

웹사이트 분석을 통한 최적화 설계 방안★

장희선*

요 약

웹3.0, 미래 인터넷 및 사물인터넷 등으로 인터넷 이용이 급증함에 따라 사용자와 웹서버들 사이의 상호 정보 교환을 통한 빅데이터들이 증가하고 있다. 이런 데이터들의 분석 결과들은 상업용 웹사이트의 경우 마케팅과 캠페인에 활용하고, 비상업용 사이트의 경우에도 사용자의 서비스 만족도 향상을 위한 기초자료로 이용된다. 본 논문에서는 웹사이트 분석을 위한 정량적 지표들을 제시하고 지표들 사이의 상관관계와 유의성 검정 분석을 통하여 최적화된 웹사이트 설계 방안을 제시한다. 138개의 웹사이트에 대한 분석 결과, 방문수와 순방문자 수, 페이지뷰 수와 평균 방문시간의 사이에는 각각 강한 양(+)의 상관관계가 존재하며, 방문당 페이지뷰 수 및 신규 방문 비율과 이탈률 사이에는 음(-)의 상관관계가 존재함을 알 수 있다. 특히, 웹사이트 방문자의 이탈률을 줄이기 위해서는 방문수와 순방문자 수를 높이기보다, 방문당 페이지뷰의 수와 신규 방문비율을 높이기 위한 전략이 요구된다.

Optimized Web Design Method by Analyzing the Websites

Hee-Seon Jang*

ABSTRACT

As the Internet usage such as Web3.0, future internet, and internet of things increases, the big data through information exchange between the users and web servers increases. Analyzing those web data, the commercial web sites use the analytic results for marketing and campaign, and non-commercial web sites also use the results to improve the user's services satisfaction. In this paper, the quantitative index is presented to analyze the web sites, and optimized web site design method is also presented through the correlation analysis of index and significance test. From the results for 138 web sites, it is observed that strong plus(+) correlation for visits-unique visitors and page views-average visit duration exists. We also observe the minus(-) correlation between bounce rate and page views per user(or ratio of new visits). In specific, to reduce the bounce rate for users, the strategy to increase the page views and ratio of new visits rather than visits and unique visitors is needed.

Keywords : Web analytics, Big data, Web measurements, Web analysis, Google analytics

접수일(2015년 2월 24일), 게재확정일(2015년 3월 20일)

* 평택대학교 컴퓨터학과

★ 이 논문은 2014학년도 평택대학교 학술연구비의 지원에 의하여 연구되었음

1. 서론

웹2.0의 개념이 정립되기도 전에 최근 웹3.0, 미래 인터넷, IoT(Inetnet of things, 사물 인터넷)라는 용어를 사용하며 인터넷 혁명의 파동에 대한 가설을 세우고 있다[4,6,14]. 그리고 웹3.0의 시대에 사용자와 웹서버 사이에서 이루어지는 빅데이터 분석을 통하여 웹사이트 설계, 온라인 마케팅 활용, 웹디자인 최적화 등을 추구하고 있다[3,11,12]. 그러나 1969년 처음 인터넷이 시작된 이래 지금까지 이용성 측면에서의 웹사이트 분석을 위해 웹로그 분석, 방문자 카운터 수 측정, 브라우저 및 운영체제 식별 등의 기본적 수준에 머무르고 있다[1,5].

본 논문에서는 최적화된 웹사이트 설계를 위해 필요한 정량적 지표들을 설정하고 138개의 웹사이트에 대하여 지표들 사이의 상관관계를 분석하며, 유의성을 검증한다. 정량적 지표는 크게 접근성, 표준성 및 구글 애널리틱스의 7가지 지표(방문수, 순방문자수, 페이지뷰수, 방문당 페이지뷰 수, 평균 방문시간, 이탈률, 신규 방문비율)를 사용한다. 특히, 상업용 사이트의 경우 웹사이트로 유입하여 들어온 사용자의 이탈률(Bounce Rate)을 줄이기 위한 주요 요인 변수를 도출함으로써 효율적인 마케팅 및 캠페인 전략 수립 방안을 제시한다.

