
사례연구

PcMeter를 이용한 인터넷 접속률 측정

Korean internet audience centric measurement: Internet index

이 상 경*

Lee, Sang Kyung

인터넷 리서치의 과학성을 담보하기 위해서는 과학적 원칙과 절차를 지켜야 한다. 그 핵심은 표본추출에 있으며, 특히 자기선택오류를 범하지 않기 위한 방안이 모색되어야 한다. 먼저 (인터넷 센서스)를 통해 네티즌의 분포와 변화상황을 주기적으로 체크하고, 그에 준한 표본추출을 실시하여 패널을 구축한다. 구축된 패널에 PcMeter를 장착케 하여 전송된 로그를 분석하여 제시하는데 이를 (인터넷 인덱스)라한다.

It is very important to keep scientific principles in a marketing research. Especially in sampling we have to select the scheme to avoid self-selection bias. Internet Index is a ready-made research service produced by analyzing the log data transferred from PcMeter installed in panelists' PC, with panelists' profile and site data base, following the scientific principles. Through the index we can figure out the actual behaviour of Korean netizen surfing various web sites and understand their cybergraphics.

I. 들어가는 말

몇 년전부터 인터넷리서치는 아주 실험적인 수준에서 시도되어 왔다. 그러다 점차 인터넷이 갖고 있는 상호 작용성, 즉시성, 응답의 편리성과 이로인한 저비용 등으로 인하여 이를 이용하여 조사를 하는 인터넷 리서치가 급증하였다. 그러나 현재 인터넷을 통한 조사는 표본설계, 응답절

* 인터넷메트릭스(sklee@internetmetrix.com)

차의 타당성과 신뢰성 등에서 많은 문제를 갖고 있다. 그로인해 리서치의 기본이 지켜지지 않은 많은 인터넷 리서치가 발표되기도 하고, 중요한 여론형성의 도구로 남용되거나 오용되고 있는 실정이다. 남용, 오용의 근본 원인은 리서치를 과학적 도구로 인식하지 못한데서 오는 것이라고 본다.

가장 중요한 원인은 표본추출에 대한 이해부족이 아닐까 싶다.¹⁾ 리서치는 과학적 원칙과 절차를 준수할때만 그 과학성을 담보받을 수 있다. 그 중에서도 표본추출에 대한 이해가 무엇보다 중요하다. 인터넷 리서치의 표본추출과 관련하여 가장 심각하게 우려해야 할 사항은 자기선택오류(self-selection bias) 문제를 어떻게 해결할 것인가이다. 인터넷조사는 “3중 자기 선택 오류”에 노출되어 있다. 첫째는 모집단에서 인터넷 사용자가 대표성을 가질 수 있는가의 문제, 둘째는 인터넷 사용자 중에서 응답자의 대표성을 어떻게 확보할 것인가, 셋째는 다중 응답자(multiple response)를 어떻게 막을 것인가²⁾이다.

과학적 인터넷 리서치를 수행하기 위해서는 무엇보다 먼저 모집단인 인터넷 사용자에 대한 이해가 전제되어야 할 것이다. 다음 절에서는 인터넷 사용자의 규정 및 현황에 대하여 살펴보고자 한다.

II. 인터넷 사용자 규정 및 현황

인터넷 사용자의 모집단에 대한 이해를 위해서는 인터넷 사용자에 대한 규정이 전제된다고 하겠다. 인터넷 사용자는 인터넷 관련 기관별로 각기 다른 기준을 사용하고 있다. 주1회이상, 월1회이상 등 이용의 양, 빈도에 따라, 연령제한 없이 할 것인가, 7세이상, 12세이상, 15세이상, 16세

1) 한국조사연구학회의 인터넷리서치 보도시 유의사항 7항목 중 5항목을 표본추출에 할애하고 있다.

2) 김용학(2000. 4), “인터넷리서치의 발전을 위한 제언”, 인터넷메트릭스 주최, 『인터넷 마케팅의 새로운 패러다임 : 인터넷 인덱스』세미나에서 발표된 자료.

