

브랜드 블로그를 위한 온라인 콘텐츠 분석 시스템

인주호^o 이성렬 김명규 채수환

한국항공대, (주)이엠넷

allpoyou@kau.ac.kr edutek61@emnet.co.kr kimmk@kau.ac.kr chae@kau.ac.kr

Online Content Analysis System For Brand Blog.

Joo-Ho In^o, Sung-Youl Lee, Myung-kyu kim, Soo-hoan Chae

Dept of Computer Engineering-Korea Aerospace University, Emnet Inc.

요 약

인터넷진흥원의 자료에 따르면 2009년 5월을 기점으로 전체가구의 80%가 인터넷을 사용하고 있으며 이 중 만 3세 이상의 사용자들도 77.2%에 달한다. 또한 전체 사용자 중 블로그 서비스 사용률은 59.7%에 달하고 있다. 블로그와 같은 온라인 커뮤니케이션 서비스의 확대는 생산적 소비자의 확대를 가져왔고 동시에 블로그를 기업 마케팅의 전략 창구로서 변화 시키고 있다. 기업이 직접 블로그를 운영하게 될 때는 콘텐츠를 블로그에 게시하는 것만으로 끝나지 않는다. 언제, 어디서, 어떻게 콘텐츠가 읽혀졌는지, 콘텐츠가 기업의 브랜드 이미지에 어떤 영향을 가지는가 등의 콘텐츠 분석 또한 필수적인 요소가 되는 것이다. 이에 본 논문에서는 브랜드 블로그의 콘텐츠를 분석하기 위한 분석 시스템을 개발하고 이를 적용시켜 브랜드 블로그 상의 콘텐츠 추적 결과를 논해보고자 한다.

1. 서 론

대한민국의 인터넷 사용률은 2009년 5월을 기점으로 전체가구의 80%가 인터넷을 사용하고 있으며 이 중 만 3세 이상의 사용자들도 77.2%에 달한다고 한다[1]. 또한 인터넷 커뮤니케이션 서비스의 이용 형태가 비약적으로 증대하고 있으며 실제 전체 사용자 중 블로그 사용자는 59.7%에 달하는 것으로 나타났다[1].

블로그는 역사가 짧음에도 불구하고 이용 범위나 이용자 수에 있어서 빠른 속도로 증가하고 있는 인터넷 매체로서 다양한 분야의 기업들에 의해 마케팅 활용이 시도되고 있는 새로운 인터넷 마케팅 수단으로 발전하고 있다(황자신 외 2인, 2006). 특히 온라인 커뮤니케이션의 발달과 함께 나타난 생산적 소비자(prosumer)의 등장은 과거 기업 중심의 마케팅에서 구전(word-of-mouth)을 중심으로 한 고객중심주의 마케팅 변화를 이끌었고, 이는 기업 마케팅의 전략적 창구로서 브랜드 커뮤니티의 활성화를 이끌었으며, 브랜드와 고객과의 새로운 관계 구축과 같은 브랜드를 사용하는 고객들 간의 커뮤니케이션을 만들었다[2]. 이와 같은 추세는 브랜드 블로그라는 개인이 아닌 기업이 운영하는 블로그의 증대로 나타나고 있다. 동시에 블로그를 마케팅의 수단으로 활용하면서 블로그 콘텐츠의 정보전달력을 측정하기 위한 이슈 또한 증가하고 있는 실정이다.

따라서 본 논문에서는 브랜드 블로그에 게시되는 콘텐츠의 분석시스템을 개발하고 이를 토대로 콘텐츠의 데이터를 분석하고자 한다.

