

웹로그 마이닝을 통한 인터넷 쇼핑몰에서의 사용자 행동 분석

이동하^a, 김성민^a, 오재훈^a, 서동렬^a, 임규건^b

Dong-Ha Lee, Sung-Min Kim, Jaehoon Oh, Dong-Yal Seo, Gyoo Gun Lim

^a ㈜ 넷스루

135-978, 서울시 강남구, 역삼동 679-5 서울벤처타운 1501호

Tel: +82-2-6243-4162, Fax: +82-505-318-3200, E-mail: {dongha, jaehoon, ksmoon, dyseo}@nethru.co.kr

^b 세종대학교 경영학과

143-747, 서울시 광진구 군자동 98

Tel: +82-2-3408-3169, Fax: +82-2-3408-3310, E-mail: gglim@sejong.ac.kr

서론

인터넷 웹 사이트 상에서 사용자 행동은 클릭(click)을 단위로 모두 로그(log)에 기록된다. 웹 서버를 통해 남는 웹로그를 가공하여 단순한 통계 수치 외에, 사용자 행동을 분석할 수가 있다. 특히 인터넷 쇼핑몰에서 사용자의 행동에 대한 분석은 중요하며, 고객의 획득, 유지 전략을 수립하기 위한 중요한 정보가 된다. 본 논문에서는 인터넷 쇼핑몰에서의 사용자 행동을 비즈니스 관점에서 분석한다. 쇼핑몰 사이트의 유입 경로 분석의 다양한 관점에 대해 논의하며, 관심 카테고리 및 상품 분석, 첫페이지 영역별 분석 등 새로운 분석 방법에 대해 소개한다. 이와 함께, 이 분석과정에서 필요한 효율적인 데이터 구조, 운영체 데이터 베이스 정보 및 이들간의 연동방안과 분석 결과의 활용 방안을 제시한다.

Keywords:

웹로그, 관심 상품 분석, 첫페이지 분석, 경로 분석

1. 소개

인터넷이 보편화됨에 따라 인터넷 기업은 자신의 사이트를 방문하는 방문자 행동을 보다 정확히 분석하고자 하는 노력을 기울이고 있으며, 온라인 기업뿐만 아니라, 공공기관이나 오프라인 기업을 포함하여, 거의 모든 웹 사이트에서 웹로그 분석을 수행하고 있다. 대표적인 온라인 기업 형태중의

하나인 인터넷 쇼핑몰의 경우, 방문자는 기업의 고객이며, 방문자의 행동을 분석하여, 매출과 직결되는 의사결정에 활용될 정보를 추출하고자 하는 욕구가 있다.

이에 따라, 웹로그 분석을 위한 다양한 상용 제품이 출시되어 있고, 웹로그 분석에 대한 연구도 점차 활발해지고 있다.

웹로그는 웹 서버가 파일에 남기는 접근 정보로서, 방문자가 접근한 하나의 파일당 한 행(line)의 로그가 단편적으로 남는다. 웹로그의 포맷(format)은 웹 서버에 따라 다양하지만, 대체적으로 유사하며, 가장 중요한 부분은 표준으로 지정되어 있다.

포맷과 관계없이 공통적인 정보로서는, 웹로그에 저장되는 접근 시간(정확하게는 요청이 완료된 시간), IP, 사용자가 요청한 파일명, 응답 상태(status □ 요청한 파일의 전송 완료 여부 등) 등이 있는데, 이들 기본적인 정보는 누가, 언제, 무엇을, 어떻게 했는지에 대한 정보를 담고 있는 것이고, 이 정보를 분석하면, 웹 이용량에 대한 통계를 얻을 수 있다.

최근 널리 쓰이고 있는 IIS 로그의 경우, 사용자 IP주소, 사용자이름, 날짜, 시간, 서비스 및 인스턴스 컴퓨터 이름, 서버 IP주소, 걸린시간, 받은 바이트수 보낸 바이트수, 서비스 상태코드, Window 상태코드,

요청형식, 작업대상 등의 정보를 담고 있다.

웹 사이트를 유지하는 기업들은 단순히 사이트를 유지하기만 하는 것이 아니라, 웹 방문자에 대한 이해를 통해 의사결정에 필요한 정보를 얻고자하는 욕구가 생겼다. 초기에는 단순히 웹 이용량에 대한 통계 정보 정도를 요구했었지만, 이제는 방문자가 어떤 의도를 가지고 있는가에 대한 분석 요구사항이 생기기 시작했다.