2. 웹사이트 분석

웹사이트 분석은 ‘정량적인 인터넷 데이터를 객관적으로 트래킹(tracking), 수집(collecting), 측정(measuring), 리포팅(reporting), 분석(analyzing)을 통해 웹사이트와 웹마케팅의 효과를 최적화하는 것’으로 정의한다[8,13]. 즉, 웹사이트 분석은 운영 중인 웹사이트의 목표를 이루기 위한 데이터 분석을 의미하며, 크게 정량적(quantitative) 방법과 정성적(qualitative) 방법으로 분류된다. 정성적 방법은 설문조사(survey), 휴리스틱 평가(heuristic evaluation), 사용성 테스트(usability test), 사용자 방문 인터뷰(interview)를 포함한다[7,9]. 그리고 정량적 분석에서는 사용자 히트, 서버 웹로그, 로그 자동 파싱, 방문자 수 카운터 등을 이용

하며, 최근에는 웹비콘(web beacon), 자바스크립트 태그(javascript tag), 패킷 스니핑(packet sniffing) 방법 등을 통해 페이지뷰, 히트, 상위 종료 페이지, 웹사이트 몰입도, 화면 해상도 등을 측정한다[10,15]. 본 연구에서는 정량적 평가에 초점을 두어 <표 1>의 지표를 사용한다. 크게 접근성, 표준성과 구글 애널리틱스[15]를 활용한 지표 값을 분석하며, 세부 항목을 요약하면 다음과 같다[16,17,18].

<표 1> 정량평가 지표

구분	평가 항목	도구
접근성	인식의 용이성 운용의 용이성 이해의 용이성 기술적 진보성	K-WAH
표준성	웹표준의 준수	W3C Validator
방문수	사이트 총 방문수	Google Analytics
순방문자수	사이트에 대한 순방문자수의 합계	
페이지뷰수	총 페이지 조회수	
방문당 페이지수	Page Views/Visits	
평균 방문시간	방문자당 평균 방문시간	
이탈률	일정시간(10초) 이내로 머무른 트래픽의 비율	
신규 방문비율	처음 방문한 방문자 비율	

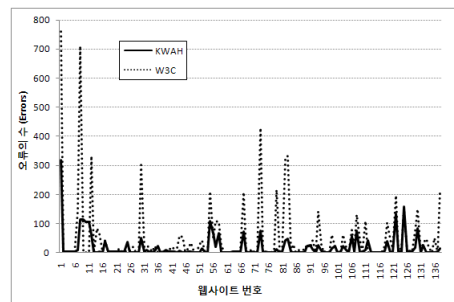
- 1) Web Accessibility(접근성): 웹접근성은 ‘장애인뿐만 아니라 모든 사람이 정보통신 기기나 서비스를 손쉽게 활용할 수 있도록 만드는 것’으로 정의하고 인식, 운용, 이해의 용이성과 기술적 진보성에 대한 오류의 수를 진단한다. 특히 한국형 웹콘텐츠 접근성 지침(K-WAH)에 따라 대체 텍스트 제공, 제목 제공, 기본 언어 명시, 사용자 요구에 따른 새 창 열기, 레이블 제공, 마크업 오류 방지의 세부 지침을 준수하는지를 점검한다[7,8].
- 2) Markup Validity(표준성): 웹표준의 준수 여부를 평가하기 위하여 W3C(World Wide Web Consortium)에서 분석 대상 웹사이트 주소를 입력하거나 HTML 파일을 업로드 하는 경우 마크업 언어의 표준 준수 여부를 점검한다[18].
- 3) Visits(방문 수): 정해진 시간 내 얼마나 많은 사람

