

# 인터넷사용자측정 조사에서의 표집틀의 문제

장 원 호\*, 박 동 선\*\*

인터넷 사용자가 과연 어떤 사이트를 주로 이용하고 그 이용 시간은 어떠한 수준인지를 확인하고 더 나아가서 각 사이트 이용자들은 어떠한 특성을 나타내고 있는지를 알아보는 인터넷사용자측정 (Audience-centric Measurement)의 중요성은 인터넷 조사의 기반이 되는 데이터를 제공한다는 점뿐 아니라 최근 활발히 연구되고 있는 데이터마이닝(Data Mining)과 그것을 바탕으로 하는 CRM (Customer Relationship Management)의 기초자료로서 그 중요성이 날로 커지고 있다.

본 연구는 기존의 인터넷사용자측정 조사의 국내외 연구 동향과 그 차이점을 살펴보고, 국내외 기존 조사가 가지고 있는 문제점, 특히 조사의 표집틀의 왜곡과 관련된 문제점을 분석하고 이를 해결하기 위한 실천적, 이론적 대안을 제시하고 있다.

## I. 서 론

정보사회의 변화과정을 혁명적이라고 표현한다면 그 혁명의 중심추가 되는 것은 인터넷의 발달이라 할 수 있다. 인터넷의 발달은 비단 기술적 변화만을 초래한 것이 아니라, 기존의 근대성의 여러 전제들의 변화를 초래하는 담론과 문화 패러다임의 변화를 동반하는 문화적, 생활세계적 변화를 초래했다는 점에서 그 영향의 심대함을 가늠할 수 있다.

최근의 인터넷 사용자의 수에 대해 여러 통계적 차이가 있지만, 우리 나라의 경우 인터넷을 사용하는 사람의 비율이 전체 인구의 40%를 넘어 섰다고 조사되고 있다. 이것은 인터넷이 이미 우리 생활세계의 가장 주된 커뮤니케이션의 수단으로 발달했다 해도 과언이 아닌 단계에 이른 것이라 할 수 있다. 더 나아가서 인터넷은 그 자체로 이미 중요한 시장을 형성하고

---

\* 서울시립대학교 도시사회학과

\*\* 인텔리서치

있으며 기존의 오프라인 시장과 경쟁 또는 보합적 관계로서 그 경제적 중요성이 더욱 커지고 있다.

이런 점에서 인터넷 사용자가 과연 어떤 사이트를 주로 이용하고 그 이용 시간은 어떠한 수준인지를 확인하고 더 나아가서 각 사이트 이용자들은 어떠한 특성을 나타내고 있는지를 알아보는 인터넷사용자측정(Audience-centric Measurement)은 그 중요성이 날로 커지고 있다 할 수 있다. 특히 인터넷사용자측정은 데이터마이닝(Data Mining)과 그것을 바탕으로 하는 CRM(Customer Relationship Management)의 기초자료로서 향후 인터넷 조사에서 차지하는 역할에 매우 클 것으로 예측된다는 점에서 그 중요성이 더욱 크다 할 수 있다.

이런 맥락에서, 본 연구는 기존의 인터넷사용자측정 조사의 국내외 연구 동향을 먼저 살펴보고, 국내의 조사가 가지고 있는 문제점, 특히 조사의 표집틀의 왜곡과 관련된 문제점을 분석하고 이에 대한 해결 대안을 제시하고자 한다.

## II. 인터넷사용자측정(Audience-centric Measurement) 조사의 국내외 연구 동향

### 1. 국외 연구 동향

인터넷 사용자 측정(Audience-centric Measurement)은 인터넷 사용자를 대표하는 표본집단을 구성하고 표본집단의 인터넷에 사용현황을 조사하여 분석하는 것을 말한다. 이는 흔히 사이트로그 분석(Site-centric Measurement)과는 다른 차원의 웹로그(Web log) 분석방식이다.

