

웹로그 분석을 적용한 웹사이트내의 웹컨텐츠 분석 연구 결과

정선경¹, 이철우²
삼성SDS¹, 전남대학교²
hera0107@orgio.net¹, leecw@chonnam.ac.kr²

Web log Data Analysis Apply to Web Contents Analysis Result data

Sun - Kyeong Jeong¹, Chill - Woo Lee²
Samsung SDS¹, Chonnam National University²

요 약

본 논문은 웹사이트를 구성하는 웹컨텐츠의 함목적성 및 사용성 평가를 위하여 웹사이트 사용자들의 흔적인 웹로그 분석 데이터를 적용하여 정량적인 평가를 한 결과에 대하여 보고하고자 한다. 웹로그파일은 사용자가 웹사이트를 이용하면 이에 대한 기록이 로그라는 형태로 흔적이 남는다. 로그분석이란 이 데이터를 기반으로 위에서 말한 다양한 정보를 추출해 내는 것이라 할 수 있다. 로그분석은 사용자에 따라 단지 로그 정보를 분석하는 것에 한정시키기도, 로그 정보를 기반으로 한 보다 다양한 정보를 분석하는 확장된 개념으로 확대시키기도 한다. 본 논문은 웹트렌즈[로그분석 툴]를 사용하여 웹사이트내의 웹컨텐츠 분석에 웹 로그 분석 결과가 같은 정량적인 개념을 활용하여 정성적인 분석으로 일관되었던 웹사이트 분석에 새로운 결과를 도출하였다. 또한 마케팅을 수행하는 데 있어서 고객의 요구에 철저히 대응하기 위해서 고객을 철저히 파악하여야 한다. 로그파일 분석을 통해서 주요 고객층, 고객의 구매 패턴, 주 구매시간, 구매탐색 경로등의 데이터를 추출할 수 있다. 로그파일 분석 데이터를 기반으로 인터페이스 설계나 상품의 레이아웃등의 설계, 고객 서비스 강화등의 다양한 대 고객 마케팅을 펼칠 수 있다. 본 논문은 단순 수치해석의 개념이었던 로그 분석에 웹컨텐츠 분석 기법을 접목하여 웹사이트내에서 사용되고 있는 컨텐츠의 사용성에 대한 연구 결과에 대해서 발표하고자 한다.

1. 서 론

인터넷 비즈니스가 지난 가장 큰 특징 중의 하나는 바로 고객의 관심과 행동 양식을 파악한 가운데 전개될 수 있다는 점이다. 특정 고객이 방문경로 및 구매 행동에 대한 수치적 결과 데이터는 향후 고객이 다시 방문했을 때 실제 구매를 일으킬 수 있도록 유도할 수도 있다. 사이트에 접속해서 나가는 동안의 모든 행위를 파악할 수 있다는 것이다.

분석 자료는 그 활용도에 따라서 인터넷 비즈니스에 가장 중요한 성공 열쇠가 될 수 있다. 이러한 분석을 위한 여러가지 로그 분석 툴을 사용한다. 고객의 정보는 입력된 IP 정보를 기반으로 웹사이트 내에서 이루어진 컨텐츠의 사용 및 획득에 대한 데이터를 *.log로 저장하고, 이 데이터를 분석 툴을 이용하여 WebLog 분석을 활용한 컨설턴트들이 효율적으로 분석할 수 있는 가공 데이터를 제공한다.

본 논문은 이러한 로그분석을 적용하여 웹사이트를 평가하고 웹사이트 내의 컨텐츠 설계에 이 분석 데이터를 적용하여 보다 합리적이고 효율적인 웹사이트 시스템을 구축하고자 한다. 이제 고객의 정보는 단순한 Log에서 벗어나 분석이 가능한 데이터로 가공되었다.

다양한 형태 포맷으로 보고서를 출력함으로써 자료 작성에 편리하다. 또한 다양한 포맷의 로그파일의 분석을 지원함으로써 별도의 로그파일에 대한 정제 작업이 필요하지 않다..