이 사이트를 방문했는지를 측정하며, 이를 위해 cookie를 사용하고 사용자가 방문하고 떠날 때 세션을 시작하고 종료하기 때문에 방문 수는 '정해진 시간 내 발생하는 모든 세션의 합'으로 정의된다. 세션은 임시 쿠키 값에 의해 식별되고, 방문자가 첫 번째 페이지를 요청하면 그들의 세션이 시작되며, ① 방문자가 웹사이트를 떠나거나 브라우저를 닫는 경우, ② 방문자가 아무 행동도 취하지 않은 채 일정 시간이 지나면 웹서버가 자동으로 세션을 종료하는 경우의 상황이 발생할 때까지의 세션으로 측정한다.

- 4) Unique Visitors(순 방문자 수): 특정시간 동안 웹사이트를 방문한 고유한 방문자를 의미한다. 이 값은 사용자의 브라우저 애플리케이션에 설정되어 웹서버(혹은 분석 자바스크립트 태그)에 의해 읽히는 영구 쿠키에 의해 트래킹 된다. 따라서 순 방문자 수란 '정해진 시간 내의 모든 고유한 cookie_id의 합'으로 정의된다. 예를 들어, 어떤 사람이 동일한 컴퓨터에서 뉴욕 타임즈 웹사이트를 하루에 한 번씩 한 달(30일) 동안 매일 방문한다면, 방문 수(Visits)는 30, 순 방문자 수는 1이다.
- 5) Page Views(페이지 뷰 수): 웹사이트의 총 페이지 조회 수로써 방문시 보여진 페이지나 방문 세션동안 요청된 페이지 수를 의미한다.
- 6) Pages/Visit(방문 당 페이지 수): 방문당 평균 페이지 조회 수로써 페이지뷰 수(Page Views) 값을 방문수(Visits)로 나누어 구한다.
- 7) Average Visit Duration(평균 방문시간): 웹사이트에 방문 당 머문 시간의 평균(방문 길이)을 의미한다.
- 8) Bounce Rate(이탈률): 전체 트래픽 중 웹사이트에 일정시간(10초) 이내로 머문 트래픽의 비율이다. 단일 페이지 방문율(전체 방문시 단일 페이지의 방문 비율)로 구하기도 한다.
- 9) %New Visits(신규 방문 비율): 전체 방문 수 중 웹사이트를 처음 방문한 방문자의 비율로써클라이언트가 서버에게 보내는 cookie_utma(방문 정보)의 값의 합으로 정의된다.

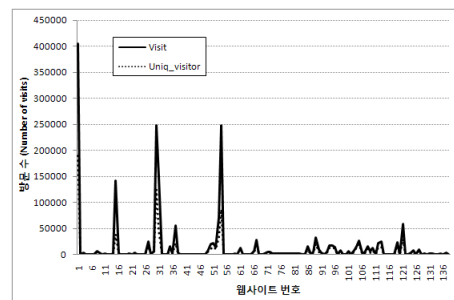
3. 상관관계 분석

<표 1>의 지표 값을 분석하기 위하여 경기도에서 운영하고 있는 138개의 홈페이지에 대한 2달 동안의 측정 자료를 이용한다[7,8]. 대표적으로 웹접근성 및 W3C 표준성 준수 여부(Errors의 수), 방문 수 및 순 방문자 수에 대한 분석 값을 나타내면 각각 [그림 1]과 [그림 2]와 같다. [그림 1]은 K-WAH와 W3C Validator를 이용해 진단한 오류의 수이다. K-WAH 오류의 수가 많은 경우 W3C로 진단한 오류의 수도 많으며, W3C 진단 도구로 평가한 오류의 수가 K-WAH 보다 다소 높음으로써 W3C에서 보다 정밀한 기술적 Errors를 진단함을 알 수 있다.