먼저 사이트로그 분석방법은 해당 사이트에 접속하는 모든 기록을 조사하는 방식으로서 전수조사의 성격을 가지고 있다. 하지만 프록시 서버를 통해 접속했을 경우는 단 한번 접속한 것으로 기록되며(현재 25-30% 정도에 이룸) 사용자의 브라우저에서 캐싱(caching)되는 경우는 서버 로그 파일에 기록되지 않는다는 단점이 있다. 반면 인터넷사용자 측정의 경우 프록시 서버나 사용자의 브라우저에서 캐싱되는 경우를 포함해서 사용자가 접속한 모든 사이트의 정보를 수집할 수 있다는 장점이 있다. 하지만 인터넷 사용자 측정 조사는 전체 인터넷 사용자를 측정할 수 없기 때문에 패널을 통한 표본조사라는 한계를 가지고 있다.

사이트 로그 측정방법의 경우 각 웹사이트가 자발적으로 스스로의 웹로그파일(Web log file)을 제 삼의 인증(Auditing) 기관에 제출해야만 조사가 진행될 수 있다. 반면 인터넷 사용자 측정 방식의 경우 웹사이트의 협조가 없어도 측정이 가능한 방식이다. 뿐만 아니라, 인터넷 사용자 측정의 경우, 어떤 사이트에 어떤 사회인구학적 특성을 가진 방문자가 방문하는지, 한 사람이 100번 방문하는지, 아니면 100명이 한번씩 방문하는지 등의 방문자의 특성에 관한 분석에 강점을 가지고 있다. 따라서, 미국에서는 많은 민간업체에서(슈퍼마켓 미디어 메트릭스, 닐슨

네티레이팅, PC DATA ONLINE, 넷벨류, alexa) 경쟁적으로 인터넷 사용자 측정 결과를 발표하고 있다.

인터넷 사용 측정 결과를 발표하는 외국 업체는 크게 alexa와 그 외 회사의 2가지 그룹으로 나누어 볼 수 있다. 첫 번째 그룹은 alexa로서 표본추출방식으로 자발적 응답 방식을 사용하고 있다. 즉, 인터넷 사용 측정에 참여하고자 하는 응답자는 alexa 홈페이지를 방문하여 사용자 측정을 위한 프로그램을 응답자의 PC에 설치하면 응답자가 접속한 웹사이트의 URL이 alexa의 서버로 자동 전송되고 이를 합산하여 인터넷 사용자 측정을 진행하고 있다.

두 번째 그룹은 슈퍼터 미디어 매트릭스, 닐슨네티레이팅, PC DATA ONLINE, 넷벨류로서 무작위 표본추출에 의해 응답자를 선정 한 후, 인터넷 사용자 측정을 위한 프로그램을 설치한다는 표본추출 방식이 alexa와 다르다. alexa는 응답자를 확인할 수 있는 ID와 같은 부호를 부여하고 있지 않기에 인구, 통계적 정보를 수집하고 있지 않다. 단지 전세계의 컴퓨터 중 alexa가 설치된 컴퓨터에서 어떤 웹사이트가 로딩(loading)되고 있는가를 측정하고 있다. 따라서, alexa에서 웹사이트 순위를 측정하는 기준은 표본으로 확보된 PC에서의 웹사이트 접속 데이터를 통해 일정기간 각각의 웹사이트에서 열람된 총 웹페이지수(Pageview)를 추정하고 이를 순위 기준으로 사용하고 있다.

