

인터넷 조사에서 민감한 정보를 얻기 위한 방안 구현

An Implementation of Method for Obtaining Sensitive Information at Internet Survey

박희창* 남기성** 이기성***

요 약

본 논문에서는 기업 등에서 사원이나 고객을 대상으로 민감한 정보를 얻기 위한 조사에서 응답자들이 정직하게 응답하기를 꺼리는 질문들에 대하여 응답자의 비밀을 노출시키지 않고서 보다 정확한 정보를 얻을 수 있는 방안으로 간접응답기법인 확률화응답기법을 인터넷 상에서 사용할 수 있도록 구현하고자 한다. 본 시스템은 DB 환경에 바탕을 두어 기존의 온라인 설문 조사 시스템과 연계하여 자료를 공유할 수 있을 뿐만 아니라 독립된 스폿 서베이(spot survey)가 가능하도록 구현하고자 한다.

I. 서론

온라인 설문조사는 대표성의 결여 등과 같은 단점이 있음에도 불구하고 김정기 외(1999)에서 서술한 바와 같이 오프 라인(off-line)의 조사에서 가지지 못한 많은 장점이 있기에 급속도

* 창원대학교 통계학과

** 창원대학교 통계학과

*** 우석대학교 전산정보학부

로 발전하고 있으며, 모집단이 제한적인 회원이나 기업체의 구성원 등을 대상으로한 조사에서는 온라인 설문조사의 단점을 보완할 수 있기에 현실적으로 많이 사용하고 있다.

또한 최근 들어 기업 등에서 자사의 고객 특히 온라인 회원을 대상으로 혹은 자회사의 사원을 대상으로 기업의 경영 자문을 얻기 위해 온라인 설문조사가 현장에서 많이 행해지고 있으나, 온라인 설문조사 역시 기존의 설문조사와 마찬가지로 조사자가 응답자들의 프라이버시나 사생활과 관련된 민감한 정보를 얻고자 할 경우에 정확한 정보를 얻기란 쉬운 일이 아니다. 한편, 응답자들은 기존의 설문조사와 마찬가지로 온라인 상에서 민감한 질문을 직접적으로 받게 되면 혹시 자신의 비밀이나 사생활이 노출될까 의심하여 혹은 응답 후의 부당한 대우를 받게될까 두려워 정직한 응답을 꺼리게 된다.

따라서, 기업 의사결정에 대한 찬·반 여부, 공적 물건의 유용량 등과 같은 민감한 정보를 얻고자 할 경우 간접응답기법인 확률화응답기법을 온라인 설문조사에 적용해 볼 수 있다. 그러면, 응답자들의 응답이 확률장치를 이용하여 간접적으로 이루어지게 되므로 응답자가 자신의 신상에 대한 불안이나 개인 정보의 유출을 이유로 정확하지 않은 응답을 할 가능성을 줄일 수 있게 된다. 그러므로 온라인 설문조사를 하는 데 민감한 정보를 얻기 위한 확률화응답기법을 이용할 수 있는 시스템이 필요하다고 생각된다.

따라서, 본 연구에서는 민감한 정보를 얻기 위한 확률화응답시스템을 구현해 보고자 한다. 본 시스템은 기존의 온라인 설문조사 시스템과 더불어 사용할 수 있을 뿐만 아니라 독립된 스팟 서베이 등이 가능하도록 구현하고자 한다. 그리고 본 시스템은 동일한 응답자가 여러 번 답하는 것을 막기 위해 로그인(log in)을 하는 사이트에서는 동일 아이디에 대하여 중복 응답을 하는 것을 막을 수 있고, 로그인을 하지 않는 사이트에서는 동일 IP에서 중복 응답하는 것을 막을 수 있도록 하며, 학교 실습실 등과 같이 여러 명이 사용하는 경우에 대비하여 응답에 제한을 두지 않을 수 있도록 구현하고자 한다.

본 논문은 1장에서 서론으로 민감한 사항에 대한 온라인 설문조사에서 확률화응답기법의 필요성을 설명하고, 2장에서는 확률화응답기법을 소개하며, 3장에서는 구현된 확률화응답시스템에 대하여 살펴보고, 4장에서는 결론과 향후 연구과제를 다루고 있다.