웹로그는 방문자에 대한 충분한 정보를 가지고 있기 때문에, 트래픽 통계만을 추출하는데 그치지 않고 고객에 대한 세밀한 분석을 수행하고자하는 요구사항이 생긴 것이다. 이를 위해, 새롭고 다양한 기법들이 도입되었다. 대표적인 방법으로는 시나리오 분석이나, 유입 분석이 있다.

또한 확장된 포맷(extended format)의 웹로그에서는 쿠키(cookie)를 통해 사용자의 로그인 아이디를 로그에 남길 수 있으며, 이를 통해 개별 사용자를 인식할 수 있다. 쿠키정보외에도 사용자의 웹사이트 로그인 아이디를 남기는 방법이 더 있으며, 사용자 로그인 아이디와 운영체 데이터베이스의 고객 DB를 조합하여 활용하는 경우, 방문자의 구체적인 행동을 분석할 수 있게 된다. 더 나아가, 개별 방문자의 구체적인 “의도”를 유추할 수 있게 되는것이다. 방문자의 의도를 분석하는 것은 단순한 웹 통계분석과는 다른 방법론이 필요하다. 좀더 세밀한 준비를 통해 사용자의 행동을 분석하여야 하며, 데이터 마이닝에 가깝다고 할 수 있다. 본 논문에서는 기존 웹로그 분석 기법들을 개괄하고, 쇼핑몰에서의 고객 행동 분석 요구 사항을 제시한다. 이와 함께 고객 행동을 분석하기 위한 분석 방법과 구현 방법을 제시하고 구현 시스템의 구조를 제안한다.

2. 웹로그 분석

웹로그는 좁은 의미로는 사용자가 접근한 파일에 대한 로그만을 담은 접근 로그(access log)만을 대상으로 하는데, 확장된 의미로는 접근한 파일 이전에 사용자가 접근한 파일(웹 분석 분야에서는 이 파일을 referrer라고 부름)에 대한 정보를 담고 있는 참조 로그(referrer log)와 사용자가 사용한 브라우저에 대한 정보를 담고 있는 에이전트 로그(agent log)를 포함한다. 한편 최근 대부분의 웹

서버에서는 쿠키(cookie)를 이용하여, 개별 방문자를 인식하기도 하고, 방문자에 대한 정보를 기록하고 있다. 본 논문에서 언급하는 웹로그 분석의 대상은, 접근 로그, 참조 로그, 에이전트 로그와 로그에 남은 쿠키정보를 포함한다.

또한 최근 웹로그 분석 기술의 발달로 인해, 운영체 DB와의 연결을 통한 추가 정보를 활용하는 경우도 있으므로, 이러한 정보가 수집되는 상황을 고려한다.

웹로그 분석의 본래적인 목적은 트래픽 분석이었지만, 최근 많은 발전을 한 결과, 누가, 언제, 어디를 통해 웹 페이지를 방문했는가, 무엇을 이용했는가 등에 대한 다양한 분석을 목적으로 하게 되었다. 웹로그 분석 보고서를 관심 대상 정보에 따라, 표 1과 같은 형태로 분류할 수가 있다.

표 1. 관심 대상에 따른 분석 결과 보고서

관심 대상	분석 보고서
방문자수, 페이지수	시간/일/주/월 기간 단위 현황과 추이 회원 속성별 현황 (나이/성별/지역...) 서비스 및 콘텐츠 그룹별 현황 사업 단위별 웹 비즈니스 성과 측정
컨텐츠	인기 컨텐츠 및 상품 인기 카테고리 주로 검색하는 키워드 주로 다운로드 되는 문서
시스템	웹 서버별 동작 현황 에러 발생 현황과 에러 종류 전송 데이터량 현황 동시 접속자 현황
사용자 행동	유입 경로분석 시나리오 분석 경로분석

이러한 분석 보고서들은 거의 모든 웹로그 분석 업체들이 구비하고 있으며, 인터넷 홈페이지를 보유한 업체들은 대부분 웹로그를 분석하고 있다. 하지만, 각 개발 업체마다 조금씩 다른 형태로 구현되고 있는데, 그 이유는 웹로그의 양이 방대하여 구현상의 여러 제약이 있을 수 있다는

근본적인 문제 외에도, 분석하고자하는 상황에 매우 복잡한 경우가 종종 있어서, 분석에 대한 표준적인 방법론이 있을 수 없기 때문이다.

그림 1과 같은 경우, 사이트의 유입에 대한 형태를 보여주고 있다. 외부 호스트로의 분석 보고서는 거의 모든 웹로그 분석 도구들이 제공하지만, 제휴사 코드를 부여하고, 제휴사 코드를 기준으로 분석할 수 있는 도구는 많지 않다.