2. WebLog 분석 툴의 종류

2.1 WebLog [www.weblog.com]

2.1.1 특징

웹로그는 서버에 직접 셋팅되어야 하는 분석툴이다. 서버에 직접 해당 프로그램을 설치하여야 하는 이유로 설치시 난이도가 있다. 국내에서 개발된 프로그램으로 구체적인 지역 정보를 제공하고 있어 IP에 대한 지역 정보를 획득하기 유리한 툴이다.

2.1.2 장점

웹로그는 단일 도메인에 복수의 서버가 연결된 경우 분석이 가능하다. 웹서버에 프로그램상의 로드가 적어 웹서버의 성능에 적은 영향을 미친다. 시도단위의 지역 정보에 대한 상세한 결과 출력이 가능하다.

2.1.3 단점

웹로그는 웹서버에만 셋팅이 가능하므로 클라이언트 컴퓨터등에서 로그만의 분석이 불가능하여 자유로운 분석 환경을 제공하지 못하고 있다.

2.2 WebTrens [www.webtrends.com]

2.2.1 특징

웹트렌즈는 웹서버에 클라이언트 컴퓨터에 동시에 설치가 가능하며, 기존의 프로그램의 셋팅처럼 설치 난이도가 높지 않다.

2.2.2 장점

웹트랜즈는 단일 도메인에 복수의 서버가 연결된 경우 분석이 가능하다. 분석하고자 하는 항목의 지정과 그 항목의 하위 구성 요소의 지정이 가능하여 원하는 형태의 결과를 출력이 가능하다. 웹사이트를 구성하는 Global menu와 Local menu의 노출에 대한 분석이 가능하며 원격으로 파일 전송이 가능하다. 콘텐츠의 활용 및 사용자의 웹사이트로의 접속에 대한 정보인 Reference & Keywords의 획득이 가능하다.

2.2.3 단점

웹트랜즈는 국외프로그램인 관계로 IP 정보에 대한 해석 능력이 떨어진다. IP는 지역적, 조직의 형태별로 고유한 값임에도 불구하고, 이러한 정보의 해석에 오류가 발생하기도 한다.

3. 웹로그를 적용한 웹컨텐츠 분석 방법론

본 논문은 웹사이트를 구성하는 가장 기본이 되는 정보인 컨텐트의 효율적인 사용에 대해서 분석하는 방법에 대해서 제시하고자 한다.

3.1 접근성 [Accessibility]

모든 사용자들이 물리적 환경에 구애받지 않고 웹사이트를 볼 수 있다는 것은 매우 중요하다. 웹 브라우저와 하드웨어 성능의 차이에 따라 어떤 사용자는 한정된 이용만 허용될 수도 있기 때문이다. 웹 사이트를 보고자 하는 사람은 누구나 볼 수 있어야 한다. 그러나 모든 사용자가 물리적으로 꼭 같은 환경 즉 꼭 같은 소프트웨어와 하드웨어를 가지고 있지는 않기 때문에 접근성을 세심하게 배려하지 않은 사이트를 보는데 장애를 느낄 수 있다. 다음의 리스트는 서로 다른 물리적 능력에 기인하는 접근성의 이슈 들이다.

3.2 성능 [Performance]

사용자들은 다양한 접속환경과 다양한 하드웨어 성능을 가지고 사이트에 접속한다. 많은 그래픽과 한 페이지에 집중된 지나치게 많은 정보는 사용자가 페이지를 다룬 받는 시간을 지연시킬 것이다. 사용자들은 20에서 30초가 걸리는 시간을 기다려 하지 않거나 주의가 분산될 것이다. 사용자는 사이트에서 필요로 하는 정보를 빨리 얻을 수 있도록 하는 것이 중요하다. 여기서는 많은 지연 요소로 인하여 사용성이 떨어지는 것을 막기 위한 모든 배려가 이루어 졌는지를 검사한다

3.3 네비게이션 [Navigation]

네비게이션 테스트는 최종적인 사용자 테스트와 병행하여 수행할 수 있다. 하지만 모든 소프트웨어 시험에서 그러하듯이 커버리지 테스트가 중요하다. 따라서 웹사이트 상의 모든 가능한 네비게이션 경로를 분석하고 하나도 빠짐없이 검사하여야 한다. 이때 패스 리버스도구를 이용하는 것이 유용하다. 많은 복잡한 사이트에서도 사용자 중심의 사이트맵이 제공되지 않는다. 따라서 사용자들은 사용하면서 자신이 이해한 앱을 만들어야 한다. 링크는 어떤 내용으로 연결 되는지 잘 설명되어 있어야 하며, 어떻게 이전에 방문했던 곳으로 갈 수 있는지도 잘 보여야 한다