II. 확률화응답기법

사회적으로 민감한 조사에서 응답자들이 응답을 회피하거나 정직하게 응답하지 않는 질문들에 대하여 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 내기 위하여 Warner(1965)는 확률장치를 이용하여 간접응답을 하도록 하는 확률화응답기법(randomized response technique ; RRT)을 처음으로 제시하였다. Warner는 응답자들에게 민

감한 질문과, 민감한 질문과 배반되는 질문으로 구성된 확률장치를 사용하여 민감한 속성에 대한 질적 정보를 얻고자 하였다. 그리고, Greenberg 외(1969)는 민감한 질문과 배반되는 질문 대신에 민감한 질문과 전혀 무관한 질문을 사용하는 무관질문기법(unrelated question technique)을 제안하였으며, Greenberg 외(1971)는 이를 양적속성기법으로 발전시켜 민감한 변수에 대한 양적 정보를 얻고자 하였다. 또한, Loynes(1976)는 Warner기법의 민감한 질문과 배반이 되는 질문 대신에 “예”라고 응답하도록 강요하는 강요질문기법(forced answer technique)을 제안하였다. 그 후 미국, 캐나다, 영국, 호주 등 서구 여러 나라와 일본, 인도 등 몇몇 아시아 국가에서도 이 분야에 대한 연구가 활발히 진행되고 있다. 특히, Fox와 Tracy(1986), Chaudhuri와 Mukerjee(1988)는 확률화응답기법들을 정리, 요약하여 체계화하였다. 최근에는 이들 이론들의 실제적 활용에 많은 관심이 집중되고 있으며, 사회학, 경영학, 의학 등 여러 학문분야에서의 조사활동에도 이의 활용이 적극 모색되고 있다.

국내에서는 류제복, 홍기학과 이기성(1993)들이 확률화응답기법에 관한 책을 출간하여 그 중요성을 강조하였고, 류제복, 이계오와 이기성(1995)은 확률화응답기법들이 적용된 사례들을 비교 분석하여 실용화를 위한 방안을 제시하였다.

대표적인 확률화응답기법들을 요약하면 <표 2.1>과 같다.

<표 2.1> 확률화응답기법의 분류

자료의 종류	종류		비고
질적 자료	관련질문기법		Warner(1965)
	무관 질문기법	무관한 속성의 모비율을 아는 경우	Greenberg 외(1969)
		무관한 속성의 모비율을 모르는 경우	
	강요질문기법		Loynes(1976)
양적 자료	무관 질문기법	무관한 변수의 모평균을 아는 경우	Greenberg 외(1971)
		무관한 변수의 모평균을 모르는 경우	

또한, 민감한 사항에 대한 실제조사에 확률화응답기법의 원활한 활용과 보다 정확한 정보를 얻기 위해서는 다음과 같은 과정을 거쳐 실제조사를 수행하는 것이 바람직하다.

- ① 조사하고자 하는 내용이 민감한가를 조사 : 확률화응답기법은 민감한 사항에 대한 조사에 효율적이므로 1차적으로 질문의 민감성을 조사한다. 민감성은 지역, 성별, 연령 등 조사대상에 따라 달라질 수 있으므로 사전조사가 필요하다.
- ② 확률화응답기법의 선택 : 확률화응답기법의 실용화에 가장 큰 걸림돌 중의 하나가 조사자와 응답자들이 이 기법에 대한 이해의 부족이라 볼 수 있다. 따라서 이들의 이해가

- 용이하고 응답자들의 익명성을 최대한 보장해 줄 수 있는 기법을 선택한다.
- ③ 조사방법의 선택 : 민감한 사항에 대한 조사를 면접조사, 우편조사, 전화조사, 인터넷조사 등에서 시간, 비용 등을 고려하여 선택한다. 그러나 조사방법에 따라 확률화응답기법과 확률장치들이 달라질 수 있으므로 이점을 함께 생각하여 선택한다.
 - ④ 확률장치의 선택 : 확률장치는 응답자들에게 친숙하고 사용에 간편한 것을 선택한다. 왜냐하면 확률장치를 사용하는 조사에 대부분의 응답자들이 익숙하지 않고 호기심과 의혹을 갖고 이것이 일종의 속임수가 아닌가 하는 생각을 갖게 된다. 그러므로 응답자들의 익명성을 충분히 보장해 준다는 믿음을 갖을 수 있는 확률장치를 선택한다.
 - ⑤ 확률화응답기법과 확률장치에 대한 올바른 설명 : 확률장치에 대한 사용방법과 확률화응답기법이 응답자들의 신분을 확실히 보장해 줄 수 있다는 설명과 추가로 이해가 부족한 사람들을 위하여 간단한 예가 첨부된 안내문을 작성한다.