슈퍼터 미디어 매트릭스, 닐슨네티레이팅, PC DATA ONLINE, 넷벨류와 같은 회사의 경우, 무작위 추출된 표본집단을 통해 일정기간 각각의 Web site를 접속한 순 이용자수(Unique Visitor)를 순위 선정의 기준으로 사용하고 있다. 즉 alexa는 전세계에서 인터넷에 연결된 컴퓨터 중 alexa 프로그램이 설치된 컴퓨터를 표본집단으로 사용하는 반면, 슈퍼터 미디어 매트릭스 및 그외의 조사회사는 전체 인터넷 이용자 중 무작위로 추출된 인터넷 이용자를 표본집단으로 사용하는 조사 설계방식이다. alexa는 전세계 50만대의 PC에 Web site 이용현황 측정을 위한 프로그램이 설치되어 있으며 한달 동안 18테라바이트에 해당하는 방대한 분석 자료를 바탕으로 조사 결과를 제공하는 것으로 알려져 있다. alexa 방식의 조사의 경우 상대적으로 간편하게 전세계 Web site의 사용 현황을 추정할 수 있는 장점이 있는 반면 각 국가별 분석 결과의 신뢰성에 의문이 제기되고 있다.

예를 들어, 2000년 10월 조사 결과 전세계 웹사이트 중 세이클럽(4위), 아이러브스쿨(5위), 다음(11위), 한메일(12위), 야후코리아(13위), 라이코스코리아(18위), 한게임(21위), 네이버(41위), 네띠앙(43위), 프리첼(50위)로 조사되었다. 이 중 세이클럽은 지난달 대비 3배나 늘어난 결과이다. 이에 대해 당사자인 세이클럽측도 “내부 집계로는 10월 한달 동안 9월과 비슷한 13억 페이지뷰 정도”라며 “회사에서도 놀라 alexa 측에 경위설명을 요청할 예정”이라고 말할 정도이다(디지털 타임스, 2000년 11월 13일).

인터넷 이용자 중 일부를 표본집단으로 사용하는 조사 설계 방식에서는 원칙적으로 표본으로 추출된 인터넷 이용자가 어떤 컴퓨터에서 인터넷에 접속하든지 인터넷 사용에 관련된 자료를 수집할 수 있어야 한다. 그러나 현실적으로 인터넷 카페, 학교, 호텔, 공공기관과 같이

특정 컴퓨터를 특정 사용자가 사용하지 않는 환경에서 표본으로 선정된 인터넷 이용자가 인터넷에 접속할 경우는 측정하기 어려운 문제점이 발생한다. 미국의 경우 인터넷 카페, 학교, 호텔, 공공기관에서 인터넷에 접속하는 트래픽(traffic)이 전체 트래픽에서 차지하는 비중이 극히 작은 수준이라고 판단하여 제외하는 방식으로 조사 설계가 이루어지고 있다.

## 2. 국내 연구 동향

국내에서도 인터넷 사용자 측정을 진행하고 있거나 준비하고 있는 많은 조사업체가 있다. 본 연구에서는 크게 두가지 그룹으로서 인터넷 메트릭스(Internet Metrix)와 웹패턴테크놀로지(Rankserv)의 조사에 대해 살펴보겠다.

인터넷 메트릭스에서는 15세에서 49세의 국내 인터넷 이용자를 대상으로 집 또는 직장에서 접속한 웹사이트를 순방문자수(Unique Visitor) 기준으로 분석하고 있다. 표본 추출 방식은 슈퍼미디어 메트릭스와 같이 무작위 추출을 통한 응답자 선정과 전화인터뷰의 방식을 통해 표본 인터넷 이용자를 추출했으며 표본 크기는 5,000명으로 밝히고 있다. 이에 반해 웹패턴테크놀로지에서는 전 연령대의 국내 인터넷 이용자를 대상으로 접속 장소에 상관없이 인터넷에 접속한 결과를 순방문자수(Unique Visitor)와 페이지뷰(Pageview) 두 가지 기준을 합산하여 분석하고 있다. 표본추출 방식은 Alexa와 동일하게 자발적 응답자가 프로그램을 설치하는 방식으로 현재 표본 규모는 25,000명으로 알려져 있다. 웹패턴테크놀로지가 alexa와 달리 프로그램을 설치하는 응답자의 간단한 사회인구학적 특성에 대한 정보를 받고 있다. 하지만 주민등록번호, 전화번호 등의 정보가 없기 때문에 응답자의 인구학적 특성에 대한 신뢰성을 확인하기는 어려운 설계방식이다.