2. 관련연구

2.1 블로그 개관

과거와 달리 정보 전달의 방향에는 많은 변화가 일어나고 있다. 도서관을 중심으로 한 소장위주의 방식에서, 개방된 네트워크에서 이용자들이 검색과 링크를 통해 정보에 접근하는 방식을 거쳐, 양방향성을 활용한 활발한 의사소통을 통해 정보에 대한 반응을 즉각적으로 얻을 수 있는 방향으로 변화하고 있는 것이다[4]. 또한 이러한 양방향성과 상호작용성은 기존 미디어 산업의 정보생산과 소비에서 나타나던 일방향성까지 무너뜨려 이용자들의 적극적인 선별, 평가가 또래집단(peer group)간 정보 확산에 영향을 미치고 있다[5]. 이 같은 정보전달방향의 변화는 개인의 정보생산능력 및 확산기회의 증가를 가져왔고 정보전달 플랫폼 다변화의 주된 배경이 되고 있다[4]. 이런 변화 속에서 블로그 이용수의 증가는 열린 공간과 참여를 핵심으로 하는 양방향성 정보전달 변화의 대표적인 예라고 할 수 있다.

블로그가 사업의 방향을 변화시킬 것이라 예상[10]한 것처럼 블로그는 개인의 경험, 지식 등을 인터넷상에 간단하게 기록하는 인터넷 일기장 형태에서 변화하여 현재 뉴스를 포함하는 매스미디어를 대체할 수 있는 효과적인 수단으로 떠오르고 있으며[4], 기업들이 직접 블로그를 운영하는 브랜드 블로그들 또한 늘어나고 있는 추세이다 [6].

블로그를 정보의 개방성과 전문성을 기준으로 구분하면 그림 1과 같이 네 가지 형태로 유형화 하는 것이 가능하다[4].

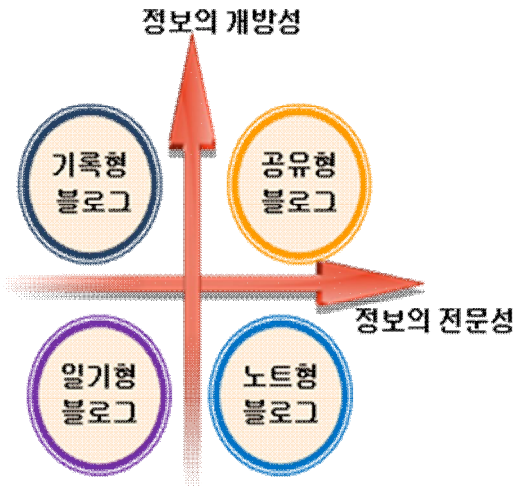


그림 1. 정보흐름의 관점으로 본 블로그 유형[4]

그림 1을 토대로 정보의 개방성과 전문성 여부를 통해 정보 네트워크로서 블로그의 본질에 가장 근접한 유형은 기록형과 공유형 블로그로 볼 수 있으며 최근의 브랜드 블로그들 또한 이 두 가지 타입으로서 만들어지고 있다.

본 논문에서는 온라인콘텐츠 분석시스템의 대상으로 이 두 가지 형태의 블로그 콘텐츠를 추적하는 실험을 하고자 한다

3. 시스템 구성

3.1 시스템 구성

다수의 브랜드 블로그 콘텐츠를 개별적으로 관리하며 콘텐츠의 상황을 추적한다는 것은 현실적으로 매우 어려운 일이다. 많은 시간과 인력이 필요로 되어지기 때문이다. 또한 콘텐츠로의 접근 경로, 실제 콘텐츠를 스크랩해간 사용자를 통해서 콘텐츠가 읽혀지고 있는가에 대한 정보는 블로그 서비스를 운영하는 매체사가 제공해 주지 않는 한 개인이 알아내기 불가능한 정보이기도 하다. 해서 본 논문에서는 효과적인 온라인 콘텐츠를 분석하기 위해 다음과 같은 분석시스템을 구현했다.

그림2는 온라인 구전 콘텐츠를 분석하기 위한 시스템 구성도이다.