이상 등 연령별로 규정할 것인가, 또한 www만 포함시킬 것인지, 혹은 TCP IP 통신규약을 따르는 www를 비롯한 모든 Application을 포함시킬 것인지 하는 인터넷의 범위나 규정 등이 다르게 적용되고 있어 아직 세계적으로 표준화되어 있지 않다. 이는 어쩌면 당연한 현상일지도 모른다. 인터넷은 아직 초기진입단계로서 그 범위가 급속히 커져가기 때문에 계속 정의가 새롭게 내려져야 할 것이다.

인터넷 사용자의 현황을 파악하기 위하여 인터넷 메트릭스는 우리나라에서 처음으로 1999년 6월에 「인터넷센서스」³⁾를 실시하였다. 이는 국내 인터넷 시장구조 및 변화추이, 인터넷 사용자 특성과 이용행태 변화를 정기적으로 추적 분석하고, 해외 인터넷 시장동향과의 비교분석을 통하여 국내 인터넷 시장의 기본지표를 마련하고 나아가 향후 변화를 예측함으로써 인터넷 리서치와 마케팅 과학화의 기초를 마련하기 위한 작업이다. 또한 내부적으로는 PcMeter를 이용한 인터넷 이용률 측정시 국내 인터넷 이용자들의 구성비대로 표본을 추출하기 위해서 시행되었다. 조사대상은 전국의 15-49세 이상 남녀 10,000명을 인터넷이 아닌 전화조사를 통하여 실시하였다. 그 이후 1999년 11월(2차, 15-49세 10,000명, 전화조사), 2000년 4월(3차, 2세 이상, 7,000명, 전화조사)에 걸쳐 세 차례의 「인터넷센서스」가 실시되었다.

처음 두 차례는 아직 인터넷 보급률이 낮다고 보아 15세-49세를 대상으로 하였으나 3차 인터넷 센서스에서는 1999년의 급속한 한국 인터넷 시장의 성장을 반영하기 위하여 2세이상으로 연령층을 확대하였고, 무선 인터넷을 포함하여 조사를 실시하였다.

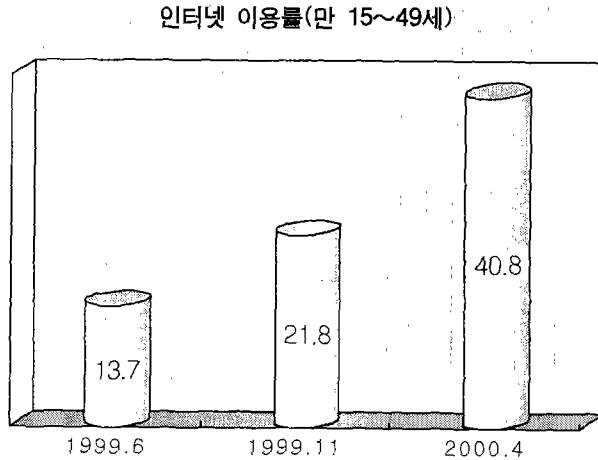
이상의 규정에 따라 우리나라 인터넷 사용자 현황을 「인터넷 센서스」

3) 센서스는 인구, 국제조사라는 뜻에서 전수조사로 발전한 개념으로 「인터넷센서스」가 표본조사임에도 인터넷 인구에 대한 조사이고 99년 상반기 국내에서 관련조사가 없는 환경에서 대규모 표본으로 진행된 조사라는 배경에서 붙여진 상표명이다. 인터넷메트릭스가 실시한 「인터넷 센서스」는 기업체 등에 지난 두차례 판매하였으며, 이번 3차 조사결과는 5월 하순경에 통합편(만2세 이상), 여성편(만2세 이상 여성), 만15-49세 Tracking편, 만 2-19세편(미취학 및 초, 중, 고생)으로 총 4권이 발간되어 판매할 예정이다.