그런데, 문제는 두 조사기관의 인터넷 사용자측정 결과가 <표 1>에 나타나 있듯이 매우 큰 차이를 보인다는 것이다.

<표 1>에서 알 수 있듯이, 두 조사기관의 조사 순위결과에서 공통되는 웹사이트는 20위까지 살펴볼 때 불과 8개밖에 되지 않는다. 이렇게 동일한 시기에 조사된 인터넷사용자 측정조사가 크게 다른 이유는 무엇일까? 본 연구에서는 두 조사기관의 인터넷사용자 측정조사에서의 표집틀의 차이에 기인한다고 생각한다. 즉, 인터넷 메트릭스는 집 또는 직장에서 인터넷에 접속한 결과를 기준으로 조사한 반면, 웹패턴테크놀로지는 접속장소에 관계없이 측정된 조사결과라는 점이 두 조사기관의 순위의 차이를 낳았다는 것이다.

<표 1> 2000년 10월 30일부터 2000년 11월 5일까지 Internet Metrix와 RankserV  
(웹패턴테크놀러지)의 순위결과

순위	Internet Metrix	RankserV (웹패턴테크놀러지)
1	* 다음(daum.net)	* 다음(daum.net)
2	* 야후코리아(kr.yahoo.com)	조이링크 (joylink.co.kr)
3	* 라이코스 코리아(lycos.co.kr)	* 모교사랑 (iloveschool.co.kr)
4	* 네이버(naver.com)	* 라이코스 코리아(lycos.co.kr)
5	네티앙(netian.com)	* 야후코리아(kr.yahoo.com)
6	슈퍼보드(superboard.com)	세이클럽 (sayclub.com)
7	한미르 (hanmir.com)	프리첼 (freechal.com)
8	* 드림위즈 (dreamwiz.com)	한게임 (hangame.com)
9	인터피아98 (interpia98.net)	* 드림위즈 (dreamwiz.com)
10	천리안 (chollian.net)	* 신비로 (shinbiro.com)
11	엠포스 (empas.com)	* 네이버 (naver.com)
12	* 모교사랑 (iloveschool.co.kr)	씨네월컴 (cinewel.com)
13	하이홈(hihome.com)	다모임 (damoim.net)
14	* 신비로(shinbiro.com)	심마니 (simmani.com)
15	msn (msn.com)	팍스넷 (paxnet.co.kr)
16	* 드림엑스(dreamx.net)	디지털조선일보 (chosun.com)
17	야후 (yahoo.com)	* 드림엑스(dreamx.net)
18	마이크로소프트 (microsoft.com)	SBS (sbs.co.kr)
19	우투 (wo.to)	옥션 (auction.co.kr)
20	하나넷 (hananet.net)	와우콜 (wowcall.com)

표집들을 실제 표본이 추출되는 목록이라 할 때, 이 표집들의 가장 중요한 요소로서는 가능한 한 연구하고자 하는 전체모집단을 모두 포함하여야 한다는 것이다. 그런데 인터넷메트릭스의 경우 집과 직장에서 접속하는 사람들만을 표집들에 포함시키고 있는데, 만약 그 외의 지역에서 인터넷을 사용하는 사람들의 비율이 높다면 이것은 표집들 작성에 매우 큰 오류를 범하고 있는 것이다.

이와 관련하여, 최근 인텔리서치에서 인터넷 이용자들의 접속 장소별 인터넷 사용시간 비율을 조사한 것이 <표 2>에 나타나 있다.

<표 2> 인터넷 사용장소별 이용시간

단위: %

접속 장소	인터넷 사용 시간
집	43.0
직장	25.0
PC방	21.3
학교	7.7
학원	1.7
기타	1.3

출처: 인텔 리서치(2000. 4.)