검색엔진에서의 유입 분석도 대부분의 도구들이 가능하지만, 다양한 검색엔진에서 어떤 검색어를 통해 유입되었는지를 분석하기에는 검색 포털의 구현형태가 매우 다양할 수 있기 때문에 어려움이 있다. 하나의 검색엔진이 여러 서버로 구성된 경우(www1.abc.com, www2.abc.com)가 있을 수 있으며, 포탈인 경우, 검색 창이 다양한 페이지에서 나타날 수 있기 때문이다.

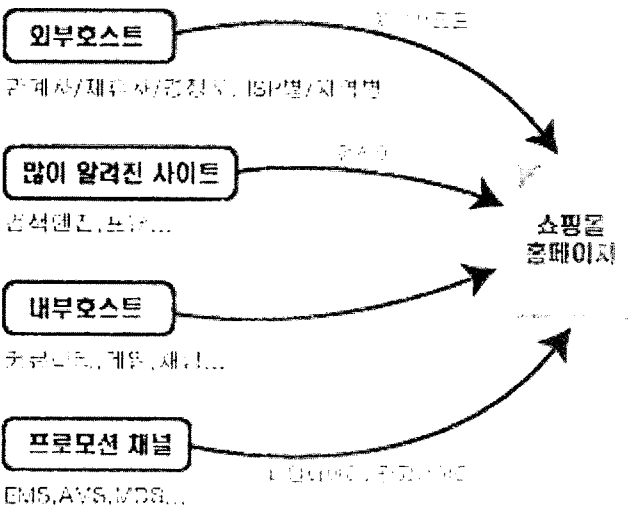


그림 1. 유입 경로 분석

3. 쇼핑몰에서의 분석 요구 사항

다양한 형태의 사이트가 운영되고, 웹로그 분석에도 복잡한 요구사항이 많이 생기고 있다. 유입 경로 분석은 일반적인 사이트에서도 상당히 복잡하지만, 인터넷 쇼핑몰의 경우에는 좀더 복잡하기도 하며, 더 중요해지기도 한다.

인터넷 쇼핑몰에서는 고객 획득과 유지를 위한 다양한 캠페인이 진행되며 사이트의 방문뿐만 아니라 구매에 이르게 되는 경로에까지 분석

요구사항이 확대되기 때문이다.

유입 경로에서는, 유입 호스트, 유입시의 제휴사 코드 및 메일아이디, 광고 아이디를 추출해낼 수 있어야하며, 시작페이지(=사이트 방문 시 방문자가 처음 본 페이지)를 구분할 뿐만 아니라 그 시작페이지가 어떤 이벤트 코드를 지니고 있는지를 구분할 수 있어야한다. 쇼핑몰 사이트인 경우 여러 프로모션이 동시에 수행되기도 하며, 여러 채널에서 이루어지고 있다. 따라서, 각 프로모션 채널마다 시작페이지가 다를 수도 있고 유입 시작페이지에 이벤트 코드를 부여해서 프로모션을 구분하는 방법을 쓸 수 있다. 유입 경로 분석을 수행하기 위해서는 여러 종류의 분석 보고서가 필요하다.

그림 2의 시나리오 분석은 가입이나, 구매 결정에 자주 사용되는 분석방법이다. 이메일을 받고 방문한 고객이 여러 단계를 거치면서 실제 구매 결정까지 도달한 비율을 보여준 것이다. 인터넷에서는 이러한 각 단계의 사용자 행동이 해당 페이지를 본 형태로 구체적으로 웹로그에 남기 때문에, 이들 구매 결정에 대한 시나리오 분석은 구매와 관련된 사용자의 의도를 나타내는 의미를 가진다.