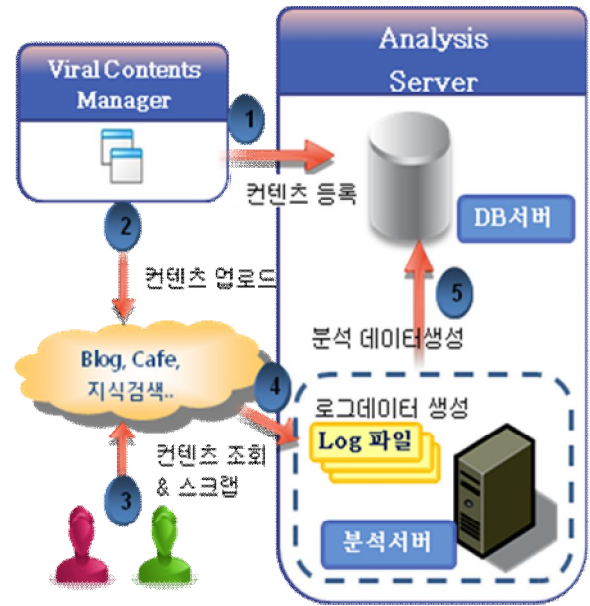


그림 2. 온라인 콘텐츠 분석 시스템

본 논문에서 구현한 브랜드 블로그 콘텐츠 분석시스템은 콘텐츠와 포스트콘텐츠의 개념을 기본으로 하고 있다. 최초 블로그 등에 게재하기 위해 제작된 콘텐츠는 임의의 블로그에 게시하기 전 단계의 데이터 이므로 이것을 콘텐츠라 명하고 블로그에 온라인상태로 게시된 콘텐츠는 포스트콘텐츠로 명명했다. 콘텐츠는 복수개의 포스트 콘텐츠로 나뉠 수 있는 것이다.

시스템은 크게 분석서버, 분석자료 보관 DB서버, 콘텐츠 관리 매니저로 나뉘며 각각의 기능은 다음과 같다.

- 분석서버: 블로그 콘텐츠의 정보 조회가 있을 경우 해당 요청은 분석서버의 로그 파일로 들어옴. 이렇게 쌓여진 로그 정보를 분석 모듈을 통해 DB서버로 저장한다.
- DB서버: 분석서버를 통해서 분석된 데이터를 보관.
- 콘텐츠 관리 매니저: 매니저 프로그램을 통해 콘텐츠 분석에 필요한 추적코드 등을 생성하고 이를 토대로 실제 블로그에 올라가는 콘텐츠에 ID를 부여하여 관리한다. 또한 DB서버에 저장되어 있는 데이터를 불러와 현재 분석 중인 콘텐츠의 상황(조회수, 확산건수, 콘텐츠 조회 검색어 등)을 확인 할 수가 있다.

3.2 분석 시스템 데이터 항목 정의

위의 시스템 구성을 통해 온라인 콘텐츠에 대해서 다음의 항목을 데이터화 해서 추적할 수 있다.

표 1 추적데이터 항목

본문 조회수	스크랩 수	확산 건수	확산 조회수	총 조회수	유입 검색어
-----------	----------	----------	-----------	----------	-----------

각 항목은 다음의 의미를 지닌다.

- **본문 조회수:** 원본 온라인콘텐츠가 읽혀진 회수
- **스크랩 수:** 다른 사용자가 원본 콘텐츠를 자신의 블로그 등으로 복사해간 회수로서 블로그 시스템을 운영하는 매체에서 제공하는 데이터.
- **확산 건수:** 원본 온라인콘텐츠가 다른 사용자에게 의해 옮겨진 회수
- **확산조회수:** 다른 사용자에게 스크랩된 콘텐츠가 조회된 회수
- **총 조회수:** 본문조회수와 확산 조회수의 합
- **유입 검색어:** 검색 매체를 통해서 콘텐츠로 유입 될 경우에 해당 검색어.

3.3 분석서버 시스템

그림 3은 분석서버 시스템의 구성도이다. 분석서버는 분석 모듈을 통해서 각각의 온라인 콘텐츠로부터 생성된 콘텐츠와 포스트콘텐츠의 데이터를 분석해낸다.