Ⅲ. 확률화응답시스템의 구현

1. 시스템 개발 환경 및 시스템 흐름

구현된 시스템의 개발 환경에서 개발 언어는 GCC, Java, HTML 등이며, 운용 환경은 Linux용으로 개발하였다. 또한 DB는 MySQL을 이용하였다.

확률화응답시스템은 관리자(조사자) 모드와 응답자 모드의 두 가지 모드로 구성되어 있다. 관리자 모드에서는 설문을 작성하는 에디터와 확률장치의 선택 및 민감한 설문이 선택될 확률을 입력하는 부분으로 이루어져 있고, 기타 계산에 필요한 기타 정보를 입력하도록 되어 있다. 응답자 모드는 실제 응답자가 응답할 수 있도록 이루어져 있으며, 무관질문기법에서 무관한 속성에 대한 모비율을 모를 때와 무관한 변수에 대한 모평균을 모를 때는 접속 순서에 따라 표본 1과 표본 2로 나누어져 응답에 참여하도록 하였다. 응답의 결과는 DB로 저장되며, 민감한 속성(변수)에 대한 모비율(모평균)의 추정값과 분산추정값을 계산한 후에 응답자에게는 민감한 속성(변수)에 대한 모비율(모평균)의 추정값만을 보여주고, 관리자에게는 모비율(모평균)의 추정값과 분산추정값에 대한 결과를 모두 보여 주도록 구성하였다.

본 시스템은 자료의 입력에서 처리, 결과를 모두 DB를 바탕으로 이루어져 있다. 이로 인하여 동일 응답자 등의 반복 측정에서도 DB를 사용함으로 인하여 기존의 설문응답시스템과 쉽게 합쳐서 사용할 수 있다. 또한 처리과정에서도 DB를 사용함으로 인하여 쿼리(query)를 사용하여 수행 속도면에서 파일시스템보다 속도에서 보다 빠르게 진행할 수 있다. 또한 기본적인 결과를 DB를 연동함으로 지속적인 조사 등에서 추세 분석이 가능하다.

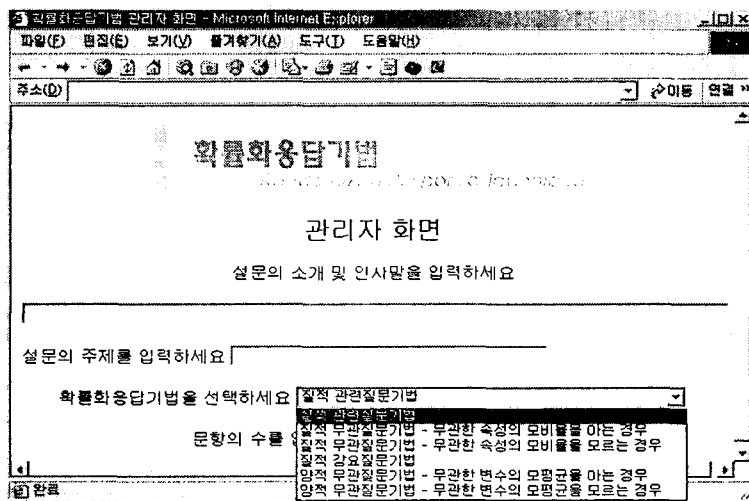
본 시스템의 구현을 위해 설계한 테이블의 항목은 <표 3.1>과 같다.