<표 2>에서 알 수 있듯이, 집(43.0%) 또는 직장(25.0%)에서 인터넷을 이용하는 이용시간은 전체 인터넷 이용 시간의 68% 정도이고 나머지는 PC방(21.3%), 학교(7.7%), 학원(1.7%), 기타(1.3%)로 조사되었다. 이 조사결과는 한국의 인터넷 사용 환경의 특수성을 반영하는 것으로 미국의 경우 집 또는 직장에서 인터넷에 접속하는 비율이 대단히 높고 인터넷 카페 등에서 인터넷에 접속하는 시간이 낮은 비율인 반면, 우리사회에서는 PC방에서의 인터넷 사용율이 21.3%에 이르고 있는 것이다. 이러한 상황을 고려해볼 때 인터넷 메트릭스의 경우, 집 또는 직장 이외의 장소에서의 인터넷 접속하는 32%가 표집틀에서 제외되고 있는 것이다.

그렇다면, 웹패턴테크놀러지 조사의 측정결과가 옳은 것일까? 웹패턴테크놀러지는 프로그램이 설치된 모든 접속장소에서 측정하고 있기에 인터넷메트릭스 보다는 표집틀의 포괄성이 높다 할 수 있다. 하지만, PC방이나 학교 등에서 프로그램이 설치된 경우, 표본으로 선출된 응답자 이외의 다른 이용자가 인터넷을 사용하는 경우에도 동일한 ID를 통해서 로그 데이터가 전송되게 된다는 점에서 응답자의 특성 파악에서 큰 오류를 범하게 되는 것이다. 즉, 웹패턴테크놀러지의 조사결과는 페이지뷰를 측정하는데는 큰 문제가 없지만, 순방문자수를 측정하는데는 신뢰도가 떨어진다고 할 수 있다. 특히, 우리 사회와 같이 다수의 방문자가 한 명의 방문자로 측정되는 PC방, 학교 등에서의 인터넷 이용이 높은 상황에서는 그 신뢰도가 더욱 떨어지게 되는 것이다.

그런데, 위에서 제기된 표집틀의 문제는 만일 집 또는 직장에서 인터넷을 사용하는 집단과 PC방, 학교에서 인터넷을 사용하는 집단 사이에 체계적인 차이가 없다면, 표집틀의 타당성이 크게 의심받지 않으므로 표집틀의 포괄성 부족이 그다지 큰 문제가 아닐 수도 있다. 이런 점에서, 본 연구에선 인터넷 사용장소에 따른 사용자 집단의 사회인구학적 특성 차이를 분석해 보았다.

### Ⅲ. 인터넷 사용장소별 집단간 사회인구학적 특성 비교

먼저 국내의 인터넷 사용자들이 과연 어느 장소에서 주로 인터넷을 사용하는 지를 알아보았다.

<표 3> 인터넷 주사용지의 분포

인터넷 주사용지	빈도	상대빈도 (%)
집	1748	44.9
직장	710	18.2
PC방	915	23.5
학교	379	9.7
학원	84	2.2
기타	59	1.5
총계	3895	100.0

자료: 인텔리서치 (2000. 4.)

<표 3>에서 알 수 있듯이 집과 직장 이외에서 인터넷을 사용하는 사람들의 비율이 40%에 이르고 있다. 이는 한국의 인터넷 사용 문화가 서구와는 달리, 경제적, 공간적 이유로 PC 사용에 제한을 받고 있는 집단들이 PC방과 같은 지역에서 인터넷을 주로 사용하고 있는 현실을 반영하고 있다. 이것이 또한 한국과 유사한 사회경제적 여건에 있는, 제 3세계의 향후 인터넷 사용 문화에 대한 큰 함의를 제공하고 있다 하겠다.