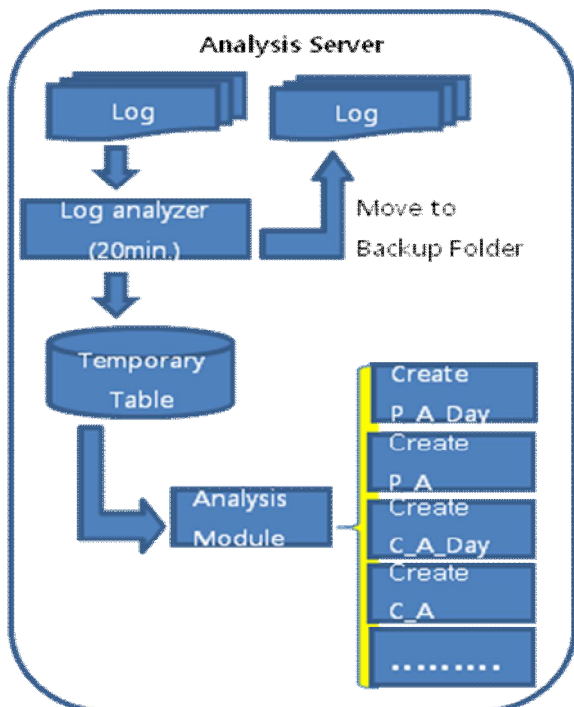


그림 3. 분석서버 시스템 구성도

브랜드 블로그에 게시된 온라인 콘텐츠의 확산 과정을 추적하기 위한 시스템 모듈은 그림 4의 방식으로 동작한다.

```

//임시 테이블로부터 포스트콘텐츠 ID, referrer 정보를 읽어 들임.
Sql1 = "SELECT 포스트콘텐츠 ID, 유입 referrer FROM Temporary Table";

//유입 referrer정보로부터 다음의 정보를 추출해냄.
저장변수 [] = 변수 추출 클래스(매체정보, 상품정보, 상품에 따른 사용자 ID);

//읽어 들인 포스트콘텐츠 ID 를 통해 확산 체크를 위한 설정값 로딩.
Sql2= "SELECT 매체정보, 상품정보, 트랜스 ID FROM 확산정보데이터 WHERE 포스트콘텐츠 ID= sql1, 포스트콘텐츠 ID";
IF (Sql2 == null)
{
    Sql1, 포스트콘텐츠 ID에 대한 확산 정보가 설정되어 있지 않은 경우.
    Sql1, 포스트콘텐츠 ID의 본문 조회수를 증가시킨다.
} ELSE
{
    IF (저장변수 []의 값 == Sql2, 매체정보, 상품정보, 트랜스 ID )
    {
        Sql1, 포스트콘텐츠 ID에 대한 정보는 본문 조회로 인해 발생한 정보임.
        Sql1, 포스트콘텐츠 ID의 본문 조회수를 증가시킨다.
    } ELSE
    {
        Sql1, 포스트콘텐츠 ID의 정보는 다른 사용자에서 조회가 발생.
        Sql1, 포스트콘텐츠 ID의 확산 조회수를 증가시킨다.
        Sql1, 유입 referrer 정보를 저장.
    }
}
    
```

그림 4 콘텐츠 확산 판단 과정

블로그에 게시되는 포스트콘텐츠들은 해당 블로그의 사용자 ID와 개별 포스트콘텐츠를 구분하기 위한 포스트콘텐츠ID를 가진다. 이 정보를 기반으로 분석시스템은 분석서버로 유입된 데이터로부터 블로그 ID, 매체정보, 상품정보, 포스트콘텐츠ID 값을 추출 후 이를 저장되어 있는 정보와 비교하여 해당 콘텐츠가 올린 사람의 블로그에서 읽혔는지 아니면 다른 사용자에게 옮겨진 후 읽혔는지를 구분해 낼 수 있다.

4 실험 방법 및 분석

4.1 실험 방법

온라인 콘텐츠 분석 시스템의 실험 대상으로 스포츠, 교육, 쇼핑물 부분에서 운영중인 브랜드 블로그를 대상으로 데이터를 수집했다. 또한 테스트 콘텐츠는 직접 업체의 상품과 관련 있는 내용을 다른 관련콘텐츠와 온/오프라인상에서 이슈내용을 다른 이슈 콘텐츠로 구분해서 관리했으며 본문스크랩이 가능하게 허용했다.