3.4 유지보수성 [Maintainability]

웹사이트를 개발하는 것 뿐만 아니라 지속적으로 유지 관리하는 것도 중요하다. 웹사이트를 설계하는 사람은 웹사이트의 생명이 다할 때 까지 웹 사이트를 쉽게 확장하고 유지할 수 있는지를 고려하여야 한다. 웹사이트 유지의 중요 이슈 중의 하나는 얼마나 쉽게 다른 서버로 이식할 수 있는냐 하는 것이다. 사이트 내에서 상대적인 링크를 사용하면 이식이 쉬워진다.

또 다른 이슈는 자주 업데이트 되는 부분을 쉽게 교체할 수 있도록 적당한 용어로 분리되어 있는가 하는 것이다. 즉 적당한 수준에서 쉽게 업데이트 할 수 있도록 구성하는 것이 중요하다. 웹사이트의 확장은 설계 단계에서부터 세심하게 고려하여야 한다. 이미지와 아이콘을 디자인 할 때부터 새로운 이미지나 아이콘을 쉽게 바꾸거나 대체할 수 있도록 고려하여야 한다. 만약 정보의 시간성이 중요하다면 모든 페이지에 최근 업데이트 날짜를 반드시 명기하여야 한다. 또한 접촉할 사람의 전화와 이메일을 명시하여야 한다. 만약 주기적으로 정보를 갱신하거나 홈페이지를 개편한다면 이러한 일정도 공시하는 것이 좋다.

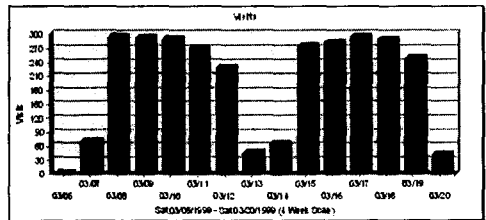
3.5 가독성 [Readability]

사용자가 산만함 없이 페이지상의 중요한 정보를 쉽게 읽을 수 있어야 한다. 그래픽이나 애니메이션이 흥미를 유발하지만 좋은 것도 너무 많으면 산만하기 쉽다. 또한 한 페이지 이상의 정보도 너무 많으면 사용자가 여러 번 스크롤해야 한다. 웹의 기능성은 하루가 다르게 좋아지고 있다. 페이지의 정보 밀도도 페이지의 가독성에 영향을 미친다. 인쇄매체의 경우에도 페이지의 여백과 강조 부분이 중요하듯이 백색 공간과 강조 확대 글씨체 부분이 중요하게 이용된다. 링크는 웹 페이지에 있어서 중요한 요소중의 하나이고 이를 잘 보이게 하는 것도 필요하다. 하지만 너무나 많은 링크와 기본 설정보다 작은 글씨들로 채워진 페이지는 점점 읽기 힘들게 만든다. 가독성은 결국 사용자들 통한 주관적이고 정성적인 평가 결과가 많이 필요한 부분이다. 따라서 가독성의 기술적 체크 부분은 단지 참고 사항으로 하여도 좋다.

4. 웹트랜즈 실험 결과를 사용한 웹컨텐츠 분석 결과

본 논문의 실험 결과는 웹트랜즈를 이용하여 웹서버에 저장된 로그를 프로그램상에서 실험하여 그 결과를 웹컨텐츠 분석 자료의 정량적인 데이터로 적용하였다.

