

# 인터넷 인덱스 방법론

2004. 10.

(주) 메트릭스 코퍼레이션

**MetriX**

[www.metrixcorp.com](http://www.metrixcorp.com)

# 목 차

- 데이터 생산 프로세스
- SCM vs. ACM
- Metrix vs. Koreanclick

# 데이터 생산 프로세스

# 데이터 생산 프로세스

## ■ 조사분석 플로우



모집단조사

조사대상 : 국내 만6세 이상, 10,000명  
조사주기 : 6개월

패널모집

성, 연령 등 인터넷 이용자 구성비를 반영하여 패널 구성비 결정  
RDD (Random Digit Dialing)로 패널 무작위 추출

로그수집

PcMeter가 패널의 웹 서핑로그를 수집, 수신서버로 자동전송

데이터분석

Log DB, Profile DB, Site DB를 통합한 다차원 분석  
웹 상에서 '인터넷 인덱스' 서비스 제공

## ■ 조사개요

- 조사명 : KRNIC, '인터넷 이용자수 및 이용행태 조사'
- 모집단기준: 주민등록인구 조사결과
- 조사대상: 국내 만 6세 이상 국민
- 표본추출: 다중 층화 무작위추출법 (Multi Stratified Random Sampling)
- 층화기준: 지역\*가구주연령
- 표본수 : 10,000여명
- 조사방법: 패널구축(면접) 후 전화조사
- 조사주기: 반기 1회 (메트릭스 1999. 6. 이후 총 12회 수행)
- 조사활용: 인터넷 이용자수 및 이용자 구성비 파악



## ■ 패널모집

- **패널대상:** 국내 만 6세 이상 만 59세 이하 인터넷이용자
- **모집단 정보:** 인터넷 이용자수 및 이용행태 조사
- **층화 기준:** 성\*연령
- **패널수:** 12,500여명
- **모집방법:** 1) RDD (Random Digit Dialing) 방식의 선발  
2) 대상자에 PcMeter 이메일 송부
- **모집 시 :** 1) 스크리닝: 인터넷이용 및 관련업체종사 여부  
2) 이용장소 및 가정 이용 시 공용여부  
3) Panel Profile
- **PcMeter 유형 :** 5종  
: 가정1인, 가정2인, 직장, 가정1인+직장, 가정2인+직장



## ■ 패널 구성

- 최근 인터넷 이용자수 조사결과를 토대로 국내 인터넷 이용자의 구성비를 반영하여 결정

## ■ 패널 선정

- 공개적인 패널모집 광고 및 이벤트 모집, 특정사이트 회원을 대상으로 한 패널 선정을 원천적으로 배제

## ■ 패널 개입 방지

- 패널가입 동의 이후 간단한 설치과정으로 로그 트래킹 SW를 따로 실행하거나 종료할 필요가 없음
- 타 어플리케이션의 운용, 컴퓨터 사용에 전혀 지장을 초래하지 않아 데이터 수집과정에서 발생할 수 있는 비표본오차 제거
- 패널 리워드 프로그램, 메일링 등이 패널의 인터넷 이용행태를 변화시킬 수 있는 여지를 배제

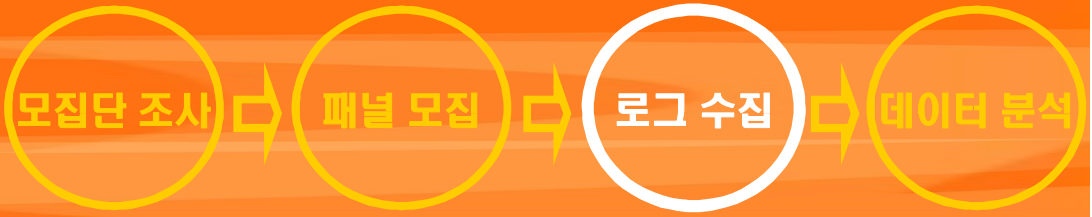


## ■ 로그 샘플

```
#PANELID=ZZZ #VERSION=2.0 #LOGDATE=2000-12-14 9:59:37  
#OSVERSION=Windows 98 #SCREEN=1024x768 #CPU=Pentium (II/Pro)  
#MEMORY=128MB
```

