

# 교육기관 홈페이지 개인정보 노출 유형 및 방지방안

- 행정안전부

## □ 추진배경

- 2011년 개인정보 노출 모니터링 분석 결과 초중고 및 대학 등 교육기관 홈페이지를 통한 주민번호 노출 사례가 상반기에 많이 발생
  - 합격자 및 신입생 명단 등 각종 명단을 홈페이지에 게시하면서 주민번호 등 개인정보 노출
- 국회, 언론 등에서 홈페이지 개인정보 노출 문제를 지속적으로 제기
  - ※ (국회)공공기관 홈페이지 개인정보 노출 심각, 대책마련 시급(국감, '11.9.20)
  - (언론)전문대학 개인정보 보호 「F학점」 (전자신문, '12.2.8)
- 교육기관의 주요 노출 사례 및 방지대책을 각급 교육기관에 알려 유사 사례 재발을 방지하고자 함

## □ 홈페이지 노출 현황

- 점검기간 : 2011. 2. 1 ~ 12. 31
- 점검대상 : 공공기관 (34천여개 사이트)
- 노출현황 : 716개 사이트에서 54,930건 노출
  - 전체 노출사이트 중 교육기관 사이트가 69.7% 차지

구분	계	중앙 행정	지자체	교육기관		공사 공단	기타
				교육행정	초중고 및 대학		
사이트수	716	42	126	27	472	43	6
(사이트비율)	100%	5.9%	17.6%	3.8%	65.9%	6%	0.8%

## □ 주요 노출유형

### ○ 각종 명단(부)

- 합격자 명단, 신입생 명단, 학부모 교육 참가자 명단 등 각종 명단을 홈페이지에 게시하면서 주민번호 등 개인정보 노출
- 특히, 인터넷 통신비 지원 대상자 명단, 장기 결식자 명단, 저소득층 학생 명단 등을 홈페이지에 게시하는 사례가 있어 민감정보의 노출 우려가 큼

### ○ 학급경영록

- 학급 운영과 관련, 학급별로 매 학기마다 제작되는 학급경영록에 학생의 주민번호, 주소와 학부모의 휴대폰 번호 등을 기재하여 홈페이지에 게시

### ○ 체험학습 및 동아리 등 참가자 명단

- 수학여행, 스키캠프, 영어캠프 등의 참석자 명단 및 동아리 회원 명단 등을 게시하면서 참가자 및 회원의 주민번호 등 포함

### ○ 대체인력 계획

- 급식 및 기간제 교사 운영계획서 등에 대체인력의 인적사항과 함께 주민번호를 포함하여 게시

- ✓ 업무담당자 부주의에 의해 학생, 학부모 인적사항 등 불필요한 개인정보를 포함하여 홈페이지에 게시
- ✓ 엑셀 자료 첨부 시 각 시트에 개인정보가 포함되어 있는지 여부, 특히 숨김 처리된 개인정보 포함 여부를 확인하지 않고 홈페이지에 게시
- ✓ 홈페이지 노출차단 프로그램 도입 기관임에도, 웹사이트 및 게시판 적용 누락, 개편 후 프로그램 미실행 등의 관리 미흡으로 개인정보 노출 사례 발생

## □ 노출 방지 방안

### <홈페이지 관리자 조치사항>

- 홈페이지 게시자료에 대한 관리 철저
  - 개인정보가 노출되지 않도록 게시판 글 및 첨부파일 지속 확인
  - 홈페이지 개편 등으로 발생한 휴면페이지 삭제 또는 개인정보 포함여부 일제 점검
- 게시판 글쓰기 화면에 개인정보 노출에 따른 법적 처벌 가능성 등을 담은 경고 문구 반드시 게재

#### ☞ 게시판 경고문구 예시

게시판에 글쓰기를 하는 경우, 본문 또는 첨부파일내에 개인정보(주민 등록번호, 성명, 연락처 등)가 포함되어 게시되지 않도록 유의하시기 바랍니다.

개인정보를 포함하여 게시하는 경우에는 불특정 다수에게 개인정보가 노출되어 악용될 수 있으며, 특히 타인의 개인정보가 노출되는 경우에는 개인정보보호법에 따라 처벌을 받을 수 있음을 알려드립니다.