<표 3.1> 시스템 DB 테이블

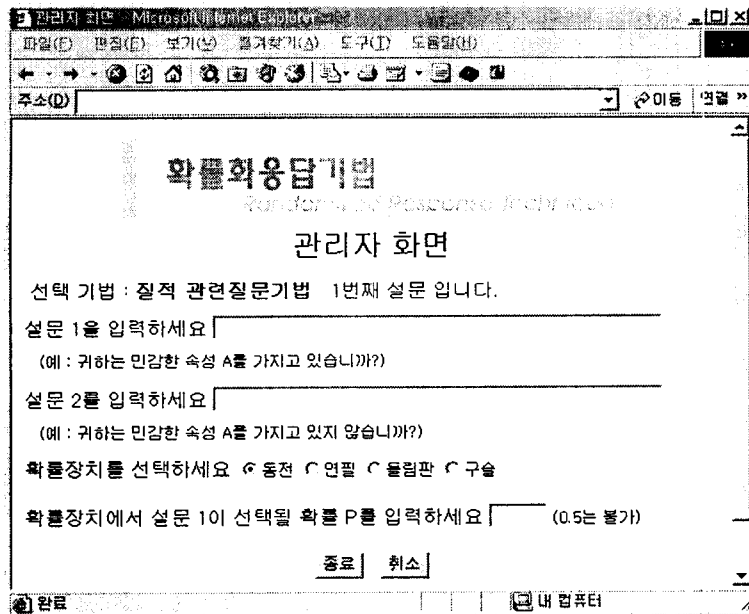
테이블 : Survey_Item (설문지별 구성 문항정보)			
Logical Name	Physical Name	Data Type	비 고
회원 아이디	Id	Varchar	PK, Index
독립 문항 Index	Spot_number	Integer	PK, Index
확률	Probability	Float	
확률장치 Type	Probability_type	Integer	
설문일자	Day	Date	
RRT type	Rrt_type	Integer	
응답	Response	Integer	

2. 시스템 예

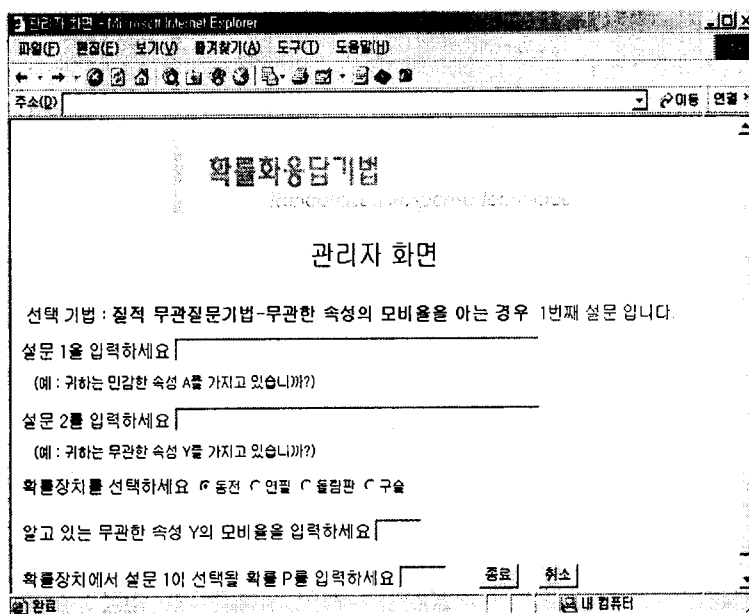
본 시스템은 관리자 모드와 응답자 모드의 두 부분으로 이루어져 있다. 관리자 모드로 접속하여 <그림 3.1>과 같이 설문지 소개, 설문지 주제, 확률화응답기법의 종류, 문항의 수를 입력한 후 계속을 누르게 되면 다음 단계로 선택된 확률화응답기법에 따라 <그림 3.2>에서부터 <그림 3.7>까지 중에 하나에서 설문 문항을 입력할 수 있는 에디터가 나타나며, 설문 내용과 확률장치 등의 내부 사항을 모두 입력하면 설문지 만들어진다.