국내기관	인터넷 이용자정의	출 처
Internet MetriX	만15세-49세로 월 1회 이상 인터넷을 이용하는자 만2세이상 (3차센서스부터)	www.internetmetrix.com
정보통신부	만 7세 이상으로 한달에 1회 이상 정기적으로 인터넷을 이용하는자	www.mic.go.kr
아이클릭	13세-49세로 이메일과 웹을 이용하는 자	www.eyeclick.co.kr
국외기관	인터넷 이용자 정의	출 처
MediaMetrix	만 2세이상 최소 월 1회이상 인터넷 이용자	www.mediametrix.com
GVU	연령제한 없음. WWW 이용자	Georgia대 부설 멀티미디어연구기관 www.gvu.gatech.edu
Computer Industry Almanac	만 16세 이상으로 인터넷을 한달에 1회 이상 꾸준히 이용하는 자	미 인터넷 컨설팅 기관 www.c-i-a.com
Commerce Net	만 16세 이상으로서 지난 한달간 인터넷을 1회이상 이용한 경험이 있고 지금도 이용하는자	미 전자상거래 컨소시엄 www.commerce.net
IDC	만 12세 이상으로 www 이용자	미 정보통신시장분석 & 예측기관 www.idc.com
FIND/SVP	18세 이상의 성인으로서 최근 3개월동안 이메일 혹은 인터넷을 이용한 자	www.findsvp.com
Nua	연령제한없음 최근 3개월동안 1회이상 인터넷을 이용한자	www.nua.ie
Nielsen Media Research	16세이상, 월1회이상, 최근 3개월이내 www 이용자	www.nielsenmedia.com

결과를 통해 살펴보면, 2000년 4월 현재 만 2세 이상 인구 중 약 1,427만 명(32.3%)이 인터넷을 이용하고 있는 것으로 나타나고 있다. 주요 경제활동인구(만 15-49세)의 인터넷 이용률은 40.8%로 지난 6월 (13.7%) 대비 10개월 여만에 +37.1%p 증가하여 급속한 성장세를 보이고 있다. 이러

한 현상은 전세계적인 것으로 세계 인터넷 시장은 최근 2년간 연평균 50%이상 매우 빠른 성장세를 보이고 있으며, 1994년 이후 전세계 인터넷 이용자의 65%를 차지하던 미국은 1998년 51%, 1999년 43%로 점차 감소하는 한편, 유럽과 아시아 및 기타국가의 비중이 상대적으로 높아지고 있다.

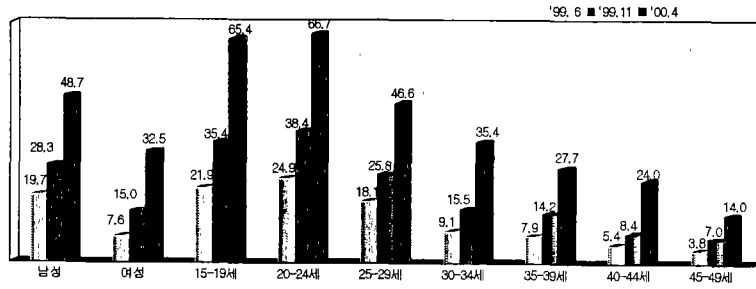


인터넷 이용의 급성장 추세는 특히 여성층에서 두드러져 남녀간 인터넷 이용률 격차가 좁혀지고 있다. 지난 6월의 27%에 비해 인터넷 이용자 중 여성의 비중이 33.7%(1999년 11월), 39.7%(2000. 4월)로 점차 커지고 있다. 이는 미국의 1997년 상황과 유사하다.