<표 2>의 결과 못지 않게 중요한 사실은 PC방에서 인터넷을 주로 사용한다고 응답한 집단의 88.9%가 자신의 인터넷 사용시간 중 70%이상을 PC방에서 사용한다고 응답하고 있으며, 더 나아가 60.5%에 달하는 554명은 오직 PC방에서만 인터넷을 사용한다고 응답하고 있다. PC방에서만 인터넷을 사용하는 554명은 전체 인터넷 사용자 표본의 14.2%에 이르는 수치인 것이다. 이러한 경향은 학교에서 인터넷을 사용하는 집단에게도 거의 동일하게 발견되고 있기에 PC방과 학교에서의 인터넷 사용자의 특성을 분석하는 작업은 향후의 일반적 인터넷 조사를 위해서도 매우 필요한 일이 아닐 수 없다.

<표 4>에서는 인터넷 주사용지를 집, 직장, PC방, 학교를 선택한 네 집단의 사회인구학적

특성을 살펴보았다. 이 분석의 취지는 인터넷 주사용자가 다른 네 집단이 과연 동일한 집단으로 평가할 수 있는지, 아니면, 체계적 차이가 있는 집단이지를 분석하기 위한 것이다. 위에서 언급한 바와 같이 인터넷 주사용자가 다른 네 집단의 사회인구학적인 특징에 큰 차가 없다면 기존의 인터넷사용자측정 조사의 타당성은 큰 문제가 되지 않겠으나, 만약 사회인구학적으로 큰 차이를 보인다면, 기존 조사의 표집들의 포괄성 부족은 타당성의 오류로 연결될 수 있는 것이라 할 수 있다.

<표 4> 인터넷 주사용자 별 집단 간 사회인구학적 특성 비교

단위: %

		집	직장	PC방	학교	전체이용자
성별	남성	52.0	71.9	66.3	41.4	57.5
	여성	48.9	28.1	33.7	58.6	42.5
연령별	만 9세 이하	2.0	0.0	0.1	3.4	1.6
	10대	30.9	2.4	51.0	53.6	33.0
	20대	26.8	37.1	42.1	36.4	32.7
	30대	24.4	38.7	5.7	4.2	20.5
	40대	12.7	19.2	1.0	1.3	10.1
	50대	2.2	2.5	0.0	0.8	1.6
	60대이상	1.0	0.0	0.1	0.3	0.5
직업별	초등학생	9.2	0.1	2.5	18.8	7.7
	중학생	11.0	0.0	17.6	7.7	10.3
	고등학생	9.2	0.3	19.7	17.2	10.6
	대학생	10.8	0.6	19.7	46.3	14.0
	사무직	14.2	67.0	11.5	5.0	21.8
	전문자유직	2.8	4.3	1.7	0.3	2.5
	생산직	1.7	2.8	2.1	0.0	1.8
	자영업	10.0	18.8	5.4	0.3	9.2
	판매서비스직	2.0	2.7	2.5	0.3	2.0
	전업주부	22.4	2.0	2.2	2.4	12.5
	농림축산어업	0.2	0.1	0.1	0.0	0.2
	무직/기타	6.5	1.3	15.1	1.9	7.3
Total		100.0 (n=1745)	100.0 (n=708)	100.0 (n=915)	100.0 (n=379)	100.0 (n=3906)

자료: 인텔리서치 (2000. 4.)