[그림 1] 웹접근성 및 표준 오류의 수

[그림 2]는 구글 애널리틱스를 이용해 측정된 방문 수(Visits)와 순 방문자 수(Unique Visitors)에 대한 결과이다. 대체적으로 순 방문자 수가 많은 웹사이트일수록 방문 수도 높음을 알 수 있다.



[그림 2] 방문 수 및 순 방문자 수

정량지표들 값에 대한 상관관계를 나타내면 <표 2>와 같다. 상관계수(r) 값이 양수(+)이면 양의 상관관계, 음수(-)이면 음의 상관관계가 존재하며 $|r| \geq 6$ 이면 상관관계가 있다고 보고, 특히 $|r| \geq 8$ 인 경우 강한 상관관계가 존재한다[2]. <표 3>은 상관관계에 대한 유의성 검정 결과이다. 유의수준 5%에서 유의확률 값이 0.05보다 작으면 통계적으로 매우 유의하다(즉, 두 변수 사이에는 상관관계가 존재)라고 본다. 상관관계가 존재($|r| \geq 6$ 인 경우)하는 요인들에 대한 유의성 검정결과 모두 유의한 것으로 나타났다. <표 2>에서 Visit-Unique visitor의 상관계수가 0.98로써 [그림 2]의 결과와 동일하게 강한 상관관계가 있는 것으로 보인다. 그리고 Page views-Avg. time(visit duration)의 상관계수가 0.878로써 페이지 뷰의 수가 클수록, 즉 사용자가 많은 페이지를 조회할수록 방문 시간이 클을 의미한다. 상관관계가 존재하는 지표들 사이의 산포도는 [그림 3], [그림 4]와 같다.

<표 2> 지표들 사이의 상관계수

	KWAH	W3C	Visit	Unique visitor	Page views
KWAH		0.722	0.445	0.489	0.183
W3C			0.430	0.473	0.154
Visit				0.980	0.712
Unique visitor					0.637

	Pages/visit	Avg. time	Bounce rate	%New visits
KWAH	-0.084	-0.035	0.025	0.001
W3C	-0.057	-0.057	-0.011	0.054
Visit	0.164	0.421	-0.179	-0.086
Unique visitor	0.147	0.337	-0.189	-0.042
Page views	0.373	0.878	-0.179	-0.079
Pages/visit		0.599	-0.744	0.368
Avg. time			-0.384	0.092
Bounce rate				-0.680

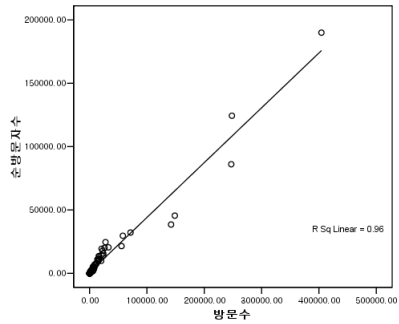
[그림 3]은 방문 수에 따른 순방문자 수에 대한 결과이다. 상관관계의 결정력을 나타내는 R square Linear 값은 0.96이고 이 값은 1에 가까울수록 두 변수 사이에는 강한 상관관계가 존재한다[2]. [그림 4]로

부터 페이지뷰의 수가 많을수록 평균 방문시간이 높으며 상관계수=0.878, R square Linear=0.77이다.

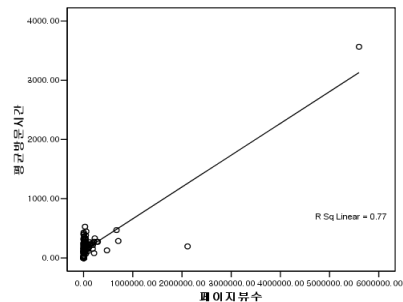
<표 3> 상관관계에 대한 유의확률

	KWAH	W3C	Visit	Unique visitor	Page views
KWAH		0.000	0.000	0.000	0.031
W3C			0.000	0.000	0.071
Visit				0.000	0.000
Unique visitor					0.000