BRWSR	TIME	WEB SITE & PAGE
IE1	6:55:57 pm	www.xxx.com
IE1	6:56:44 pm	shopping.xxx.com/index.html
IE2	7:03:58 pm	www.yyy.com
IE2	7:04:58 pm	login.yyy.com/BIN/login.cgi





## ■ 로그 수집 by PcMeter

- 수집정보: 환경정보, 시간, URL
- 환경정보: OS, Memory, CPU, 화면해상도, IP 등
- 수집 URL File 확장자: html, htm, cgi, php, jsp, asp, gif, jpg, asf, asx, wmv
- 미수집
  - http, https로 시작하지 않는 로그(ex. ftp)
  - About:blank
  - <http://localhost/>
  - URL 길이 400자 이상
  - img, embed태그를 통해 웹페이지에 삽입된 이미지, 플래쉬, 배너, 동영상 등
  - 브라우저: Internet Explore (Netscape, 전용 브라우저 X)



## ■ 로그 전송

- 일시 저장 후 일정 시간 간격으로 전송
- 시간보정: Client ~ Server 시간 차이 조정
- 네트워크 장애 시 클라이언트에 로그 저장
- 수신서버: 4대 (3대@IDC, 1대@Metrix)
- 방화벽 통과 : http 80 port 이용



## ■ 데이터 분석



```
#PANELID=ZZZ #VERSION=4.0 #LOGDATE=2002-12-14 9:59:37
#OSVERSION=Windows XP #SCREEN=1024x768 #CPU=Pentium III
#MEMORY=512MB
```

BRWSR	TIME	WEB SITE & PAGE
IE1	6:55:57 pm	www.xxx.com
IE1	6:56:44 pm	shopping.xxx.com/index.html
IE2	7:03:58 pm	www.yyy.com
IE2	7:04:58 pm	login.yyy.com/BIN/login.cgi



## ■ 측정 단위

	shopping.xxx.com/index.html
• 웹사이트 (Website)	xxx.com
• 서브도메인 (Sub Domain)	shopping.xxx.com
• 채널 (Channel)	쇼핑
• 프로퍼티 (Property)	[주] XXX
• 카테고리 (Category)	검색포탈 / 포탈
• 메인페이지 (Main Page)	www.xxx.com, www.xxx.com/index.html, xxx.com, www.xxx.co.kr/index.html,...
• 유사도메인 (Synonyms)	xxx.co.kr, xxx.net, x3.com, x3.co.kr, x3.net, 3x.com,... => xxx.com,