표 2는 분석시스템을 적용해서 운영중인 전체 브랜드 블로그(80개) 중 대표 세 개 업체의 이슈콘텐츠, 관련 콘텐츠 각 1개의 추적 결과이다.

표 2. 대표 사례를 통한 콘텐츠 추적 데이터

업종	구분	콘텐츠 내용	전체조회수 (등록 일의 조회수)	댓글 수	스크랩 수	확산 건수	확산 조회수
스포츠	이슈	축구 선수 골 동영상	2506(2324)	2	0	0	0
스포츠	관련	차사의 기능성 야구용품 설명	205(53)	0	1	1	2
교육	이슈	개봉영화 리뷰	2736(548)	4	1	1	796
교육	관련	해당 기관 신입생모집 설명	504(10)	0	0	0	0
쇼핑몰	이슈	가상스타커플의 결혼이야기	3582(1665)	9	5	0	0
쇼핑몰	관련	제품 비교	837(216)	0	3	1	1

표 3. 대표 사례를 통한 콘텐츠 유입 검색어 관련 데이터

업종	구분	검색회수	검색어 수 (일별중복 제거)	검색매체(일별중복 제거)				
				네이버	다음	야후	구글	네이트
스포츠	이슈	2446	117	116	0	1	0	0
스포츠	관련	413	200	188	2	3	2	5
교육	이슈	1817	827	750	47	28	0	2
교육	관련	97	76	71	3	2	0	0
쇼핑몰	이슈	3349	208	166	42	0	0	0
쇼핑몰	관련	336	253	242	11	0	0	0

4.2 실험 결과 분석

표2를 통해서 다음 몇 가지의 특이 사항을 발견 할 수 있었다.

1. 이슈 콘텐츠는 관련 콘텐츠 보다 일반적으로 더 많은 조회수를 기록하고 있으며 특히 등록 당일 날의 조회수가 관련 콘텐츠 보다 많이 조회 되고 있음을 알 수 있다.
2. 이슈 콘텐츠와 같은 공동 화제에 대해서는 방문자들이 댓글을 통한 공감을 표현하는 반면 사이트와 관련 있는 제품정보, 업체홍보 등의 콘텐츠에는 댓글을 거의 달지 않음을 확인 할 수 있다.
3. 블로그 매체에서 집계되는 게시물의 스크랩 수와 추적 시스템을 통해 조사한 확산 건수는 차이가 나타난다. 콘텐츠를 스크랩해간 사용자의 블로그에서 콘텐츠가 워힐 때 해당 콘텐츠에 삽입된 추적 태그를 통해서 분석서버로 정보를 보내오는 시스템의 방식을 고려한다면 해당 콘텐츠는 다른 사용자에게 스크랩 되기만 하고 그 사용자의 블로그를 통해서만 읽혀 지지 않고 있다는 것으로 분석을 할 수 가 있다. 이런 현상은 해당 블로그에 접근하기 위해 사용자가 주로 검색매체를 활용하는데 이때 원본 콘텐츠가 올라와있는 브랜드 블로그 사이트들이 검색매체에서 검색상위에 노출되기 때문으로 해석할 수가 있다.
4. 콘텐츠가 조회된 회수와 검색어를 통해 유입된 수치가 비슷한 정도를 가지고 있다. 이는 대부분의 사용자가 검색매체를 통해서 콘텐츠에 유입하는 것으로 분석 할 수 있다. 또한 3번의 내용과

비교해 볼 때 사용자들이 검색된 결과 중 상위의 검색 결과로 먼저 유입함으로써 브랜드 블로그에서 스크랩해간 사용자들의 사이트보다 브랜드 블로그 쪽의 콘텐츠로 유입되는 것으로 분석 할 수가 있을 것이다.