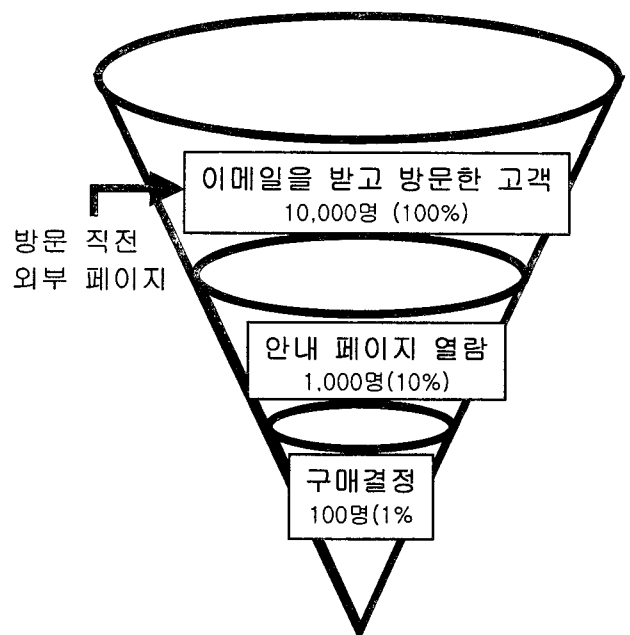


그림 2. 구매 결정에 대한 시나리오 분석

또한 현재의 구매 여부 뿐만 아니라, 방문자가 관심있어하는 상품과 카테고리에 대한 분석을

수행할 수 있다. 실제 구매 내역은 운영계 DB에 남는 것이 보통이며, 구매 내역 분석에는 반품 및 운송비용 등 여러 정보를 고려해야하므로 본 논문의 범위를 벗어난다. 실시간 또는 하루 단위의 분석을 수행하는 웹로그에 비해, 구매완료까지는 15일 이상이 걸릴 수 있는 길고 복잡한 과정이 필요하며, 이에 따라, 웹로그에 남은 행동 분석과 실제 구매 완료까지의 행동을 통합하여 분석하는 것은 근본적으로 해결할 수 없는 부분들이 있다.

한편 관심 상품 또는 카테고리 분석은 웹로그 분석에서 수행해야하는 작업이다. 방문자의 클릭은 매우 빈번한 행동이며 이들 각각의 로그를 직접 DB에서 처리하는 것은 구매내역과 같은 정보와 함께 처리할 내용이 아니다. 또한 이 정보는 사이트 개편에도 영향을 받는다. 대형 쇼핑몰도 사이트 개편이 아주 드문일이 아니며, 몇년 주기로는 행해지고 있다.

방문자가 관심있어하는 상품과 카테고리 정보는 직접적인 구매 페이지가 아닌, 상품상세, 상품확대, 상품FAQ, 상품추천, 상품평클릭, 상품평보기 등의 페이지를 보았던 기록을 수집한다. 각각의 페이지마다 구매 의사에 대한 중요도가 다를 수 있다.

수집한 방문자 관심 정보는 향후 이벤트에 사용될 수도 있고, 방문자에 대한 특성을 파악하는데도 활용된다.

표 2. 쇼핑몰에서의 주요 분석 요구사항

유입 경로 분석	방문자 세션 시작 페이지에 해당되는 로그로부터 외부 유입 호스트, 유입 패러미터, 시작 페이지, 시작 페이지 패러미터 등을 분석하고 모든 세션에 대해서, 외부 호스트, 외부 호스트 패러미터, 시작 페이지 구분, 시작 페이지의 이벤트 구분 작업을 한다.
구매 경로 분석	물건이 장바구니에 삽입되는 Action이 로그에 나타날 경우, 그 이전에 사용자가 접근한 페이지들은 분석하여, 구매에 이르게 된 페이지들을 찾아낸다.
관심 카테고리 분석	특정 페이지들을 관심 카테고리 페이지로 정하고, 해당 페이지를 접근한 사용자를 분석한다.

첫페이지 위치별 분석	첫페이지 방문후 사용자의 행동을
-------------	-------------------

4. 분석 방법

4.1 유입 분석

세션 시작 페이지에 해당되는 로그의 참조 페이지로부터 외부 유입 호스트의 실제 URL을 추출할 수 있다. 이 URL에서 호스트이름과 도메인 이름, 유입 패러미터 등을 알 수 있다. 유입 파라미터는 검색 사이트에서 사용한 검색어 정보를 가지고 있을 수도 있으며, 프로모션 코드가 들어 있을 수 있다.

호스트 정보로부터 관계사인지, 검색포탈인지, 프로모션 채널인지를 구분해야한다. 이후에 유입 패러미터를 추출하여 각 문자열이, 검색어 인지 프로모션 코드 인지를 구분해서 별도의 보고서를 생성해야한다.

방문자가 처음으로 보게되는 시작 페이지 정보도 중요하지만, 시작 페이지에 같이 나타나 있는 패러미터도 이벤트 코드 등 중요성을 담고 있을 수 있는 정보이다. 이 이벤트 코드 정보는 패러미터가 아닌 정적인(static) URL형태로도 나타날 수 있기 때문에 추출이 어려울 수 있다. 따라서, URL내부에 섞여있는 코드를 추출할 수 있는 방법이 필요하며, 이 추출된 코드를 기준으로 분석이 되어야한다.