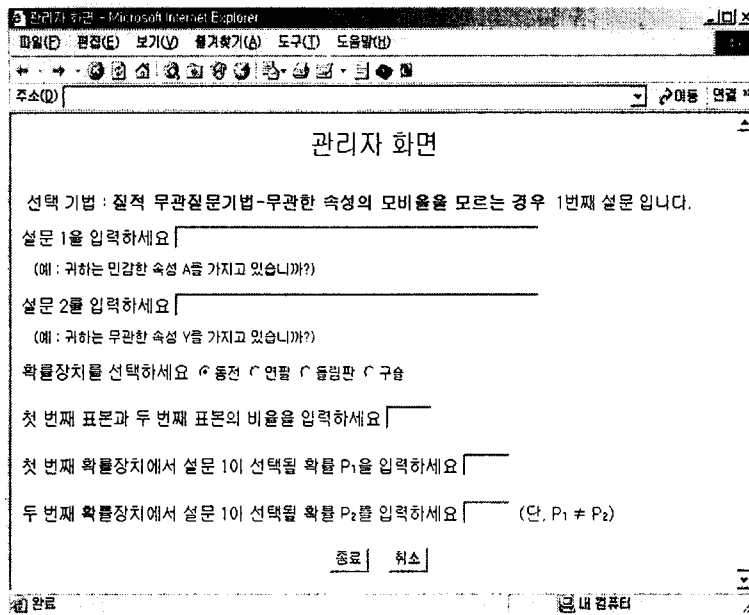
<그림 3.1> 관리자 초기화면



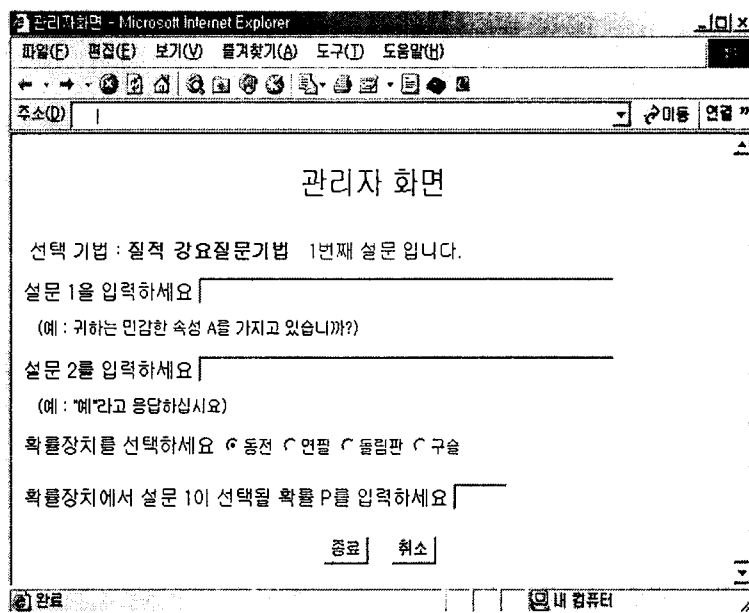
<그림 3.2> 질적 관련질문기법 관리자 화면



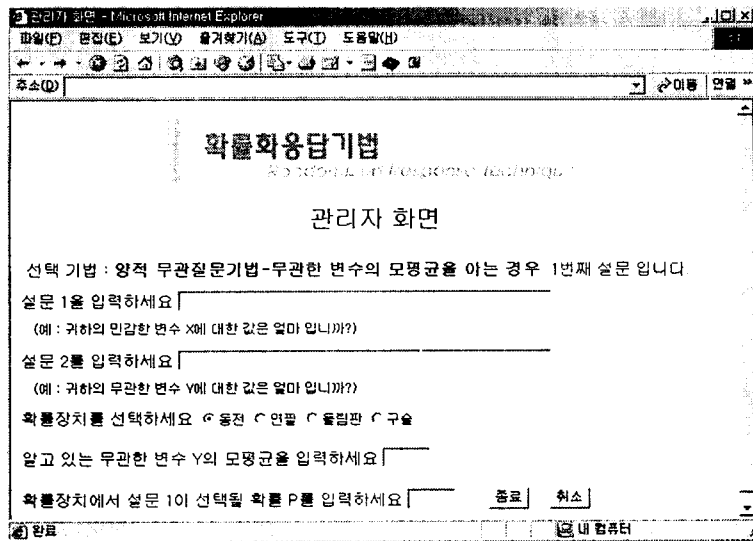
<그림 3.3> 질적 무관질문기법(무관한 속성의 모비율을 아는 경우) 관리자 화면