4.1 웹사이트의 전반적인 상황 해석을 위한 General Statistic

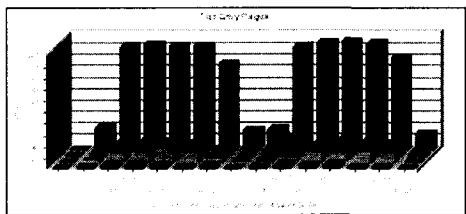


General Statistics		Report Range: 01/06/1999 00:00:00 - 03/20/1999 21:59:59
Hits:	1	26,700
Average per Day:	1	1,780
Home Page:	1	2,840
Page Views:	1	25,285
Average per Day:	1	1,665
Average per Unique Visitor:	1	12
Duplicated Views:	1	10,012
Visits:	1	3,022
Average per Day:	1	201
Average Visit Length:	1	00:31:13
Median Visit Length:	1	00:26:50
International Visits:	1	0%
Visits of Unknown Origin:	1	100%
Visits from United States:	1	0%
Visits Referred by Search Engine:	1	35
Visits from Spiders:	1	120
Unique Visitors:	1	2,044
Visitors Who Visited Once:	1	1,701
Visitors Who Visited More Than Once:	1	674

[General Statistic]

General Statistic은 웹사이트의 전반적인 상황 해석에 필요한 Hits / Page Views / Visits: 웹사이트의 평균 사용 시간 / Visitors: 2회이상 방문객의 정보로 웹사이트에 대한 충성도 해석을 위한 기초 데이터이다.

4.2 웹사이트내에서 가장 노출횟수가 많은 콘텐츠 해석을 위한 Top Entry Pages

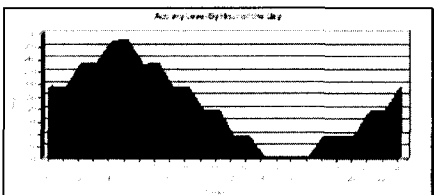


Page	Visit Total	Visits
Welcome to the Oregon Marine	15,725	14,012
Comments and Feedback on Oregon Marine	15,015	14,012
Home - Renewable in Oregon Marine	15,015	13,970
Customer Service and Support	13,451	13,451
Oregon Marine Customer Service and Support	13,451	13,451
Oregon Marine Sewer Homes	12,945	12,945
Latest News - Oregon Clean Energy	11,842	11,842
Home - Renewable	11,842	11,842
Customer Service and Support	11,842	11,842
Oregon Marine Customer Service and Support	11,842	11,842
Renewable Sewer Support	11,842	11,842
Sewer Service Support	11,842	11,842
Home - Renewable	11,842	11,842
Home - Renewable	11,842	11,842
Customer Service and Support	11,842	11,842
Customer Service and Support	11,842	11,842

[Top Entry Pages]

웹사이트내에서 가장 노출 횟수가 많다는 것은 사용자에게 가장 많은 클릭을 받았다는 결과가 된다. 이는 웹사이트를 기획하고 개발한 담당자들의 의도와 일치 할 경우 웹사이트 정보에 대한 접근성과 가독성 대비게이션에서 높은 점수를 얻게 된다. 이는 웹사이트를 설계하는 가장 중요한 정보가 되며, 추후 웹사이트를 Renewal 하고자 하는 경우 콘텐츠설계에 중요한 데이터로 활용할 수 있다.

4.3 1 일 기준으로하여 웹서버에 접속한 사용자의 변화량에 대한 결과 데이터 Activity Level by Hour of the Day

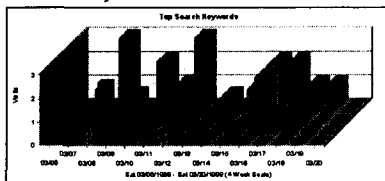


Hour	# of Hits	% of Total Hits	# of Visits
00:00 - 01:00	1,120	0.8%	1,120
01:00 - 02:00	1,120	0.8%	1,120
02:00 - 03:00	1,120	0.8%	1,120
03:00 - 04:00	1,120	0.8%	1,120
04:00 - 05:00	1,120	0.8%	1,120
05:00 - 06:00	1,120	0.8%	1,120
06:00 - 07:00	1,120	0.8%	1,120
07:00 - 08:00	1,120	0.8%	1,120
08:00 - 09:00	1,120	0.8%	1,120
09:00 - 10:00	1,120	0.8%	1,120
10:00 - 11:00	1,120	0.8%	1,120
11:00 - 12:00	1,120	0.8%	1,120
12:00 - 13:00	1,120	0.8%	1,120
13:00 - 14:00	1,120	0.8%	1,120
14:00 - 15:00	1,120	0.8%	1,120
15:00 - 16:00	1,120	0.8%	1,120
16:00 - 17:00	1,120	0.8%	1,120
17:00 - 18:00	1,120	0.8%	1,120
18:00 - 19:00	1,120	0.8%	1,120
19:00 - 20:00	1,120	0.8%	1,120
20:00 - 21:00	1,120	0.8%	1,120
21:00 - 22:00	1,120	0.8%	1,120
22:00 - 23:00	1,120	0.8%	1,120
Total Visits during Week Hours (8:00am-9:00pm)	1,120	0.8%	1,120
Total Visits during 24-Hour Operation	21,084	75.4%	21,084

