

Poll System using E-mail¹⁾

Yon Hyong Kim²⁾ Min Gweon Oh³⁾

Abstract

In this paper we propose a poll system using e-mail. This system expects to increase the response ratio because of including a questionnaire inner e-mail. Especially, this system automatically provides a general paper which is a result of categorical data analysis.

Keywords : CGI, e-mail, internet, sampling, system, survey

1. 서 론

인터넷 여론조사는 면접조사, 우편조사, 전화조사 방법에 비해 전 세계로 연결되어 있는 네트워크를 이용하여 시간이나 장소의 제한에서 벗어나 신속하게 많은 응답자를 확보할 수 있다는 장점 때문에 국내 유관기관 및 여론조사 기관에서 인터넷 여론조사를 시행하고 있다. 그러나 인터넷이라는 특수한 매체를 통해 조사가 이루어지므로 지역별, 연령별, 직업별 등의 여러 범주들에 대한 분석을 하기에는 표본의 대표성 문제가 제기되고 있다.

김연형, 오민권(2000)은 설정된 모집단에 대해 표본오차와 허용오차를 이용하여 표본수를 결정한 후 인터넷 홈페이지를 통해 응답한 표본들 중에서 비례배분 방법으로 배정된 표본들만 분석 자료로 활용함으로서 여론조사에 대한 신뢰성 및 표본의 대표성 문제를 보안한 여론조사 시스템을 제안하였다. 또한, 기존의 표본조사 방법에 입각한 인터넷 여론조사 시스템을 제안하였는데, 제안된 시스템은 표본추출방법에 의한 응답자 확보, 이차원 자료에 대한 분할표 및 독립성 검정결과 제공, 고차원 자료에 대한 범주형 자료분석 결과가 실시간에 제공된다는 장점이 있다.

이러한 홈페이지를 이용한 여론조사 방법은 인터넷 이용자수가 1천만명이 넘어섰다고는 하지만 인터넷 여론조사는 싸이버 공간이라는 한정된 범위 안에서 조사가 이루어지므로 인터넷의 이용률에 많은 제한을 받을수 밖에 없다. 특히, 홈페이지를 이용한 여론조사는 네티즌이 관련 홈페이지에 접속하여 응답해야하므로 많은 응답을 확보하기 위해서는 경품행사 등의 부가 서비스를 제공해야 한다.

국내 인터넷 이용자중 한달에 한번 이상 사용하는 이메일 보유율은 60.4%(한국인터넷정보센터 2000년 11월 기준)로 인터넷을 이용하는 네티즌은 대부분이 이메일을 보유하고 있으며 특히, 많은

1) This research was supported by the Jeonju University Research Fund in 2001.

2) Professor, Department of information statistics, Jeonju University, Jeonju, 560-759, Korea.
E-mail : yhkim@www.jeonju.ac.kr

3) Additional Instructor, Dept. of information statistics, Jeonju University, Jeonju, 560-759, Korea.

닷컴(dot com) 회사들은 기업 홍보 및 마케팅 전략의 일환으로 회원들의 이메일을 적극적으로 이용하고 있다.

여론조사에서도 이메일이 이용되고 있는데, 현재 사용되는 방법은 여론조사 의뢰자가 응답요청 이메일을 작성한 후 설문지(아래한글 또는 워드 문서)를 이메일 문서에 파일로 첨부하여 발송하면, 이메일 수신인이 해당 문서편집 소프트웨어에서 설문지를 오픈하여 각 문항에 응답/저장한 후 의뢰자에게 응답한 설문문서를 첨부하여 이메일을 발송해야 한다. 의뢰자는 메일로 접수된 각각의 응답된 설문을 출력 후 코딩과정을 거쳐 SAS, SPSS 등의 통계 패키지를 이용하여 분석하는 복잡한 절차를 거쳐야 한다.