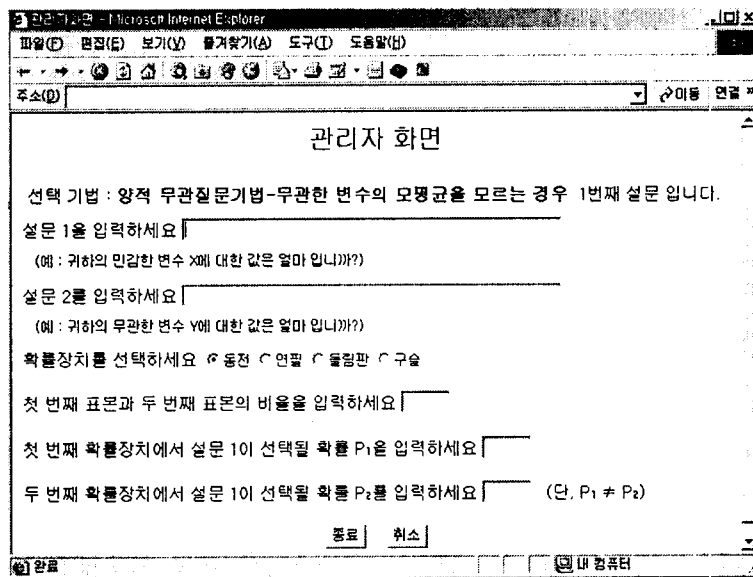
<그림 3.4> 질적 무관질문기법(무관한 속성의 모비율을 모르는 경우) 관리자 화면



<그림 3.5> 질적 강요질문기법 관리자 화면



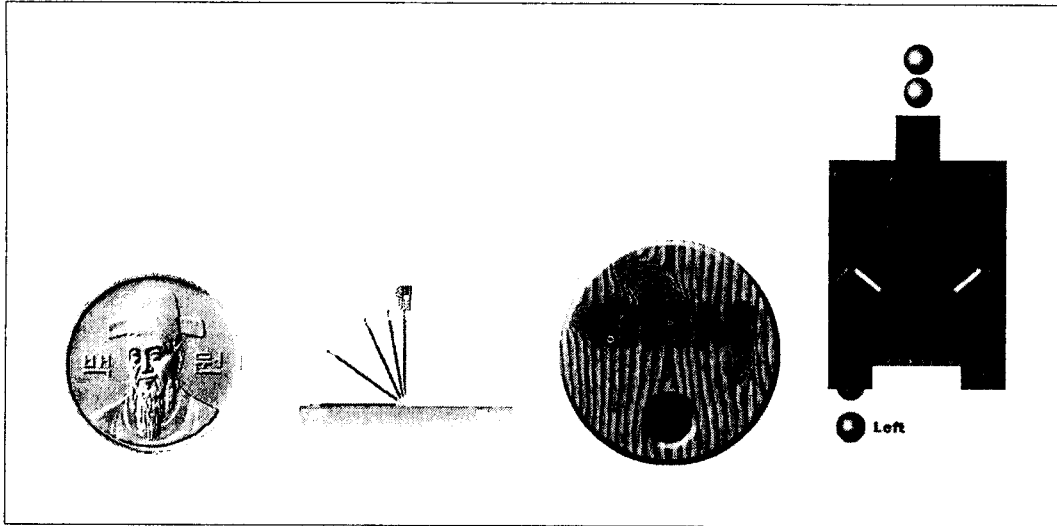
<그림 3.6> 양적 무관질문기법(무관한 변수의 모평균을 아는 경우) 관리자 화면