<표 4>에 나타나 있듯이, 각 사용자의 사회 인구학적 특성과 그들이 주로 인터넷을 사용하는 장소는 밀접히 관련되어 있음을 알 수 있다. 먼저 성별 인터넷 주사용지를 비교해 보면, 남성은 직장이나 PC방에서의 사용비율이 높은 반면, 여성은 학교에서 사용비율이 상대적으로 높은 것을 알 수 있다. 연령별 차이를 보면, PC방과 학교에서의 10대와 20대의 사용비율이 매우 높은 점이 두드러진 반면, 직장에서는 30대 이상의 사용비율이 상대적으로 높은 것을 알 수 있다. 직업별 특징을 보면, 집에서 주로 사용하는 직업별 비율은 전체 비율과 매우 유사하나, 전업주부의 비율이 상대적으로 매우 높다. 직장은 당연한 얘기이지만, 사무직 종사자의 비율이 압도적으로 높은 것을 알 수 있다. 반면 PC방은 중, 고, 대학생들의 비율이 상대적으로 높고, 학교에서는 초등학생과 대학생의 사용비율이 매우 높은 것을 알 수 있다. 이러한 결과는 앞에서 제기되었던 표집틀의 문제가 기존의 조사 방법으로는 해결될 수 없는 심각한 것임을 반증하고 있다. 다시 말하자면, PC방과 학교에서 인터넷을 주로 사용하는 집단과 집/직장에서 인터넷을 주로 사용하는 집단과의 체계적 차이가 없다면, 기존의 조사 방법에서 PC방, 학교에서 인터넷을 주로 사용하는 집단이 표집틀에서 빠져 있다 하더라도 타당성에 큰 문제를 제기하지 않겠지만, <표 4>에 나타난 바와 같이 두 집단의 사회인구학적 차이가 매우 큰 상황에서 기존의 조사는 모집단을 올바르게 대표하는 표집틀에 입각한 조사라고 할 수 없는 것이다.

#### IV. 인터넷 사용자 측정 조사의 표집틀 문제 해결을 위한 대안

현재 PC방의 현황에 대한 공식적인 통계가 제공되고 있지 않지만 업계에서는 전국적으로 20,000여 곳의 PC방이 있고 PC방마다 평균 30대씩으로 계산하면 600,000대의 PC가 있는 것으로 추정하고 있다. 학교의 경우도, 2000년 10월 말 현재 전국 1,165 초·중·고등학교의 66%인 7,664개 학교에 초고속 국가망을 이용한 인터넷 서비스가 이루어지고 있다(정보통신부와 한국전산원 발표). 또한 정보통신부는 올해 안에 인터넷이 연결돼 있지 않은 3,986 학교에 대해서도 인터넷 서비스를 개통할 예정이라고 밝혔다. 이런 점을 고려할 때, 한국의 인터넷 사용 현황과 관련하여 집과 직장 뿐만 아니라 PC방과 학교에서 인터넷 접속은 그 트래픽 규모에서 큰 비중을 차지할 뿐만 아니라, 사용자 특성에서도 차이를 보이고 있기 때문에 인터넷사용자 측정에서 이들 집단을 어떠한 형태로든 고려하지 않을 수는 없다.

먼저, 현실적인 어려움을 고려하지 않고 이론적 차원의 대안은 어떤 방법을 통해서든지 PC방과 학교에서 인터넷을 주로 사용하는 집단을 표집틀 내에 포함시켜야 한다는 것이다. 이를 위해서는 먼저 PC방이나 학교에서 인터넷을 사용하는 집단에 대한 목록을 작성하는 것이 필요하다. 이렇게 작성된 PC방/학교에서 주로 인터넷을 사용하는 집단의 목록에서 기존 조사의 패널수와 PC방/학교에서 주로 사용하는 집단이 차지하는 비율을 고려하여 표본의 수

를 정한다. 예를 들어 인터넷 메트릭스의 경우, 집, 직장 사용자 집단에서 3,000명의 패널을 보유하고 있다면, PC방과 학교 사용자 집단에서 1,500명 정도를 추가로 확보하는 것이 필요하다 하겠다. 그런 다음, PC방이나 학교에서 인터넷을 사용하는 집단에 대해서는 컴퓨터 자체가 ID로 인식되는 방법이 아닌 방법으로 그들을 인식할 수 있는 솔루션을 개발해야 할 것이다. 예를 들어, PC방이나 학교에서 인터넷을 사용하는 표본에게는 타당한 인센티브를 제시하고 그들이 PC방이나 학교에서 인터넷을 사용할 때, 사용하기 전과 사용을 마친 후에는 꼭 특정 사이트를 방문하도록 하는 것이다. 즉, 인터넷을 사용할 때 특정 사이트에 먼저 로그인하고 인터넷을 사용하고 사용을 마친 후에는 그 사이트에서 로그아웃하도록 하는 방법이다. 특정 사이트는 그들의 개인 ID와 방문 사이트, 사용시간 등이 자연스럽게 집/직장에서 사용하는 집단과 동일한 형태로 전환되는 기능만 있으면, 표집들의 문제는 해결된다고 할 수 있다. 하지만, 이 방법은 기존의 인터넷 사용자측정 조사를 위한 패널을 구할 때, 간단한 소프트웨어 한번만 설치하면 되는 데도 불구하고 표본 패널을 구축하기 어려웠던 현실을 생각할 때는, 매번 인터넷을 사용할 때마다 특정 사이트에 로그인과 로그아웃을 해야 한다는 점에서 패널 구축의 실현성이 그다지 높은 대안이라고는 할 수 없다.