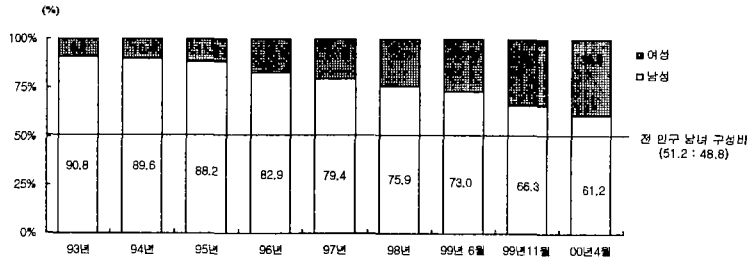
또한 연령별로 전체 인터넷 인구의 50%이상을 차지하고 있는 10-20대를 포함하여 전 연령대에서 급속한 증가를 보여 전 연령층으로 확산되는 경향이 보인다. 지역별로는 서울 및 수도권이 대체로 높은 이용률을 보이고 있지만 인터넷의 특성상 지역간 격차는 그리 크지 않은 것으로 보인다.

그러나 여전히 지역규모별, 성별, 연령별로 정보화 격차(Digital Gap)가 나타나고 있는 점을 유의해야 할 것이다. 대도시가 전체 인터넷 인구 중 차지하는 비중이 감소경향을 보이고 있으나(60.7%→61.9%→52.9%), 아직

성별, 연령별 인터넷 이용률 추이

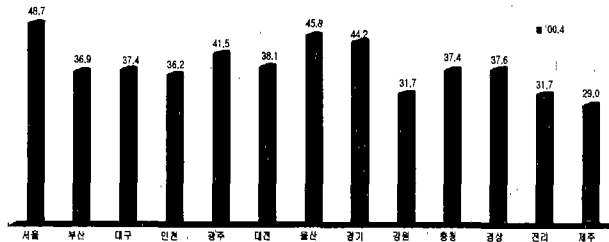


인터넷 이용자의 성별 구성비 추이



까지 도시지역에서 인구 대비 인터넷 이용률이 상대적으로 높은 편이다 (대도시 43.1%, 중소도시 41.1%, 군단위 30.0%). 성별로는 남성 중 인터넷 이용률이 19.7%(1차), 28.3%(2차), 48.7%(3차), 여성은 7.6%(1차), 15.0%(2차), 32.5%(3차)로 「주부인터넷교실」 등 정부의 적극적 지원시책에도 남

지역별 인터넷 이용률 추이



너 인터넷 이용률 격차는 해소되지 않고 있다. 연령별 이용율도 여전히 2000년 4월 현재 10대 후반(65.4%), 20대 초반(66.7%)이 60%를 상회하고, 20대 후반(46.6%), 30대 초반(35.4%), 30대 후반(27.7%), 40대 초반(24.0%), 40대 후반(14.0%)로 세대간 격차가 상존하고 있다.

Ⅲ. "PcMeter"를 이용한 측정방식

인터넷 리서치는 인터넷에 대한 조사(인터넷을 중심으로한 공급자와 수요자, 즉 웹사이트와 네티즌에 대한 조사)와 인터넷을 조사수단으로 이용한 조사로 구분해 볼 수 있다. 첫 번째 인터넷을 대상으로 한 조사는 인터넷 경제의 규모, 파급효과, 성장 속도, 인터넷 상거래 규모 등 새로운 사회과학 탐구 영역의 출현으로 볼 수 있는 영역과 사용자 행동조사로서 누가 서핑을 많이 하며, 어느 곳에 들르며, 인터넷 이용자의 움직임 경로를 조사하는 것과, 두 번째로 인터넷을 이용한 조사는 우리가 흔히 인터넷 리서치 할 때 주로 이 조사를 의미하는데 일반 서베이를 인터넷이라는 매체를 이용하여 실시하는 조사이다. 인터넷과 이메일이 전화 수준의 보급률에 도달할때까지는 인터넷이 조사수단으로 이용될 수 있는 분야는 조사의 아이템이 컴퓨터나 네트워크, 인터넷 관련 분야에서만 주로 한정될 수 있으리라 본다. 그렇다면 인터넷에 대한 조사의 경우를 살펴보자.