[Activity Level by Hour of the Day]

웹사이트는 1일 24시간동안 무제한적으로 사용자에게 노출되어 있다. 그러나, 프로모션을 하거나 웹서버의 성능 향상을 위한 별도의 조치를 취하고자 할 때 가장 안정적인 시간대에 작업하는 계획이 필요하다. 웹로그 분석 데이터 중에 1일/1주일/1개월 사용 시간에 대한 정보는 프로모션과 시스템 관리에 중요한 정보가 된다.

4.4 Top Search Keywords



Keywords	Keywords	% of Total
renewable	14	18.18%
sewer	14	18.18%
clean	14	18.18%
energy	14	18.18%
oregon	14	18.18%
marine	14	18.18%
customer	14	18.18%
support	14	18.18%
home	14	18.18%

[Top Search Keywords]

웹사이트가 특정 검색엔진들에서 어떤 키워드가 사용되어 검색되어 지는가에 대한 연구 결과이다. 검색엔진에서 사용자가 희망하는 키워드 입력 후 해당 사이트로 이동 할 경우 웹로그는 검색엔진의 명칭과 검색 키워드를 기억하여 통계 데이터를 뿌려줌으로 인하여 웹사이트 관리자는 해당 사이트를 일반인이 어떤 키워드를 사용하여 접속하는 지에 대한 정보를 획득하여 웹사이트 등록시 중요한 정보로 활용할 수 있다.

5. 결론 및 추후 연구

웹사이트 평가와 인터넷 비즈니스평가가 구분되지 않을 정도로 웹사이트 평가기준이 바뀌었다. 이런 배경에는 첫번째는 웹사이트 평가기준에 참조하는 성공한 웹사이트를 수익을 올린 사이트, 기업가치를 결정적으로 올린 사이트에 국한하였기 때문이다. 따라서 기존의 웹사이트 평가 요소들인 디자인, 성능, 기술, 콘텐츠 뿐만 아니라 소비자 보호, 사업전략, 서비스(오프라인포함)수준, 방문자 수 중은 회원 수, 매출 등의 비즈니스적인 평가 요소들이 추가되고 비중이 높아졌다. 다른 표현을 한다면 투자자 중심의 웹사이트 평가로 바뀐 것이다. 기존의 사용자 중심의 웹사이트 평가는 웹 사이트의 질적 향상을 유도하고 도움을 주었다. 로그분석을 활용한 웹사이트 분석은 '방문자수, 매출, 수익의 관점에서 이익의 창출여부에 대해서 평가하는 것이다. 단순히 분석 당사자의 사견이 아닌 정확한 수치 데이터에 의한 정량적인 분석으로 웹사이트를 구성하는 웹컨텐츠의 효율적인 활용 및 관리에 적용할 수 있다. 앞으로 웹로그 분석틀의 정확도와 웹로그에 보다 많은 중요한 데이터가 저장 될 수 있는 기술 발전이 필요하다.

참고 문헌

- [1] Eric L Reiss, Practical Information Architecture, ADDISON-WESLEY 2000
- [2] Jakob Nielsen & Marie Tahir, Home Page Usability, New Riders, 2001
- [3] Mai-lan Tomsen, Kiiler Contents ADDISON-WESLEY 2000
- [4] 김형택, 민욱길, 효과적인 인터넷 마케팅을 위한 웹로그 분석, 비비컴 2001
- [5] 정선경, 웹사이트 리뉴얼, 삼성멀티캠퍼스, 2002
- [6] 정선경, 웹로그 분석을 적용한 웹사이트내의 웹컨텐츠 분석 방법론, 2003 HCI 학회