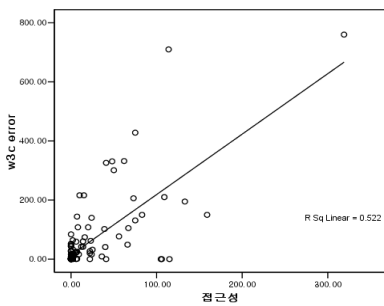
	Pages/visit	Avg. time	Bounce rate	%New visits
KWAH	0.328	0.680	0.769	0.990
W3C	0.508	0.509	0.897	0.528
Visit	0.054	0.000	0.036	0.317
Unique visitor	0.086	0.000	0.027	0.627
Page views	0.000	0.000	0.036	0.356
Pages/visit		0.000	0.000	0.000
Avg. time			0.000	0.283
Bounce rate				0.000



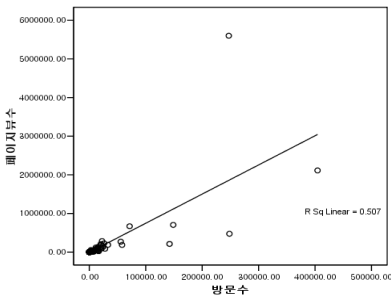
[그림 3] 방문수에 따른 순 방문자 수



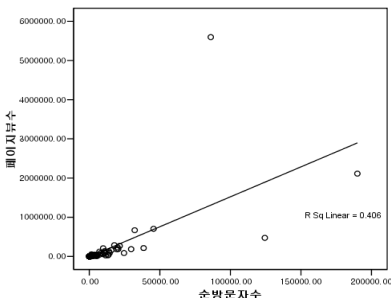
[그림 4] 페이지뷰 수에 따른 평균 방문시간
그 외 <표 2>에서 상관관계가 존재($|r| \geq 6$)하는 지표들 사이의 관계는 [그림 5]~[그림 9]와 같다. [그림 5]에서 접근성 오류가 많을수록 W3C validator에서 진단한 웹표준 오류의 수가 많으며, 한국에서 개발한 K-WAH의 웹접근성 진단도구는 국제 표준인 W3C에서 지정한 기술적 검증도구와 비슷한 웹사이트 평가 결과를 나타냄을 의미한다.



[그림 5] 접근성 오류의 수에 따른 웹표준 오류의 수

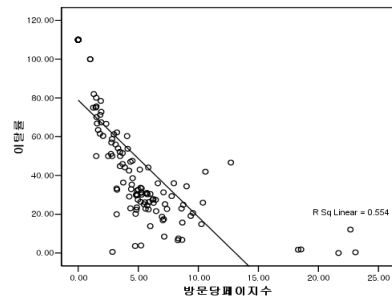


[그림 6] 방문 수에 따른 페이지뷰 수

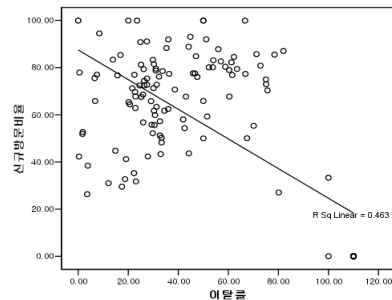


[그림 7] 순 방문자 수에 따른 페이지뷰 수

[그림 8]에서 방문당 페이지수와 이탈률 사이에는 음(-)의 상관관계가 존재한다. 이는 웹사이트 이용자의 방문당 페이지 조회수가 많을 수록 이탈률이 적음을 의미한다. 그리고 [그림 9]에서 이탈률과 신규방문비율 사이에도 음의 상관관계가 존재하며, 이탈률이 높은 웹사이트일 수록 신규 방문비율이 낮음을 나타낸다. 따라서, 웹사이트에 대한 이탈률을 줄이기 위해서는 사용자의 방문당 페이지 조회수를 늘이기 위한 웹사이트 콘텐츠의 최적 설계 방안이 마련되어야 하며, 단지 방문 수나 순 방문자 수를 늘이는 것 보다 처음으로 방문한 방문자의 비율을 높이는 것이 바람직하다.