<표 1> 성별, 연령별, 직업별 이메일 보유율 및 이용개수

		보유율	이 용 개 수				
			1개	2개	3개	4개	무응답
전 체		60.4	58.8	18.6	8.1	7.1	7.3
성 별	남 성	60.1	56.5	19.0	9.1	7.7	7.7
	여 성	60.8	61.8	18.2	6.8	6.3	6.9
연령 별	7-19세	58.3	64.0	16.3	6.7	6.1	6.9
	20대	70.0	51.5	22.6	10.2	9.4	6.3
	30대	56.0	61.7	15.8	6.8	5.0	10.7
	40대 이상	43.9	65.5	15.9	7.1	4.8	6.7
직업 별	초·중학생	45.7	65.1	15.1	6.6	5.3	8.0
	고등학생	69.4	65.7	17.9	5.7	5.1	5.6
	대학생	81.4	52.5	21.4	11.3	8.6	6.2
	농립어업	23.3	100.0	.0	.0	.0	.0
	자영업	41.8	63.1	17.3	5.5	5.4	8.7
	블루칼라	49.2	66.8	13.2	2.5	8.6	8.9
	화이트칼라	62.7	53.4	19.8	9.3	9.7	7.7
	주부	49.3	66.9	14.4	6.4	2.3	10.0
	무직/기타	67.9	58.0	24.6	6.4	3.3	7.7

<출처 : 한국인터넷정보센터>

본 연구에서는 여론조사 의뢰기관과 전문 여론조사 기관에서 보유하고 있는 회원 DB에서 설문 조사의 목적에 부합한 표본을 추출한 후 자동으로 설문 이메일이 발송되고, 메일 수신인이 이메일 문서내에 존재하는 설문에 응답하면 자동으로 서버에 응답자료가 전송되어 실시간에 의뢰자에게 통계적 분석 결과가 제공되는 이메일 여론조사 시스템을 제안하고자 한다.

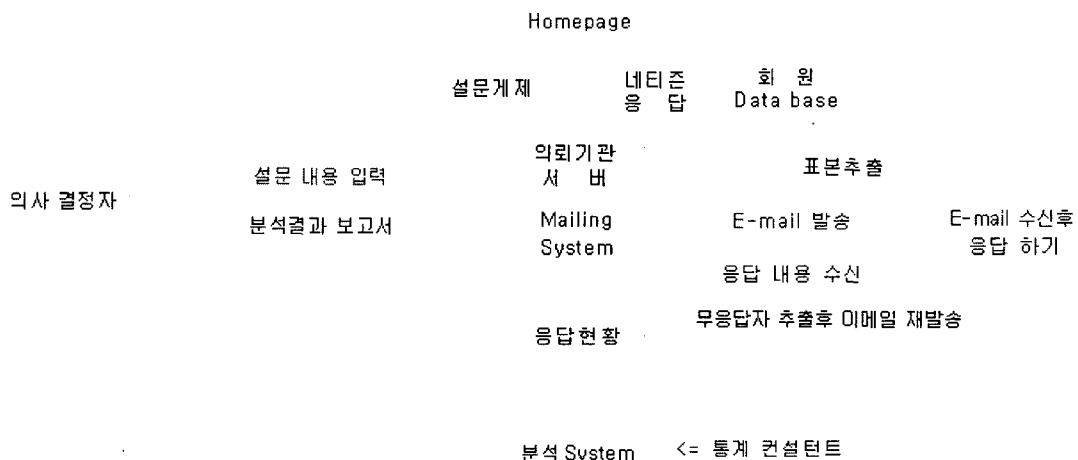
2. 여론조사 시스템

개발된 시스템은 널리 이용되고 있는 NT 서버환경에서 MySQL 데이터베이스와 웹 프로그래밍

언어인 ASP, Perl 그리고 Java Script를 사용하여 구현하였다.

<그림 1>은 개발된 시스템의 구성도로서 이메일 여론조사 시스템은 크게 회원DB에서 표본을 추출하는 부분과 추출된 표본 회원들에게 자동으로 이메일을 발송하는 메일링 부분 그리고 메일 수신인이 응답한 자료의 분석부분으로 구성된다.

개발된 이메일 여론조사 시스템은 9개의 원도우(설문작성, 설문발송, 응답현황, 통계분석, 보고서 생성, DB검색, 회원현황, 메일발송, 도움말)로 구성되어 있는데, 각 원도우에 대한 세부적인 내용은 다음과 같다.



