V. C., Scott G. and P. Smyth, "A general probabilistic framework for clustering individuals and objects", Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp.140-149, 2000.

[20] Jaime Teevan, Susan T. Dumais, Eric Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities", In the proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp.449-456, 2005.

[21] <http://memcached.org/>

[22] <http://tocplus.co.kr>

[23] <http://www.http.or.kr>

[24] <http://www.liveweb.co.kr>

[25] <http://www.olark.com>

[26] <http://www.whoson.com>

[27] <http://www.wibiya.com>

[28] <http://www.zookeeper.com>