## ■ 측정 지표

순방문자수

- 일정기간 특정 웹사이트를 방문한 중복되지 않은 이용자수

방문일수

- 일정기간 특정 웹사이트 방문자들이 사이트에 방문한 일수

방문횟수

- 특정 웹사이트의 총 방문횟수
- 30분 내 재 방문한 경우 별도 방문으로 계산하지 않음

페이지뷰

- 일정 기간 특정 웹사이트 방문자들이 확인한 웹페이지 총수
- 3초 이내에 생성되는 URL은 페이지 산정에서 제외

사이트 체류시간

- 특정 웹사이트의 방문자들이 사이트에 체류한 총 시간
- 5분 이상 추가적인 웹로그 발생이 없으면 그 이후는 카운트하지 않음



## ■ 분석 사이트 기준

- 표본조사에서의 통계적 유의수준에 근거하여 기간 내 인터넷 접속 패널의 0.3% 이상의 접속률을 보이는 사이트를 분석대상으로 함

# SCM vs. ACM

# SCM vs. ACM

## ■ 인터넷 측정방식 비교

	웹서버 측정 방식	패널 측정 방식
개요	<ul style="list-style-type: none"> <li>· 사이트의 웹서버 로그 분석</li> <li>· 가장 보편적으로 수행</li> <li>· <u>전수조사</u></li> </ul>	<ul style="list-style-type: none"> <li>· 표본 추출된 네티즌의 웹로그 분석</li> <li>· TV시청률 조사로부터 발전된 모델</li> <li>· <u>표본조사</u></li> </ul>
장점	<ul style="list-style-type: none"> <li>· 자사 사이트에 대한 풍부한 데이터</li> </ul>	<ul style="list-style-type: none"> <li>· 동일한 기준으로 여러 사이트 비교</li> <li>· 인구통계분석 가능</li> </ul>
단점	<ul style="list-style-type: none"> <li>· 고비용, 시간 소요</li> <li>· 사이트 간 비교 어려움</li> <li>· 인구통계 분석에 어려움</li> </ul>	<ul style="list-style-type: none"> <li>· 표본수 =&gt; 통계적 Fluctuation</li> <li>· 조사회사 기준에 따라 데이터 상이</li> </ul>



# SCM vs. ACM

## ■ 인터넷 측정방식 비교

	웹서버 측정 방식	패널 측정 방식
과대측정	<ul style="list-style-type: none"> <li>· 검색Agent 등 Robot의 활동</li> <li>· Internal Usage (내부개발자)</li> <li>· 프레임/광고, 팝업</li> <li>· Sent Data 기준</li> </ul>	<ul style="list-style-type: none"> <li>· 패널의 적극적 이용자 성향</li> <li>· (프레임/광고, 팝업)</li> </ul>
과소측정	<ul style="list-style-type: none"> <li>· Proxy, Cache 제외</li> </ul>	<ul style="list-style-type: none"> <li>· PC방 등 공용장소 제외 됨</li> <li>· 대상 연령 제한</li> <li>· 국외 이용자 제한</li> </ul>

# SCM vs. ACM

## ■ 데이터 불일치에 대한 연구

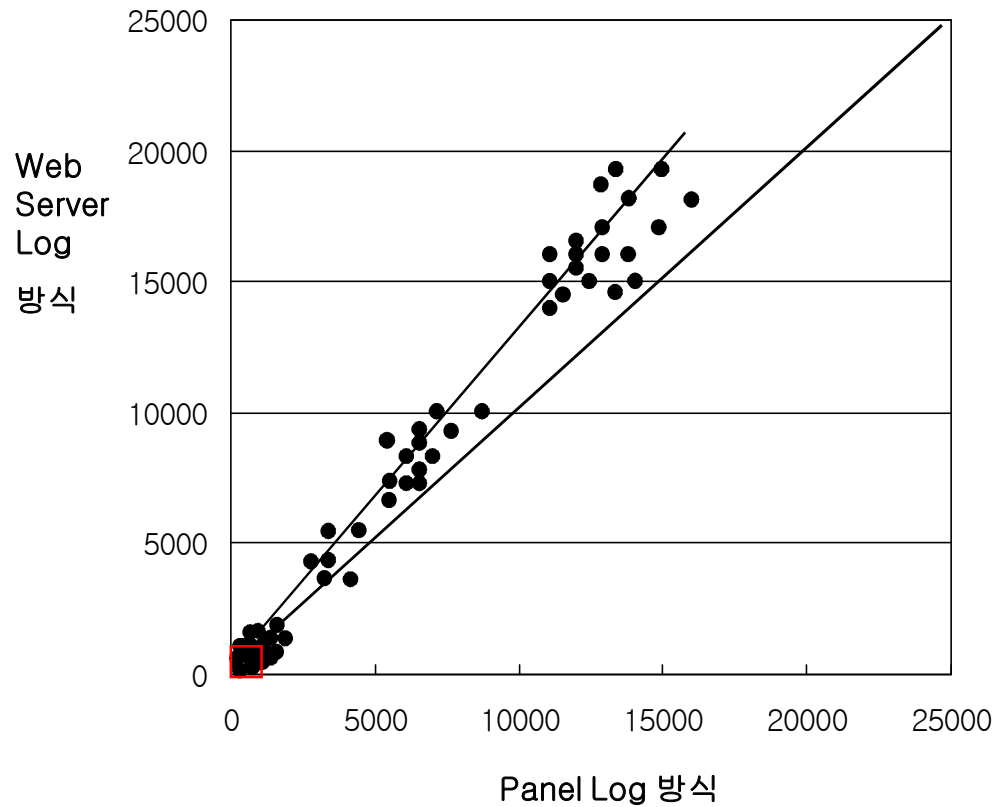
- **연구목적: 패널방식과 웹서버측정 방식의 데이터 비교**
- **측정시기: 1998년 11월~12월**
- **측정사이트**
  - Dell, Excite@Home, LA Times, Economist, AOL, Playboy, CNN, MSN, Yahoo!, USA Today 외 총 35개 사이트
- **연구참여자**
  - Yahoo!, Santa Clara Univ., Advertising Research Foundation, Internet Advertising Bureau
- **측정기관**
  - 패널 데이터 : Media Metrix, Nielsen//NetRatings
  - 웹서버 Audit: ABC Interactive, I/Pro