- 구글 등 포털의 검색로봇 수집에 따른 개인정보 노출이 가능하므로 게시판 글쓰기 부분을 검색 로봇이 수집하지 못하도록 검색 차단 기능(Robot.txt) 설정 (붙임2 활용)

#### ☞ 로봇배제표준 적용 시 유의할 점

로봇배제표준 적용을 홈페이지 전체 내용에 대해 검색엔진에 노출되지 않도록 설정할 경우, 일반 검색 포털사이트에서 기관의 홈페이지가 검색되지 않을 수 있으므로, 외부에 비공개되어야 할 정보를 포함하고 있는 폴더(또는 페이지) 혹은 인증절차를 거쳐 접근되는 폴더(또는 페이지) 등에 제한적으로 로봇배제 표준을 설정하여 사용

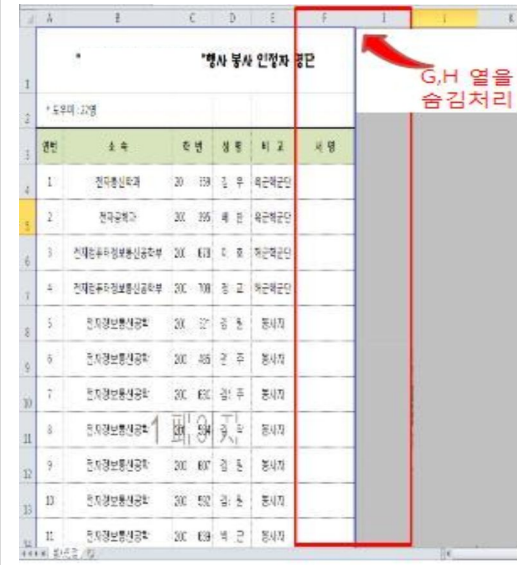
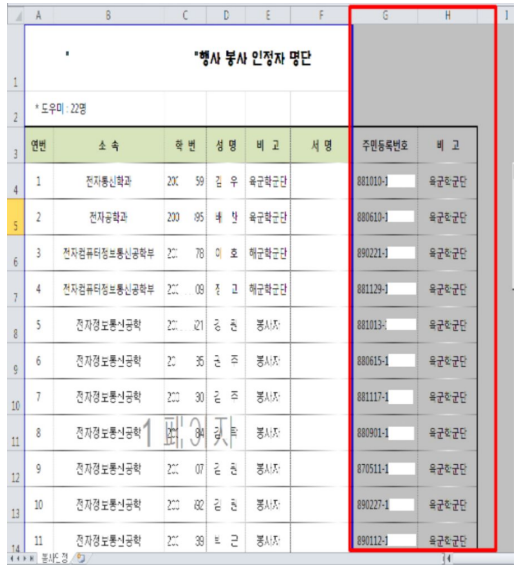
- 게시판을 운영하는 홈페이지는 개인정보 노출차단 시스템 도입
  - 도입이 어려운 경우, 홈페이지 게시판 글 등록 시 주민번호 등 개인정보가 포함되어 있는지 점검하는 방안 적용 (붙임3 활용)
- 이외, "공공기관 홈페이지 개인정보 노출방지 가이드라인([www.privacy.go.kr](http://www.privacy.go.kr)의 자료실의 지침자료에서 확인)을 참조하여 노출방지 조치

## <업무담당자 조치사항>

- 업무상 홈페이지에 글을 게재하는 경우에는 개인정보가 포함되어 있는지 철저히 검사하고 부서장의 결재(확인)를 거쳐 게시
  - 주요 노출유형을 참고하여 명부·명단 등 개인정보가 대량 포함된 파일(엑셀, 한글 등)을 게시하지 않도록 주의
  - 엑셀파일을 게시하는 경우, 첨부파일을 열었을 때에는 보이지 않지만 숨김 처리된 개인정보가 포함되어 있는지 확인

### < 중점 확인사항 >

- 엑셀파일 내의 다른 Sheet에 개인정보가 포함되어 있는지 확인
- 엑셀파일의 숨김 처리된 Sheet(또는 셀)에 개인정보가 포함되어 있는지 확인