4.2 구매 경로 분석

인터넷 쇼핑은 장바구니에 물건을 담아 결재하는 과정을 거친다. 기존의 경로 분석은 단순히 장바구니 페이지에 도달하게 되는 페이지를 여러단계에 걸쳐 URL기준으로 단순히 셀 뿐이다. 그러나 실제 구매에 이르게했던 페이지를 파악하는 것이 필요하며, 이는 실제 장바구니에 물건을 담기 전에 어떤 페이지를 보았는지가 중요하게 된다. 따라서, 실제 장바구니 페이지를 보기 직전에 보았던 페이지를 거슬러 올라가서 분석해야만한다. 대부분의 쇼핑몰의 경우, 구매에 이르게 되는 페이지는 이벤트 페이지이거나, 검색페이지, 기획전 페이지 또는 사이트를 서핑중에 장바구니에 담는 경우이다. 이러한 경우 각각의 URL을 그룹화(grouping)해주는 것이 필요하며, 그룹화뿐

아니라, 해당 페이지에서 코드를 추출하는 것도 필요하다.

구매 직전에 검색을 수행하는 경우, 상품 구매에 어떤 검색어를 사용했는가는 중요한 정보이다. 검색어 정보는 일반적으로 검색 페이지에 패러미터 형태로 남겨져 있다. 이벤트 페이지인 경우, 쇼핑몰에서 동시에 행해지고 있는 여러 이벤트들중 어떤 이벤트와 관련된 페이지인지를 코드 형태로 URL이 표현하고 있게 된다.

또한 개별 사용자들마다 구매에 이르게 되는 주요 경로를 추출하면 개별 사용자를 좀더 이해할 수 있다.

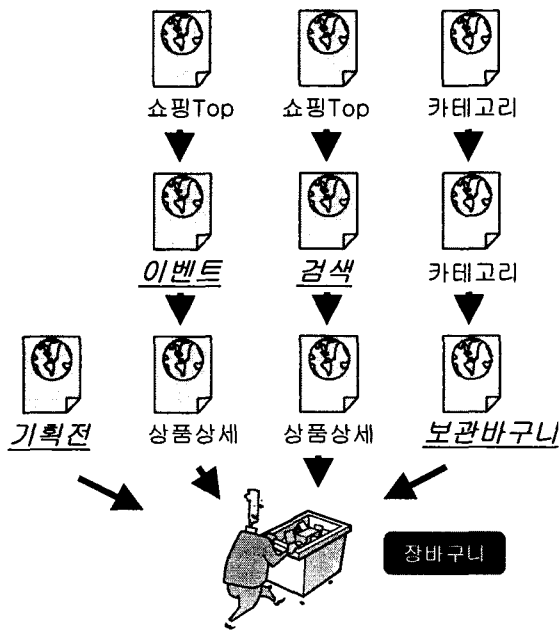


그림 3. 장바구니까지의 경로

4.3. 관심 상품 분석

특정 페이지들을 구매의사 결정 페이지로 보고, 해당 관심 카테고리를 분석하여, 쇼핑몰에서 인기있는 상품을 분석할 수 있다.