<그림 1> 이메일 여론조사 시스템 구성도

2.1 설문지 작성

본 연구에서 제안한 시스템에서는 사용할 수 있는 설문 문항은 기타 의견 수렴으로 주관식 입력을 받는 1개의 문항을 포함하여 총 11개이며, 10개의 일반 문항은 각각 5개 이하의 범주를 가질수 있도록 설계하였다. 만일 10문항 이상의 여론조사를 실시할 경우에는 일정 기간을 두고 설문 문항을 나누어 설문 조사를 수행할 수 있는데, 이 경우 이메일 주소가 키워드로 사용된다.

사용자가 시스템이 제공하는 일정 형식에 입력한 설문 내용은 시스템의 CGI 프로그램에 의해 자동으로 HTML 문서로 변환되어 특정 코드번호를 파일명으로 하여 저장되는데, 코드번호는 응답 자료와 추후 설문 문항에 대한 통계적 분석시 분류기준이 된다.

2.2 표본설계 및 메일 발송

많은 회원을 보유하고 있는 닷컴회사 뿐만 아니라 대부분의 인터넷 관련 기업들은 회원가입 시 성별, 연령, 거주지, 거주형태, 직업, 관심분야, 결혼 여부 등 다양한 회원 정보를 요구하여 체계화

된 회원 데이터 베이스를 보유하고 있다. 회원 DB는 여론조사의 표본 설계에 있어 매우 유용하게 사용될 수 있는데, 제안된 시스템에서는 세 가지 방법으로 회원 DB가 표본 설계에 사용된다.

첫째, 모든 회원을 표본으로 활용하는 방법
 둘째, 이메일을 발송할 때 표본 추출 단위로 활용하는 방법
 셋째, 응답회원에 대해 비례배분 방법으로 Resampling 하는 방법

성별	<input checked="" type="radio"/> 남자 <input type="radio"/> 여자
연령	<input checked="" type="radio"/> 10대 <input type="radio"/> 20대 <input type="radio"/> 30대 <input type="radio"/> 40대 <input type="radio"/> 50대이상
직업	<input type="text" value="==선택하세요=="/>
거주지	<input type="text" value="==선택하세요=="/>
학력	<input type="text" value="==선택하세요=="/>
거주형태	<input type="text" value="==선택하세요=="/>
거주주택	<input type="text" value="==선택하세요=="/>
결혼	<input checked="" type="radio"/> 미혼 <input type="radio"/> 기혼

전체 회원에게 메일 보내기

<그림 2> 작성한 설문 이메일의 수신인(표본)을 선택하는 예제

본 연구에서 구현된 시스템에서 표본선택은 사용자가 직접 표본 추출률, 표본크기, 표본배분, 표본추출 과정을 고려하여 위의 세 가지 방법을 적절하게 선택하여 사용할 수 있도록 설계하였다. 사용자가 설문 작성 후 <그림 2>에서 이메일 발송 표본을 추출한 후 『메일보내기』 버튼을 클릭하면 작성된 설문 내용이 이메일 문서내에 자동으로 포함되어 선택한 회원들에게 이메일이 발송된다.

2.3 응답현황 및 이메일 재발송

구현된 시스템은 자동으로 응답현황에 대한 정보를 실시간에 제공하는데, 설문이 포함된 이메일을 수신한 회원들이 설문에 응답했는지 여부를 파악하여 응답하지 않은 회원에게는 자동으로 이메일을 재 발송할 수 있다. <그림 3>은 수행중인 여론조사 설문에 대한 응답 현황에 대한 결과로 각 인구통계변수들에 대한 항목별 응답자수와 응답률이 저시되어 있다.