개인정보 숨김처리를 한 경우											개인정보 숨김을 취소하면 개인정보 노출										
																					

- 개인정보 위에 이미지를 덮거나 글자색을 변경하여 안보이게 처리한 경우, 글자 색 등을 변경하면 개인정보 노출
- 엑셀의 치환함수를 적용하여 개인정보를 마스킹(\*) 처리한 경우, 해당 개인정보가 어딘가에 숨김처리 되어 있고, \* 처리된 정보도 검색으로 개인정보 노출

## ○ 문의처

- 행정안전부 개인정보보호과 : 김지희 주무관, (02)2100-3345
- 한국인터넷진흥원 개인정보침해대응팀 : 여돈구 주임, (02)405-5632

## 참고 1 개인정보 노출 사례화면

### 1) 학급경영록

라. 학생 조사 일람표

( 2 )학년 ( 1 )반 담임( 교 )

세대수 : ( 16 ) - 동거위학자가 있으면 최고학년만 세대수로 계산

번호	이름	성별	주민등록번호	부	모	주 소	직업	전 화	동거위학자
1.	배근	남	0301	배	주	사천 101-1	회사원	010-	·
2.	김도	·	0303	김	순	사천	자영업	017-	·
3.	고한	·	0304	고	숙	선구동	회사원	010-	선4-1
4.	오금	·	0305	오	경	서동동	회사원	010-	·
5.	이철	·	0306	이	자	선구동 C동4	회사원	010-	·
6.	이대	·	0306	이	민	사천	어업	011-	·
7.	서철	·	0307	서	주	선구동	회사원	010-	·

### 2) 교육과정

VI. 부록

1. 학생명부

순	성명	학급주민등록번호	보호자		전화번호		실거주주소	직업명	비고
			부	모	자택	휴대폰			
1	김민	99000	김	김	041-	010-	김무오스	불도아원	3-무구
2	김민	99005	김	김	070-	010-	김정민 6	회사원	·
3	김민	99003	김	김	063-	010-	김무오스	수업장사	4-고문
4	김민	99005	김	김	049-	010-	김정민 6	공무원	·
5	김민	99001	김	김	043-	010-	김무오스	자영업	·
6	김민	99003	김	김	049-	010-	김무오스	자영업	·
7	김민	99008	김	김	072-	010-	김무오스	·	3-철호
8	김민	99009	김	김	049-	010-	주영민정	·	·
9	김민	99003	김	김	010-	010-	김무오스	회사원	·
10	김민	0001	김	김	070-	010-	김무오스	자영업	·
11	김민	99002	김	김	042-	010-	김무오스	조경사	·
12	김민	0001	김	김	049-	010-	황진 154	자영업	·
13	김민	0001	김	김	032-	010-	김무오스	회사원	3-수진
14	김민	9911	김	김	049-	010-	주영민정	회사원	·
15	김민	9910	김	김	070-	010-	김무오스	경찰	·
16	김민	9910	김	김	049-	010-	김정민 36	음향비	4-정재
17	김민	99006	김	김	000-	010-	김무오스	자영업	4-영민
18	김민	9911	김	김	049-	010-	김정민 36	수업장	·
19	김민	99004	김	김	047-	010-	김무오스	회사원	4-정민
20	김민	99007	김	김	049-	010-	김무오스	김정민정	4-민수
21	김민	99006	김	김	049-	010-	김무오스	김정민정	4-정민
22	김민	99006	김	김	049-	010-	황진 154	김정민정	4-정민
23	김민	99006	김	김	070-	010-	현대부	학생	·

### 3) 대체인력 계획서

문제에 대응하고, 조리원의 특별휴가,보건휴가,산재,연차휴가,병가등 휴가 이용시 복지면에서 부담을 느끼지 않고 휴가를 보내 학교업무에 충실할 수 있도록 함.