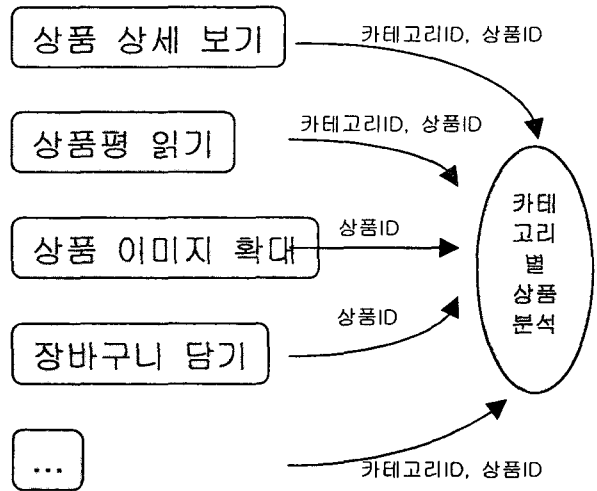


그림 4. 관심 상품/카테고리 분석

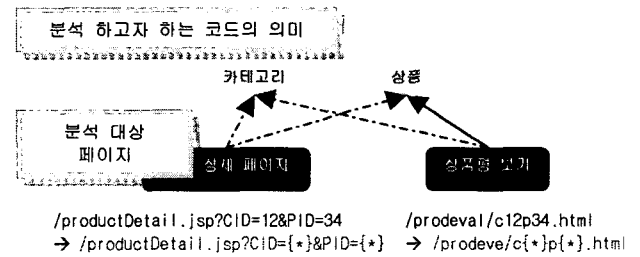


그림 5. 상품/카테고리 코드 기준 분석

카테고리	상품	히트
103668	217056	69
	236863	4
	222951	1
102952	202162	63
	171723	29
		8

카테고리	히트
종수산/식품	4,300
105787	2,901
104001	2,554
카메라/캠코더	2,353
농산&Food	2,341
105821	2,146
가전	2,124
e-슈퍼마켓	2,120
휴대폰/PCS	2,116
냉장고/김치냉장고	2,083
노트북	1,973

그림 6. 관심 카테고리 분석 보고서

카테고리, 상품은 쇼핑몰 사이트에서 분석의 기준이 될 수 있는 가장 중요한 정보중의 하나인데, 그림 5에서는 이들을 기준으로 분석이 어려운 형태를 보이고 있다.

/prodeval/c12p34.html 은 카테고리 와 상품 정보가 URL 내에 숫자형태로 포함되어 있다. 일반적으로 카테고리 와 상품 정보는 패러미터 형태로 나타나지만 이렇게 URL 내부에 나타나기도 하는데, 미리 만들어진 URL인 경우도 있고, 최근 CMS(contents management system)이 널리 쓰이면서 URL을 동적으로 생성해서 패러미터가 없는 형태로 제공하는 경우도 있다. /productDetail.jsp의 경우는 CID와 PID을 이름으로 가지는 두 개의 패러미터가 존재한다. 그런데, 이들은 서로 다른 이름을 사용하지만 각각 “카테고리”와 “상품”을 의미하며, 단지 웹 사이트 구축 시에 파라미터 이름을 통일하지 않고 구현한 것이다. ‘관심 상품’에 대한 분석을 위해서는 2개의 URL을 함께 지정하여 집계해야 하지만 기존의 파라미터 분석은 파라미터 이름을 지정해야 하기 때문에 2개의 URL을 독립적으로 분석하고 그 결과를 합산하는 복잡한 과정이 요구된다.

대부분의 대형 쇼핑몰들이 다양한 제품을 다룰 뿐더러, 해당 시스템 개발에도 여러 과정을 거친 이유로 인해, 같은 의미로 분석되어야 하는 URL이 다른 구조를 가지고 있는 것이다.

그림 5의 하단부와 같이 URL을 그룹화하면서 분석 대상 코드를 추출하여 분석을 해야하고, 분석 결과 보고서의 예를 그림 6에 보였다.

또한 개별 사용자들마다 관심 상품이나 카테고리를 추출하면 개별 사용자를 좀더 이해할 수 있다.

4.4 첫페이지 분석

쇼핑몰에서의 첫페이지는 매우 비싼 자원이다. 사이트에 따라 하루에 수백만 히트 이상의 노출이 되며, 첫페이지에서의 이동을 분석하기만 해도 방문자의 관심도를 상당히 파악할 수 있기 때문이다.

최근 쇼핑몰의 첫페이지에는 여러가지 형태의 이벤트가 동시에 이루어지고 있으며, 매번 접근할때마다 같은 위치에 다른 광고가 나타나기도 한다. 이를 분석하기 위해서는 해당 배너가 클릭될

때 호출될 URL에 다양한 정보를 표현해야할 필요가 있다. 그림 7은 특정 배너에 나타난 URL의 형태이다 다양한 정보가 URL의 패러미터 영역에 조합되어 저장된다. 이들 정보 각각에 대해 분석을 수행하면, 각 배너의 실제 효과를 알 수 있을 뿐만 아니라, 배너의 위치나, 크기, 모양에 따른 방문자의 반응을 분석할 수 있게 된다.

http://shop.abc.net/html/event/eventAppliances_040115_1.html?Sid=0001_52010000_03_07&_top_G=shop=Gimg3