■ 현재 2001년 2월 11일 19시 전체회원 1956 명중 983명 응답으로 50.25 % 응답률

1. 성별에 따른 응답 현황

구 분	회원수	응답자수	비 율	히스토그램
남자	1376 명	759 명	55.15 %	
여자	580 명	224 명	38.62 %	

증략

8. 결혼유·무에 따른 응답 현황

구 분	회원수	응답자수	비 율	히스토그램
미혼	1423 명	575 명	40.4 %	
기혼	533 명	408 명	76.54 %	

[\[돌아가기\]](#) [\[무응답 회원에게 이메일 재발송\]](#)

<그림 3> 설문 응답 현황 예제

2.4 응답자료에 대한 통계분석

시스템의 통계분석 부분은 크게 <그림 4>와 같이 기초분석과 세부분석의 두 부분으로 이루어져 있는데, <그림 5>와 같이 각 문항별 응답도수와 비율이 제공되는 일차원 분석이 기초분석에 해당된다. <그림 6>와 <그림 7>은 각각 시스템이 제공하는 세부분석 결과의 예로서 기초 통계변수(성별, 연령별, 직업별, 거주지별, 주거형태별, 결혼 유무 등)와 각각의 설문 문항에 대한 이차원 분할표와 이차원 범주형 자료분석(독립성 및 동질성 검정)의 결과이다. 삼차원 이상의 고차원 범주형 자료분석이 요구될 때에는 홍종선, 김동욱, 오민권(1999)이 제안한 순수 자바로 구현된 범주형 자료분석 시스템을 이용할 수 있다.

설문 제목	종료일자	기초분석	세부분석
1. 우리나라 IT 산업에 대한 전망	2001년 2월 10 일	[보기]	[보기]

<그림 4> 수행중인 여론조사의 통계분석 목록

▣ 설문 제목 : "우리나라 IT 산업에 대한 전망"에 대한 분석자료 결과

1. 우리나라 IT 산업의 발전 전망에 대한 귀하의 의견은?	응답자	비율	히스토그램
① 매우 밝다	139 명	14.14 %	
② 밝다	329 명	33.46 %	
③ 보통이다	422 명	42.92 %	
④ 어둡다	51 명	5.18 %	
⑤ 아주 어둡다	42 명	4.27 %	
2. 앞으로 가장 각광 받는 IT 산업은?	응답자	비율	히스토그램
① 소프트웨어 산업	429 명	43.64 %	
② 게임 산업	322 명	32.75 %	
③ 기타 컨텐츠 사업	232 명	23.6 %	

[\[돌아가기\]](#)

<그림 5> 1차원 분석 결과 예제

성별 ▶ 우리나라 IT 산업의 발전 전망에 대한 귀하의 의견은? ▶ 분석하기

빈도수 전체비율 열비율 행비율	매우 밝다	밝다	보통이다	어둡다	아주 어둡다	Total
남자	115 명 11.69 % 82.73 % 15.25 %	224 명 22.78 % 68.08 % 29.7 %	331 명 33.67 % 78.43 % 43.89 %	46 명 4.67 % 90.19 % 6.1 %	38 명 3.86 % 90.47 % 5.03 %	754 명 76.7 % 100 %
	24 명 2.44 % 17.26 % 10.48 %	105 명 10.68 % 31.91 % 45.85 %	91 명 9.25 % 21.56 % 39.73 %	5 명 0.5 % 9.8 % 2.18 %	4 명 0.4 % 9.52 % 1.74 %	229 명 23.29 % 100 %
	139 명 14.14 % 100 %	329 명 33.46 % 100 %	422 명 42.92 % 100 %	51 명 5.18 % 100 %	42 명 4.27 % 100 %	983 명 100 %