# SCM vs. ACM

## ■ 데이터 불일치에 대한 연구

- 전체사이트 : 불일치 존재 (패널로그 방식 과소측정)

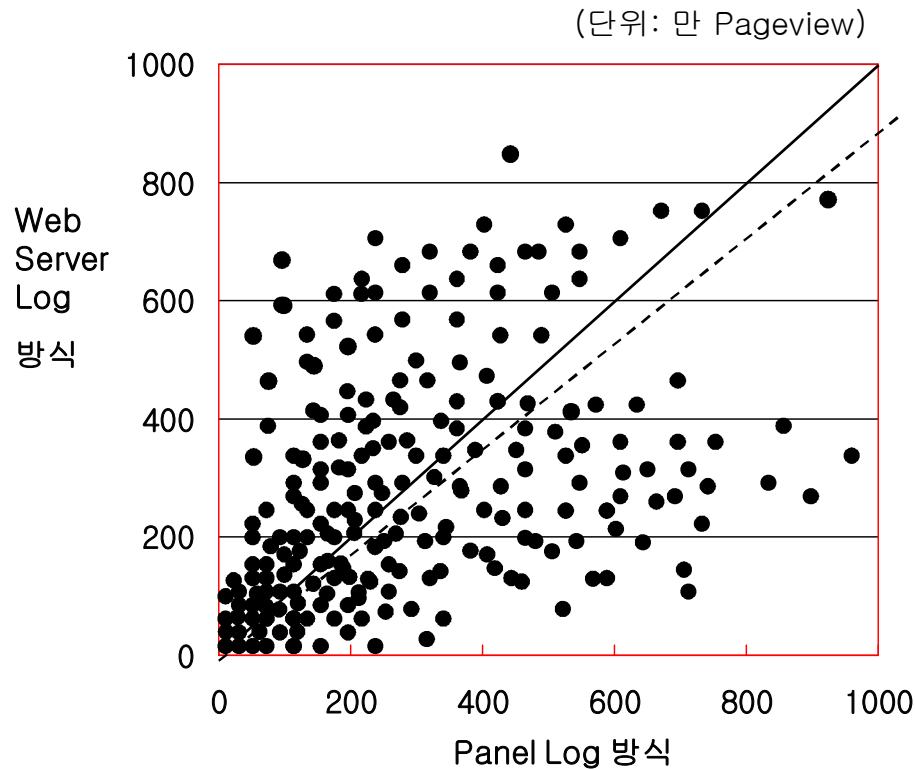
(단위: 만 Pageview)