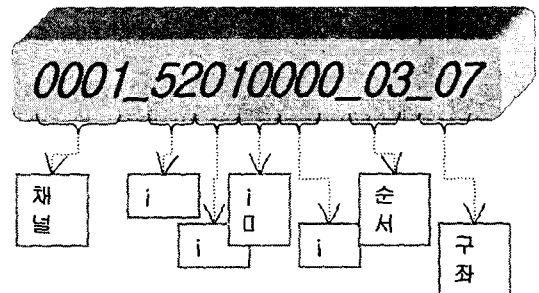


그림 7. 첫페이지 분석

5. URL 패턴 표현식

위와 같은 분석을 가능하게 하기 위해서는 URL에서 코드를 추출하는 표현식이 필요하고, 이 코드를 중심으로 분석을 수행해야한다. 이 표현식을 URL 패턴 표현식이라고 하며, 이의 BNF(Backus Naur Form)는 표 3와 같다. URL에서 코드를 추출한 뒤에는 단일값(single value)으로 처리되는 코드를 중심으로 URL 집계를 수행하므로, 코드 추출 후 더 이상 문자열 연산을 하지 않기 때문에 전체적인 분석 속도가 획기적으로 향상된다.

표 3 - URL 패턴 표현식의 BNF

```
URL_EXP ::= [service://host:port]/PATH[?QUERY][;CODE_LIST]
PATH ::= PATH_SEGMENT[/PATH]
QUERY ::= QUERY_SEGMENT[&QUERY]
CODE_LIST ::= [code_name][, CODE_LIST]
```

URL 패턴 표현식(URL_EXP)은 네트워크 서비스 이름과 호스트이름, 포트이름으로 시작하고, 경로(PATH)와 질의(QUERY), 코드 리스트로 구성되어

있다. 서비스 이름과 호스트이름, 포트 이름 부분은 생략될 수 있고, 이 경우 미리 정해진 분석 대상 기본 웹사이트를 의미하게 된다. 또는 질의와 코드 리스트만으로 구성될 수도 있으며, 이 경우 모든 URL에 나타날 수 있는 질의 부분에 대한 표현이 된다

경로(PATH)는 웹 서버에서 파일이 있는 위치를 지정한다. 경로는 경로 세그먼트(PATH_SEGMENT)들로 구성되고, 경로 세그먼트는 디렉토리 구분자(/, slash)로 구분된다.

질의(QUERY)는 지정된 URL에 전달할 문자열을 지정하는데, 일반적으로 파라미터 (parameter) 부분이라고도 불린다.

경로와 질의 사이에는 물음표(? , question mark)를 사용한다.

질의 세그먼트(QUERY_SEGMENT)들로 구성되고, 질의 세그먼트는 앰퍼샌드(&, ampersand)로 구분된다.

단, 질의 세그먼트는 /prodeval/c12p34.html 다음과 같은 구조를 가진다.

QUERY_SEGMENT ::= paraname=paravalue

paraname은 파라미터 이름이고 paravalue는 파라미터 값이다.

URL 패턴 표현식에서는 PATH_SEGMENT나 QUERY_SEGMENT부분에 코드로 추출할 문자열을 지정할 수 있는 패턴(코드 패턴)이 하나 이상 포함될 수 있으며, 이렇게 추출된 문자열은 코드 리스트(CODE LIST) 부분에 나타나는 코드에 대해 순서대로 연결된다. 코드 이름은 중괄호({, }, brace)로 묶여있고, 각각의 코드 이름은 콤마(,)로 연결된다. 코드 패턴은 앞뒤가 중괄호로 묶이며 중괄호 안에는 아스키(ascii) 문자의 대소문자, 숫자(0-9) 등 일반 URL 에서 사용하는 문자를 그대로 사용할 수 있고 와일드 카드 문자를 이용할 수도 있다.

그림 8은 관심 상품 분석에 사용된 예이다.

URL명	URL 패턴 표현식
상품보기	/ShowProduct?CID={*}&PID={*};{카테고리};{상품}
장바구니담기	/CartAdd?Category={*}&Product={*};{카테고리};{상품}

웹로그URL
/CartAdd?Category=0001&product=aaaa234

그림 8. 코드 패턴을 포함하는 URL 패턴 표현식

6. 시스템 구조

제안된 시스템의 구조를 그림 9에 제시했다. 전처리 과정에서 URL에 내재되어 있는 상품 코드와 유입 사이트 정보 등을 상세하게 추출하는 방법이 필요하며, 이 부분을 처리하는 부분이 Log Convert 이다.

처리된 내용은 운영계 데이터베이스에서 추출된 상품코드, 이벤트 코드들과 함께 분석되며, 향후 분석계 데이터베이스로 취합하여 분석을 수행할 수도 있다.