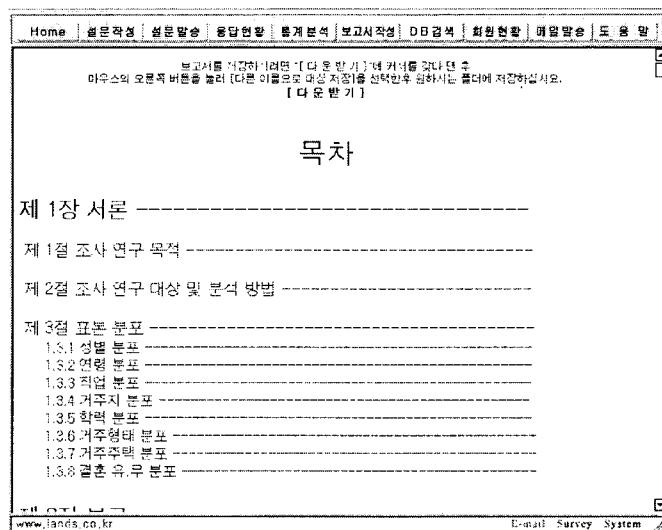
<그림 6> 2차원 분석 결과 예제

당신이 선택한 두 변수의 독립성 검정(χ^2 -검정) 결과	
지금까지 응답한 누적된 자료를 토대로 당신이 선택한 성별과 설문문항의 각 항목[범주]에 대한 비율이 같은지를 판단하기 위해 다음과 같은 가설을 설정할 수 있습니다.	
귀무가설 : [성별]과 [설문문항]의 각 항목에 대한 비율이 같다(또는 독립이다). 대립가설 : [성별]과 [설문문항]의 각 항목에 대한 비율이 같지 않다(또는 종속이다).	
위의 가설에 대한 자유도(DF: Degrees of Freedom)는 4이고, χ^2 은 26.8676752753947이며, 이에 대응되는 p-값은 2.11417171833821e-05이다.	
따라서 유의수준(Alpha)을 0.05로 정하는 경우 유의수준(0.05) > p-값(2.11417171833821e-05) 이므로 귀무가설을 기각합니다.	
즉, 두 변수의 각 항목에 대한 비율이 같다고 할 수 없습니다[통계적 종속].	

<그림 7> 독립성 가설 검정 결과

2.5 자동 보고서 작성

일반적으로 여론조사 결과를 보고서로 작성하는 과정은 분석결과를 도표화하거나 그림으로 재구성하고 유의한 결과에 대해 서술해야 하기 때문에 많은 시간이 요구된다. 이러한 실무적인 문제를 해결하기 위해 시스템에서는 『보고서 작성』 기능을 제공하는데, 이는 사용자의 요구에 의해 자동으로 목차, 서론, 본론(각 문항별 1차원 범주형자료분석 결과와 인구통계 변수에 대한 각 문항별 2차원 범주형자료분석 결과 포함), 결론(유의한 통계분석 결과 포함)으로 구성된 보고서를 제공한다. 시스템에서 제공하는 보고서는 일반화된 문서 형태로 제공되기 때문에 사용자는 보고서를 훈글 문서로 다운받아 보고서 내용을 수정 및 보완할 수 있다. <그림 8>은 솔루션에서 제공하는 보고서 생성 기능의 예제이다.



<그림 8> 보고서 자동생성 기능의 예제

2.6 여론조사 결과의 데이터베이스

시스템의 검색 기능은 크게 설문지 DB 검색, 여론조사 결과 DB 검색, 각종 보고서 검색 및 다운로드로 분류된다. 사용자가 여론조사에 착수하기 전에 비슷한 연구목적의 여론조사가 수행되었는지 여부를 파악하고 선행 조사에 대한 설문지, 보고서, 보도자료를 열람할 수 있다면 사전조사에 필요한 많은 유용한 정보를 얻을 수 있을 것이다.

<그림 9>은 국내 언론사 및 각 기관에서 수행한 약 5,000개의 여론조사 결과를 14개의 분야(연예, 예술, 여가, 산업, 정치, 사회, 지역, 컴퓨터, 언론, 교육, 직업, 건강, 환경, 의식)로 분류하여 구축된 데이터 베이스의 검색 화면으로, 약 3만개의 다양한 단어에 대한 검색 기능을 제공한다.

여론조사 결과 검색 단어				찾기버튼	
구 분	자료수	구 分	자료수	구 분	자료수
자동차	246	연예/모탁	304	예술/문화	278
여가/스포츠	272	산업/경제	378	정치/행정	369
사회/생활	528	지역정보	304	컴퓨터/인터넷	436
뉴스/언론	249	교육/학습	523	직업/취업	259
건강/의학	269	환경	233	의식	237

<그림 9> 여론조사 결과 데이터베이스 검색 목록

또한, 분석자는 종료된 여론조사의 보고서를 작성할 경우 시스템의 검색기능을 이용하여 비슷한 목적의 선행조사를 검색 후 결과를 비교하여 시계열적인 의미가 부여되도록 보고서를 작성하는데 유용한 정보를 얻을 수 있을 것이다.