[그림 8] 방문당 페이지 수에 따른 이탈률



[그림 9] 이탈률에 따른 신규방문비율

4. 결론

본 논문에서는 최적화된 웹사이트 설계를 위한 정

량적 평가 방법을 분석하였다. 평가 지표로써 접근성, 표준성 및 구글 애널리틱스의 지표값을 사용하였으며, 지표 결과값들에 대한 상관관계와 유의성 검정을 통하여 웹사이트 설계 방안을 제시하였다. 138개의 웹사이트들에 대한 분석 결과, 방문 수와 순방문자 수 그리고 페이지뷰 수와 평균 방문시간 사이에는 강한 상관관계가 존재하고, 방문 수·순방문자 수와 페이지뷰 수와 페이지뷰 수·신규방문비율과 이탈률 사이에도 상관관계가 존재함을 알 수 있었다. 특히, 페이지뷰 수·신규방문비율과 이탈률 사이에는 음(-)의 상관관계가 존재하며, 사용자의 웹사이트 방문시 이탈률을 줄이기 위해서는 방문 수나 순방문자 수를 늘이는 것보다 방문 당 페이지뷰 수를 높이기 위한 차별화된 웹콘텐츠 서비스 제공과 신규 방문 비율을 높이기 위한 웹마케팅 전략이 요구된다. 향후 정량적 평가 외에 정성적 평가 방법과 결합된 웹사이트 설계 방안에 대한 연구가 필요하다.

참고문헌

- [1] 강은정, 웹 기획 기초와 설계, 한빛미디어, 2012.
- [2] 노형진, 정한열, 한글SPSS-기초에서 응용까지, 2012.
- [3] 아비나쉬 카우척, 실전 웹사이트 분석, 에이콘, 2012
- [4] 에릭 피터슨, 웹사이트 분석의 기술, 한빛미디어, 2012
- [5] 오종혁, 웹기획 & 웹프로젝트 매니지먼트, 영진닷컴, 2012.
- [6] 이정기, 웹2.0시대-생각하는 웹기획자가 세상을 바꾼다. 비비컴, 2012.
- [7] 장희선, 박종태, “웹콘텐츠 서비스 평가,” 서비스연구, 제3권, 제2호, pp.33-44, 2013년 9월.
- [8] 장희선, 박종태, “정량적 지표를 이용한 웹사이트 분석,” 주간기술동향, 제1563권, pp.1-12, 2012년 9월.
- [9] 장희선, “Web Vulnerability Scanner를 이용한 취약성 분석,” 융합보안논문지, 제12권, 제4호, 2012년 9월.
- [10] Acunetix, <http://www.acunetix.com>.
- [11] A.V. Hausman and J.S. Siekpe, “The Effect of Web Interface Features on Consumer Online Purchase Intention,” Journal of Business Research, 2009.
- [12] Avinash Kaushik, Web Analytics 2.0, SYBEX,

2009.

- [13] Digital Analytics Association, <http://www.digitalanalyticsassociation.org>.
- [14] Eric Peterson, Web Site Measurement Hacks, O'REILLY.
- [15] Google Analytics, <http://www.google.com/analytics>.
- [16] R.L. Oliver, Satisfaction: A Behavioral Perspective on The Consumer, McGraw-Hill, 1996.
- [17] V. Mummalaneni, “An Empirical Investigation of Web-Site Characteristics Consumer Emotional States and On-line Shopping Behaviors,” Journal of Business Research, 2005.
- [18] World wide web consortium(W3C) validator, <http://validator.w3.org>.

[저자소개]



장희선 (Hee-Seon Jang)

KAIST 산업공학과(공학박사)

평택대학교 컴퓨터학과 교수
관심분야: 트래픽 엔지니어링