<그림 3.7> 양적 무관질문기법(무관한 변수의 모평균을 모르는 경우) 관리자 화면

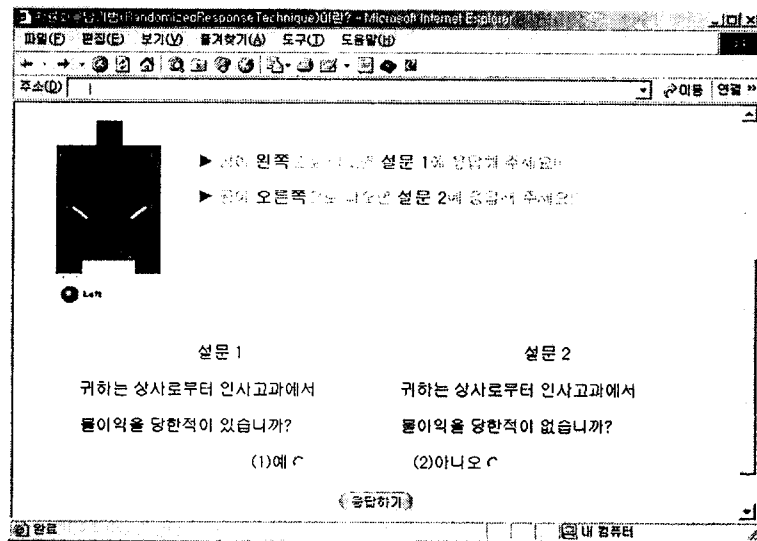
본 시스템은 4가지의 확률장치 중 하나를 선택하도록 되어 있으며, <그림 3.8>은 이들 4가지를 나타낸 것이다. 이들 확률장치는 관리자 화면에서 선택할 수 있으며, 확률장치에서 설문 1이 선택될 확률을 입력하게 되면 그 확률에 따라 동전의 앞면 혹은 뒷면, 오른쪽 혹은 왼쪽

과 같이 두 가지 중에 하나의 형태로 나타난다.



<그림 3.8> 4종류의 확률장치

<그림 3.9>는 응답자가 설문 사이트에 접속하였을 때 나타나는 화면이다.

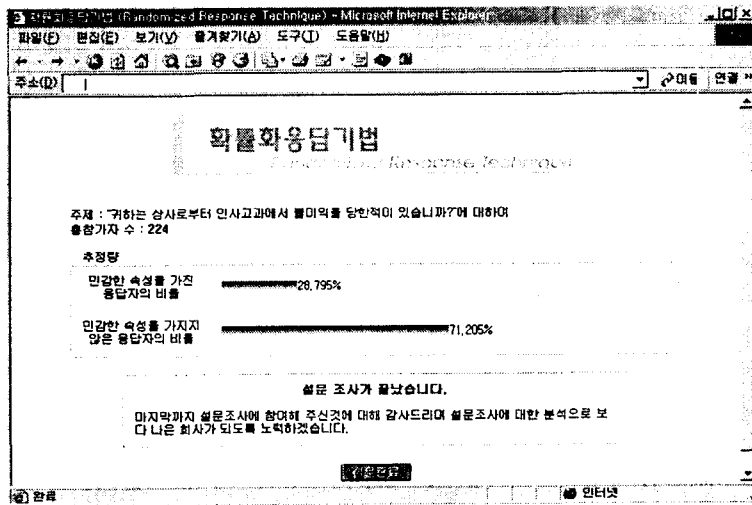


<그림 3.9> 관련질문기법의 경우 응답자 화면

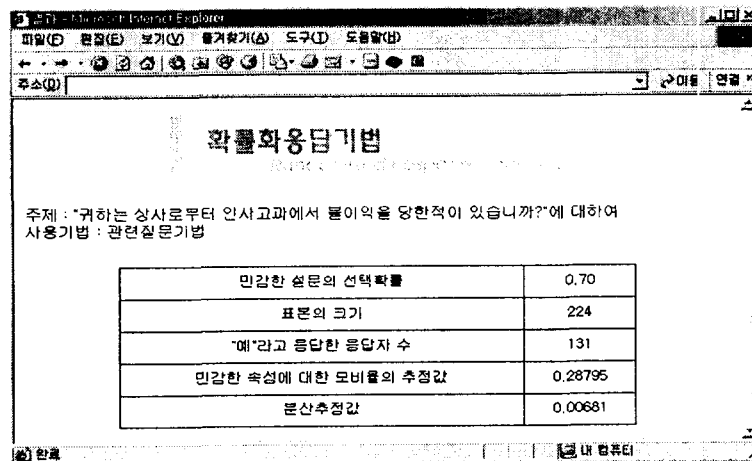
조사자는 응답자가 설문 1에 대하여 “예” 라고 응답을 하였는지 설문 2에 대하여 “예”라고 응답을 하였는지 알 수 없도록 설계되어 응답자가 진실되게 응답할 수 있도록 하였다.