2.7 회원현황

표본을 추출하는데 있어 가장 중요한 것 중에 하나는 기초통계변수들(성별, 연령, 거주지, 거주형태, 직업, 결혼 여부 등)에 대한 회원들의 현황을 파악하는 것이다. 이를 변수들에 대한 현황은 시스템에서 최초 이메일을 발송하기 위해서 표본을 추출하거나 전체 응답회원들 중에서 총화추출이나 네이만 최적배분 등의 방법으로 Resampling 할 경우 기준이 된다.

<그림 10>는 시스템이 제공하는 회원현황의 예제로 각 인구통계 변수들에 대한 항목별 회원수 및 비율이 제공된다.

1. 성별에 따른 회원 현황

구 분	회원수	비 률	히스토그램
남자	1370 명	70.36 %	
여자	577 명	29.63 %	
:			
중략			

2. 결혼유무에 따른 회원 현황

구 분	회원수	비 률	히스토그램
미혼	1414 명	72.62 %	
기혼	533 명	27.37 %	

<그림 10> 현재까지 등록된 회원 현황 예제

2.8 도움말

도움말 원도우는 본 연구에서 제안한 솔루션 사용자에게 솔루션의 여러 기능들에 대한 설명과 사용방법을 제공한다. 특히 혼동하기 쉬운 여러 통계용어들에 대해서 한국통계학회에서 발간한 통계용어사전의 알파벳순으로 구성된 영문용어와 국문용어를 찾아볼 수 있도록 하였다.

3. 결 론

본 연구에서는 이메일을 이용한 여론조사 시스템을 제안하였다. 제안된 시스템은 기존의 표본추출방법에 의한 표본추출, 표본에 대한 자동 이메일 발송, 이메일을 통한 응답자료의 자동 전송, 실

시간 응답자 현황 파악, 무 응답자에 대한 자동 이메일 재발송, 응답자료에 대한 다양한 실시간 분석결과제공(이차원 자료에 대한 분할표 및 독립성 검정결과 제공) 등의 기능을 포함하고 있다. 제안된 시스템은 회원 DB를 이용하여 사용되기 때문에 회원수가 부족한 기관에서 사용하기에는 부적절하고 많은 회원수를 확보하고 있는 사이트와 연계하여야 하는 부담이 있지만 인터넷 홈페이지를 이용한 여론조사보다는 기능적인 측면과 방법론에서 많은 이점이 있다. 본 논문에서 제안한 여론조사시스템은 <http://www.mystat.co.kr/solution/poll/index.html>로 접속하여 직접 사용할 수 있다.

참고문헌

- [1] 남궁 평, 홍종선 (1996). 범주형 자료분석, 탐진.
- [2] 이규석 (1997). CGI 프로그래밍 언리쉬드, 정보문화사.
- [3] 이계오, 이해용, 김기환, 김영원, 변종석, 김광용, 홍종필, 제병환, 김창영, 오민권, 김연형 (2001). 인터넷 조사, 나남출판사.
- [4] 홍종선 (1995). 대수선형모형, 자유아카데미.
- [5] 김연형, 오민권(2000). *An Internet Poll System*, 「한국통계학회논문집」, 제7권 3호, 927-936.
- [6] 오민권(2000). 인터넷 여론조사의 향후 발전 방향, 「조사통계연구회 발표 논문집」
- [7] 이해영, 김기환(2000). *Internet Survey Methodology*, 「한국통계학회논문집」, 제7권 3호, 945-953.
- [8] 홍종선, 김동욱, 오민권(1999). 인터넷에서 범주형자료분석 시스템개발, 응용통계연구, 제12권 1호, 83-95.
- [9] Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, John Wiley and Sons.