# SCM vs. ACM

## ■ 데이터 불일치에 대한 연구

- Small sites : 불일치 심화 (패널로그 방식 과대측정)



$$Y=0.8423X$$

$$R^2=0.3424$$

\* Small Sites :  
일1,000만 페이지뷰 이하 사이트

# Metrix vs. Koreanclick

# Metrix vs. Koreanclick

## 패널 구성 비교

	Metrix			Koreanclick	
패널구성	만6~59세 인터넷 이용자			만10~65세 인터넷 이용자	
Top 30 (1월 기준)	순위	웹사이트	방문자수(000)	웹사이트	방문자수(000)
	1	daum.net	25,134	daum.net	23,012
	2	naver.com	23,311	naver.com	22,350
	3	kr.yahoo.com	19,197	yahoo.co.kr	18,913
	4	nate.com	16,526	nate.com	18,261
	5	bugs.co.kr	13,677	bugs.co.kr	15,113
	6	auction.co.kr	11,827	hanafos.com	14,962
	7	dreamwiz.com	11,323	dreamwiz.com	13,187
	8	hanafos.com	11,105	chol.com	13,011
	9	empas.com	9,983	empas.com	12,922
	10	msn.co.kr	9,519	auction.co.kr	12,181
	16	com.ne.kr	7,245	interpark.com	9,629
	17	netmarble.net	7,204	damoim.net	8,810
	18	hangame.com	6,924	imbc.com	8,722
	19	imbc.com	6,773	lgeshop.com	8,666
	20	freechal.com	6,761	netpia.com	8,535
	21	cafe24.com	6,744	freechal.com	8,473
	22	sbs.co.kr	6,741	com.ne.kr	8,382
	23	hihome.com	6,426	sbs.co.kr	8,164
	24	kbs.co.kr	6,345	okcashbag.com	8,044
	25	interpark.com	6,324	e-station.com	7,923
	26	damoim.net	6,238	kbs.co.kr	7,906
	27	chosun.com	5,995	netmarble.net	7,680
	28	buddybuddy.co.kr	5,124	chosun.com	7,630
	29	011e-station.com	5,123	hmall.com	7,619
	30	hmall.com	5,092	cjmall.com	7,575

# Metrix vs. Koreanclick

## ■ 조작적 정의

### - 방문 횟수

: 30분 이내 재방문은 동일 방문으로 처리

### - 체류 시간

: 5분 이상 페이지 이동이 없는 경우 체류시간을 5분으로 한정

### - 프레임에 대한 처리

: 3초 이내에 동시에 올라오는 URL은 페이지뷰 산정에서 제외하는  
방식으로 프레임을 제거

# Metrix vs. Koreanclick

## ■ 페이지뷰 지표 산출 조작적 정의



2004-04-02 17:57:17	UW	IE16	<a href="http://www.hanafos.com/hdredir.htm?hdttitle=HF_Man-t00_HF-h">http://www.hanafos.com/hdredir.htm?hdttitle=HF_Man-t00_HF-h</a>
2004-04-02 17:57:18	UW	IE16	<a href="http://www.hanafos.com/main/default_hfzone.asp">http://www.hanafos.com/main/default_hfzone.asp</a>
2004-04-02 17:57:18	UW	IE16	<a href="http://mainiframe.hanafos.com/popqueryHF.asp">http://mainiframe.hanafos.com/popqueryHF.asp</a>
2004-04-02 17:57:18	UW	IE16	<a href="http://image.hanafos.com/ad_banner/text_add_zone.html">http://image.hanafos.com/ad_banner/text_add_zone.html</a>



# Metrix vs. Koreanclick

## ■ 채널 분류

- 서브도메인 단위의 지표 산출이 아닌 보다 하위의 URL들을 묶어 지표 산출