기존의 설문 방식으로는 네티즌들이 자신의 무수한 클릭들을 기억하는 것은 불가능한 일이어서 설문에 의한 답변을 취합해 온 기존의 조사 방법으로 네티즌들의 인터넷 이용행태를 파악하는 데는 한계가 있다. 따라서 TV 시청률 조사에서처럼 표본 추출한 네티즌들의 PC에 로그 트래킹 소프트웨어를 설치하여 방문 사이트를 측정하는 방법이 이에 대한 대안이라 생각한다. 국내에서는 최초로 도입되는 인터넷 접속률 조사는 기존의 TV 시청률 조사기법을 인터넷에 적용한 방식으로 네티즌의 컴퓨터에 접속 추적 소프트웨어(Log Tracking Software)를 설치한 후 이들의 인터넷 방문 기록을 수집하여 네티즌의 인터넷 이용행태를 분석하는 서비

스이다. 국내에서는 인터넷 메트릭스가 최초로 개념을 도입하여 최초로 서비스를 실시 중에 있다. 세계적으로는 미디어 메트릭스(MediaMetrix)가 1996년 최초로 서비스를 시작하였고, 닐슨넷레이팅스(Nielsen-Netratings), 엡피씨데이터(@PCdata), NetValue 등이 서비스 중에 있다.

모든 네티즌들의 PC에 동일한 소프트웨어를 설치(전수조사) 할 수 없는 현실에서 어떤 네티즌을 어떻게 표본 추출할 것인가가 중요한 문제이다. 지원자들을 모집해서 패널을 구성하면, 적극적인 인터넷 이용자와 경품 등 인센티브에 민감한 층의 가입율이 높아질 것이다. 그리고, 특정 웹사이트를 통해 지원을 받는다면 해당 사이트의 방문자가 과다하게 나올 가능성이 있고, 조사 결과에 민감한 특정 이해집단의 참여를 통제할 수 없게 될 것이다.

국내 인터넷 이용자를 대표하는 집단을 표본으로 추출하기 위해서는 모집단에 대한 조사가 선행되어야 한다. 이를 위해 인터넷메트릭스에서는 「인터넷 센서스」를 실시하여 인터넷 이용행태에 관한 기본적인 사항들을 주기적으로 전화조사하고 있다. 통계청에서 전국민을 대상으로 전수 조사하여 5년마다 발표하는 인구센서스에 근거하여 성, 연령, 지역별 인구 구성비에 따라 10,000명이 무작위 추출된다. 본 조사를 통해 성, 연령, 지역, 학력, 직업, 이용장소 등을 기준으로 한 네티즌의 구성비를 알 수 있고, 이에 비례하여 패널을 구성한다. 예를 들어, 통계청의 인구주택총조사에 의하면 15-49세 남녀 성비는 51.1 : 48.9이지만, 패널의 남녀 구성비는 최근 3차 인터넷 센서스 결과에 따라 61.2 : 38.8의 비율로 구성된다.

패널의 동의를 받은 후 로그 추적 소프트웨어('PcMeter')를 전자우편으로 송부하는데, 이용장소 및 PC공유 여부에 따라 PcMeter도 조금씩 달라진다. 이는 두 개 이상의 PC를 통해 인터넷을 접속하는 경우와 하나의 PC를 여러명이 사용하는 경우에도 실제 이용자별로 로그 데이터를 분리하여 현실을 보다 정확히 반영해 내려는 노력의 산물이다. 또한 전자우편을 이용하지 않는 네티즌에게는 ID와 비밀번호를 부여한 후 웹사이트에서 프로그램을 다운로드 받게 하고 있다.