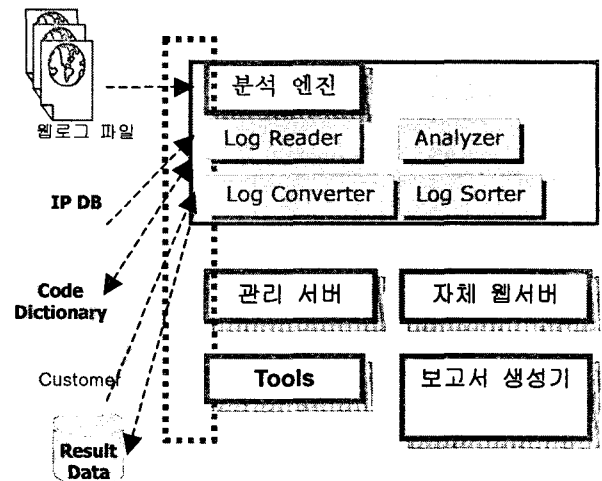


그림 9. 시스템 구조

분석계 DB로 Export 할 수 있는 정보로는 방문자별 URL, 방문자별 외부 URL, 로그인했던 사용자의 방문 내용을 정리한 정보 등이 있으며, 이들 정보를 매일 Export 한다.

제시된 시스템은 현재 다수의 국내 대형 인터넷 쇼핑몰에 적용되고 있다.

6. 결론

사용자 행동은 클릭(click)을 단위로 모두 로그(log)에 기록된다. 이 정보를 활용하는 분석 기법은 점차 발전하고 있는데, 특히 인터넷 쇼핑몰에서 사용자의 행동에 대한 분석은 고객의 획득, 유지 전략을 수립하기 위한 중요한 정보가 된다.

본 논문을 통해서 우리는 유입 분석, 구매 경로 분석, 관심 상품 분석, 첫페이지 위치별 분석은 기존의 웹로그 분석 방법과 다른 방법론이 필요함을 보였고, 이를 분석하는 방법론과 URL에 내재된 코드를 추출하는 새로운 기법을 제시하였다. 본 논문에서 제시한 시스템에 대한 타당성 평가 등의 향후 연구과제가 남아 있다.

7. 참고 문헌

- [1] 김재경, 이건창 권순재(2003). “ 웹로그 데이터 및 퍼지인식도를 이용한 웹쇼핑몰 사용자 구매패턴 분석,” 경영학연구, 32권, 2호, 567p, 한국경영학회.
- [2] 김종달, 김성민, 남도원, 이동하, 이진영. “ 웹 마이닝 시스템 설계 및 유용한 접근 패턴 정의”, 한국지능정보시스템학회, 2000년 6월 23일, p283-291.
- [3] 이건창, 정남호, 김재경, 조운호, 권순재(2002) “ ,클레멘타인 데이터마이닝 솔루션을 이용한 웹 로그 분석,” Information Systems Review, 4권, 1호, 47p, 한국경영정보학회.
- [4] 용환승, 김지현(2003). “ 웹 로그 분석을 위한 OLAP 시스템 및 성능 평가,” 멀티미디어학회논문지, 6권, 5호, 909p, 한국멀티미디어학회.
- [5] 이영재, 임현성(2002). “ 웹 로그 분석을 통해 연관성 규칙을 활용한 개인화 상품 추천의 효과성 연구,” 2002 춘계학술대회 논문집, 749p, 한국경영정보학회.
- [6] 이희석 임미희 최수영(2004). “ 웹 사이트 성공 영향요인 도출과 웹 사이트 유형별 비교 분석,” 경영정보학연구회지, 14권, 2호 한국경영정보학회.
- [7] 정옥란, 강미정, 조동섭(2002). “ 클릭스트림 분석을 위한 웹 서버 시스템의 설계 및 구현,” 정보처리학회논문지D, 9권, 5호, 945p, 한국정보처리학회.
- [8] Cooley, R. (2000) “ Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data,” Ph.D. Thesis.

University of Minnesota. May 2000.

- [9] S.-M. Kim, J.-D. Kim, J.-H. Hong, D.-W. Nam D.-H. Lee, J.-Y. Lee. “ A System for Association Rule Finding in an Internet Portal Site” INFORMS KORMS SEOUL Conference, June/18/2000 ~June/21/2000, pp.999~1005.
- [10] D.-H. Lee, D.-Y. Seo, N.-H. Kim, J.-Y. Lee, “ Discovery and Application of User Access Patterns in the World Wide Web,” Proceedings of the 4th World Congress on Expert Systems , March 16-20, 1998, pp.321-327.
- [11] B. Mobasher, N. Jain, E.H. Han, J. Srivastava(1996), “ Web Mining : Pattern Discovery from World Wide Web Transactions,” Technical Report 96-050, Dept. Computer Science University of Minnesota, Minneapolis.