보다 현실을 감안한 대안으로는 PC방과 학교에서 인터넷을 사용하는 집단의 측정결과를 하나의 별개의 측정결과로 분류하여 분석하는 방법이 있을 수 있다. 위에서 제기된 이론적 대안이 용이하지 않다면, 일단 집과 직장에서의 인터넷 사용자 측정 방식과 PC방, 학교의 조사방식을 분리해서 서로 다른 방식의 조사 방법을 통해 조사하고 이를 가능한 범위 내에서 통합하여 분석하는 방식을 선택하는 대안을 모색할 필요가 있다. 예를 들어, 집과 직장에서 주로 인터넷에 접속하는 층의 경우, 인터넷 이용자를 표본추출하는 방식을 사용하여 조사하고 PC방, 학교의 경우 인터넷에 접속된 PC를 표본 추출하는 방식으로 조사하는 대안을 모색할 필요가 있다. 그런 다음, PC방이나, 학교에서 인터넷을 주로 사용하는 집단들의 사회인구학적 특성을 고려하여, PC방과 학교에서의 인터넷사용자 측정 결과를 재해석하는 방법이 있을 수 있다. 물론 이 방식으로는 현재 제기된 표집들의 문제를 해결했다고는 할 수 없다. 그러나, 현재의 기술적 수준과 현실적인 조사 진행의 어려움을 고려한다면 하나의 대안이 될 수 있다.

## V. 결 론

본 연구는 인터넷사용자 측정 조사에 국한하여 인터넷 조사에서 제기될 수 있는 표집들의 문제를 제기하고 이에 대한 가능한 해결책을 제시하고자 하였다. 그럼에도 불구하고 본 연구가 문제의 인식 차원에 머물러 있지, 문제의 해결과 관련된 대안이 미흡하다는 점은 필자들

도 부인할 수는 없다. 하지만, 앞으로 인터넷 조사의 필요성이 더욱 커질 것은 명백한 사실이고, 이와 관련하여 인터넷사용자측정은 향후의 인터넷 조사의 기반을 이룰만한 중요한 자료를 이루는 조사라는 점에서 그 표집틀의 포괄성과 타당성 문제는 시급히 해결하지 않으면 안되는 현안인 것이다. 더 나아가서 인터넷사용자측정 조사의 데이터가 궁극적으로는 데이터마이닝이나 CRM의 기초자료로 사용된다는 점에서, 본 연구가 제시한 표집틀 왜곡 현상은 비단 조사관련 학계나 업계만의 문제가 아니라, 이를 이용한 데이터마이닝이나 CRM을 추구하는 업계가 공동으로 협력하여 해결할 문제인 것이다.

## 〈참 고 문 헌〉

디지털 타임스, 2000년 11월 13일

이상경, 2000. “PcMeter를 이용한 인터넷접속률 측정” 『조사연구』 제 1권 1호

홍종필, 2000. “인터넷 마케팅에 있어서 웹 오디언스 트래킹 데이터의 활용”

한국사연구학회 인터넷서베이 워크샵 발표논문, 2000. 8. 22.