패널이 자신의 PC에 PcMeter를 설치한 후 인터넷에 접속하면 방문한

웹사이트의 URL 및 방문시간이 인터넷 이용환경(OS, 해상도, CPU, Memory 등)에 대한 정보와 함께 인터넷메트릭스 서버로 자동 전송된다. 수집된 로그 데이터를 방문자의 프로파일과 교차시켜 분석하면, 각 웹사이트에 어떤 사람들이 방문하는지 알 수 있으며 반대로 특정 인구통계적 성격을 갖춘 사람들이 어떤 사이트를 주로 방문하는 지를 파악할 수 있다. 이런 과정을 거쳐 만들어진 '인터넷 인덱스'(인터넷 접속률 조사)는 페이지뷰(pageview) 뿐만 아니라 방문자수(unique visitor), 방문회수(visit), 체류시간(duration) 등 다양한 기준으로 사이트별 인구통계적 변수 별로 집계하여 매주 온라인에 제시된다. 방문자수는 특정기간 동안 해당 사이트를 한번이라도 가본 적이 있는 사람들의 수에 대한 추정이고, 방문횟수는 특정기간 동안 전체 네티즌들이 그 사이트에 방문했을 횟수를 합한 숫자이다. 한사람이 여러 사이트를 번갈아 반복 사용하면 짧은 시간에 방문횟수가 계속 늘어 날 수 있기 때문에 일정시간 내에 다시 사이트를 방문할 경우 별도의 방문으로 간주하지 않는다. 또한 LAN 환경의 네티즌들이나 정액제로 인터넷에 접속하는 네티즌들에게는 인터넷 접속 후 다른 용무를 보는 경우가 있는데, 데이터의 왜곡을 막기 위해 한 페이지당 체류시간의 최대치를 두고 있다.

인터넷 인덱스는 웹사이트 순위 이외에도 프로퍼티(property, 동일 기업이 운영하는 웹사이트 집합체)나 포털, 언론, 금융, 증권, 쇼핑몰 등 카테고리별 순위를 제공하며, 특정 사용자 그룹(예를 들어, 20대 대졸 여성 등)의 웹사이트 방문 현황도 보여 준다. 그리고, 이와는 별도로 고객들에게는 보다 심화된 분석 결과들이 제공되는데, 여기에는 특정 웹사이트를 방문하는 경로, 경쟁사와의 방문자 중복, 장기적인 변화 추이 등에 대한 자료가 포함될 예정이다.

리서치는 '과학'이다. 자연과학적 원칙들에 비교하면 엉성해 보이기도 하지만, 그건 인간, 사회가 갖고 있는 특성에 연유하는 것으로 그만큼 어렵고 실은 많은 투자가 필요한 작업이다. 그러니 더욱더 과학적 원칙에 충실해야 하지 않을까 한다. 그 과학성이 가장 지켜져야 할 부분은 표본 추출에 있다. 지난 1년여간 자원과 시간의 대부분을 투자하였지만 아직도 나갈 길이 멀다고 생각한다. 인터넷 매체의 다양성을 반영하기 위하

여 앞으로 2000년말 까지 패널을 2만명으로 확대할 계획으로 있으며, 패널 구축과 유지, 관리에 여러 가지 시도와 자원을 지속적으로 투입할 예정이다. 또한 조사 대상 연령을 2세 이상으로 확대할 것이고 PC방과 학교에 대한 측정과 인터넷이 가능한 모든 플랫폼에 대한 측정을 준비 중에 있다.

인터넷 세상의 강력한 힘은 투명성, 개방성에서 나온다고 한다. 그 기반이 되는 접속률 데이터가 필요한 시점이며, 생산된 데이터를 활용한 다양한 수요, 다양한 문제상황에 적용되길 바란다. 인터넷 인덱스는 국내 기술력과 마케팅 노하우에 의해 구축된 최초의 인터넷 접속률 조사이고, 국내 인터넷 시장의 현황을 객관적으로 보여주는 틀이다. 인터넷 인덱스를 활용하여 인터넷 세상을 이해하고 인터넷 성장의 기초가 되길 바라며, 우리의 이러한 시도가 학계에서도 공유되길 바라며 이 글을 마친다.⁴⁾

연구의 활성화를 위하여 연구주제 등 관련 협의를 거쳐 인터넷센서스와 인터넷 인덱스의 데이터를 학계와 공유하고자 합니다. 연락처는 다음과 같습니다.

전화: 02-561-5595; E-mail: ischo@internetmetrix.com

4) 인터넷 인덱스는 주간단위로, Off-line Customized Report는 월단위로 클라이언트의 주문에 따라 생산, 판매된다.