이 상의 실험 데이터를 통해서 단순히 블로그 매체에서 제공하는 수치 데이터보다 더 많은 사실을 유추 해 낼 수가 있다. 콘텐츠로 유입되는 대부분의 방문자들이 검색매체에서 검색어를 이용한다는 점, 콘텐츠 스크랩 수치가 실제 콘텐츠의 확산을 의미하진 않는다는 점 등 이다.

5. 결론

브랜드 블로그에서 상호작용성을 크게 느낄수록 이용자들의 브랜드 블로그에 대하여 긍정적인 태도를 형성한다는 연구결과[7]는 많은 점을 시사한다. 블로그 운영에 있어서 방문자들을 이끌 수 있는 정책이 필요하다는 것이다. 본 논문에서 구현한 온라인 콘텐츠 분석 시스템은 사용자의 유입을 이끌 수 있는 수단을 제공 하는 것은 아니다. 하지만 분석 된 데이터를 통해서 블로그 매체에서 제공하는 단순 수치 이상의 내용을 알아 낼 수 있음은 실험을 통해 보여주었다. 이런 데이터를 토대로 브랜드 블로그의 운영 방향에 지표가 될 수 있을 것이라 본다.

향 후 과제는 블로그 콘텐츠의 댓글 분석을 통해 블로그에 대한 이용자의 반응을 분석하는 추가 방안에 대해 연구가 진행 되어야 할 것이다.

- 이 논문은 ㈜이엠넷의 주관으로 연구 개발된 논문임.

참고문헌

[1] 한국인터넷 진흥원(2009) (인터넷이용실태조사 최종보고서)

[2] 황장신·김은혜·조정식(2006). 웹사이트에 대한 태도에 영향을 미치는 요인으로서의 인지된 상호작용성, 인터넷 이용 동기 및 관여도 (한국광고홍보학회 8(1):159-186)

[3] McWilliam, Gil (2000). Building Stronger Brands Through Online Communities. Sloan Management, 20:197-208

[4] 강선미, 김영기 (2009). 정보흐름의 관점에서 본 블로그 연구(인문학논총 14(2) 147~171)

[5] 이호영,정은희 (2008). 블로그를 중심으로 본 디지털 콘텐츠의 사회적 확산(KISDI 이슈 리포트, 정보통신정책연구원, 8(12))

[6] 김윤식,정규엽 (2009). 호텔브랜드 블로그 특성이 구매의도 및 온라인 구전 커뮤니케이션에 미치는 영향: 블로그 태도를 매개변수로 (호텔경영학연구 18(3):1~22)

[7] 김용현(2005). 블로그 마케팅의 효과에 관한 연구. 고려대학교 언론대학원, 석사학위 논문

[8] 김지영, 윤영민(2010), 온라인 구전 커뮤니케이션. 사용 후기 평가에 미치는 댓글의 영향에 관한 실험 연구(한국방송학보 24-1)

[9] Dr. Ralph F. Wilson(2005), The Six Simple Principles of Viral Marketing

[10] Business Week(2005), "Blogs Will Change Your Business"(May 2005).

[11] 안경환, 곽봉철(2007), 디지털 미디어의 상호작용 WOM 마케팅 효과에 관한 연구 (韓國컴퓨터情報學會論文誌, Vol.12 No.2, [2007])

[12] 김영철, 강명구(2008), 웹로그 분석방식의 시스템 확장성 연구(전자상거래학회지, Vol.9 No.2, [2008])

[13] 박소연, 이준호(2002), 로그 분석을 통한 이용자의 웹 문서 검색 행태에 관한 연구. (情報管理學會誌, Vol.19 No.3, [2002])

[14] 전주호, 구회군, 최병선, 이원구, 이재광(2001). 리눅스 기반 통합 로그 분석도구의 설계 및 구현. (한국인터넷정보학회 학술발표대회 논문집, Vol.2 No.2, [2001])

[15] 정강용, 박나연(2007), 웹서버의 로그파일 분석에 의한 웹 서비스 활용에 관한 연구. (韓國컴퓨터情報學會論文誌, Vol.5 No.1, KCI등재[2007])