3. 응답결과

응답결과는 응답자용과 관리자용으로 구분되어 있으며, 응답자용은 <그림 3.10>과 같이 민감한 속성 A의 모비율의 추정값만을 보여주고 있다. 관리자는 <그림 3.11>과 같이 확률화응답기법을 이용한 응답자 수, 민감한 속성 A의 모비율의 추정값, 모분산의 추정값 등을 볼 수 있도록 구현하였다. 또한 다른 설문과 같이 복합적으로 이용할 때는 문항에 따른 즉, 남·녀별 모평균 등을 비교할 수 있도록 구현되었다.



<그림 3.10> 응답자용 결과



<그림 3.11> 관리자용 결과표

IV. 결론 및 향후과제

본 논문에서는 민감한 정보를 얻기 위한 조사에서 응답자들이 정직하게 응답하기를 꺼리는 질문들에 대하여 직접응답 대신에 간접응답을 통해 응답자의 비밀을 노출시키지 않고서 보다 정확한 정보를 얻을 수 있는 간접응답기법인 확률화응답기법을 인터넷 상에서 사용할 수 있도록 구현하였다. 이는 기업체 등에서 개인의 비밀이나, 응답후의 불이익을 염려하여 거짓 응답하는 것을 조금이나 방지하여 의사결정을 위한 보조 도구로서도 사용할 수 있을 것으로 예상된다.

본 시스템은 기존의 설문조사 시스템과 연계하여 민감한 질문에만 확률장치를 이용할 수 있도록 하여 다른 속성에 따라 민감한 질문에 대한 차이도 볼 수 있을 뿐만 아니라 독립된 단일문항 질문으로도 사용이 가능하도록 하였다.

향후의 과제로는 다양한 표본추출방법을 적용하여 회원이나 구성원들을 대상으로 목적에 맞는 응답자를 추출하는 기능을 부과하여 기업 등에서 사원이나 고객을 대상으로한 조사에서 실용성을 보다 높이는 방법이 필요하다고 생각된다. 또한 다양한 질문에 대한 대처 방법과 응답자에게 흥미를 유발할 수 있는 디자인을 도입한 확률장치의 고안이 필요하다.

참고문헌

1. 김정기 · 김희재 · 남기성 · 박희창 · 이성철 · 정정현. 1999. 《사회조사분석론》 (1판). 창원대학교출판부.
2. 류제복 · 홍기학 · 이기성. 1993. 《확률화응답모형》 (1판). 자유아카데미.
3. 류제복 · 이계오 · 이기성. 1994. "확률화응답기법의 실용화 방안." 《응용통계연구》 8(1) : 9-26.
4. Chaudhuri, A. and Mukerjee, R. 1988. *Randomized Response : Theory and Techniques*(1st ed.). New York : Marcel Dekker, Inc..
5. Coomber, R. 1997. "Using the Internet for Survey Research." *Sociological Research Online*, 2(2). <<http://www.socresonline.org.uk/socresonline/2/2/2.html>>
6. Fox, J. A. and Tracy, P. E. 1986. *Randomized Response : A Method for Sensitive Survey*, Sage Publications.
7. Greenberg, B. G., Abul-Ela, Abdel-Latif A., Simmons, W. R., and Horvitz, D. G. 1969. "The Unrelated Question Randomized Response Model : Theoretical Framework." *Journal of the American Statistical Association* 64 : 520-539.
8. Greenberg, B. G., Kubler, R. R., Abernathy, J. R., and Horvitz, D. G. 1971. "Applications of the RR Technique in Obtaining Quantitative Data." *Journal of the American Statistical Association* 66 : 243-250.
9. Schwarz, C. J. 1997. "StatVillage : An On-line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling." *Journal of Statistics Education* 5(2). <<http://www.amstat.org/publications/jse>>
10. Warner, S. L. 1965. "Randomized Response ; A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60 : 63-69.