**2. 운영방법**  
조리종사원 부재시 보건증이 준비된 대체인력을 확보하여 운영

**3.운영기간**  
2011.03.02 - 2012.02.28

**4. 소요경비: 1,076,940원**  
-조리사 46,150원\*10일(대체인력사용일수)=461,500원  
-조리보조원 43,960원\*14일(대체인력사용일수)=615,440원

**5. 대체인력 명단**

성 명	주민번호	계좌번호	건강진단결과	연락처	비 고
김정	690801	835046- (농협)	양호 (2011.03.07)	9 010- (010- )	

### 4) 저소득층 지원

연월일				성명	법적보호 사실명 현황(해당란에 "1"로 표기)				주요지원내역	비고
연	월	일	성명	국민기초 수급자	한부모	자살위험 계층	특수교육 대상자	주요지원내역	비고	
1	1	1	오영		1			000026	사기	
2	1	2	김영	1				000026	김이	
3	1	2	김영	1				001026	김이	
4	2	1	조영				1	001113	조이	
5	2	1	최영			1		020716	최이	
6	2	1	이영		1			020905	이이	
7	3	1	김영		1			011216	김이	
8	3	1	박영		1			010209	박이	
9	3	2	송영		1			010408	송이	
10	3	2	이영	1				010420	이이	
11	3	2	전영			1		010616	전이	
12	3	2	김영	1				020102	김이	
13	3	2	김영					020102	김이	
14	4	2	김영	1				000522	김이	
15	4	2	김영					001102	김이	
16	4	3	송영		1			000011	송이	
17	4	3	장영	1				001204	장이	
18	5	1	김영				1	991004	김이	
19	5	1	박영		1			990505	박이	
20	5	2	김영				1	981206	김이	

## 5) 합격자 명단

2010학년도 신입생 전형 합격자 명단								
순	수험 번호	출신 학교	성명	생년월일	성별	전형	합격 여부	비고
1	1	군포	김대	941103-2	여	각학	합격	
2	2	관동	김대	941005-2	여	각학	합격	
3	3	관동	박대	940418-2	여	각학	합격	
4	4	관동	박대	941023-2	여	각학	합격	
5	5	관동	최대	940713-2	여	각학	합격	
6	6	관동	김대	940104-1	남	각학	합격	
7	7	관동	김대	940707-1	남	각학	합격	
8	8	관동	김대	930916-1	남	각학	합격	
9	9	배움	오대	941117-1	남	각학	합격	
10	10	배움	최대	940616-1	남	각학	합격	
11	11	배움	김대	950113-2	여	각학	합격	
12	12	배움	박대	950122-2	여	각학	합격	
13	13	배움	문대	940312-2	여	각학	합격	
14	14	배움	이대	950126-2	여	각학	합격	
15	15	배움	김대	940916-2	여	각학	합격	
16	16							

## 6) 동아리 명단

‘온샘골 사랑나눔’ 동아리 명단						
순	학년 (계 위)	성명	주민등록번호	역할분담	연락처 (휴대폰)	
1	교사	정	641025-	지도교사	019-	
2	3	박	950511-	회장	010-	
3	3	권	951018-	부회장	010-	
4	3	박	960304-	3학년 단원	010-	
5	3	이	950313-	3학년 단원	010-	
6	3	김	951115-	3학년 단원	010-	
7	2	김	960921-	2학년 단원	010-	
8	2	김	961130-	2학년 단원	010-	
9	2	박	960712-	2학년 단원	010-	



## 7) 캠프 참가자 명단

스키캠프 방배정 명단						
일시: 2010년 2월 5일-2월6일						
순	방번호	학년	이름	주민등록번호	연락처	비고
1	08	유치원	김지민	040713		봉곡유치원생
2			김지민	031101		
3			김지민	041115		
4			김지민	040512		삼서유치원생
5			김지민	040108		
6			김지민	050511		
7			김지민	040824		인송교사
8			김지민		010-	
9			김지민		017-	
10	09	1	김지민	021211		인송교사(해동축)
11			이민	020921		
12			김지민	020301		
13	3 방정	1	김지민	020711		
14			김지민	001206		
15			김지민	020311		
16	남	1	북	020421		
17			김지민	020911		
18			김지민	000302	010-	
19	방정 1	1	이민	020821		
20			이민	020821		
21			김지민	020301		
22	남	1	해	021111		인송교사
23			해	020301		
24			해	020301		
25	방정 3	해	해	000117		
26			해	010711		
27			김지민	011110		
28	남	2	김지민	010721		
29			김지민	010220	010-	
30			김지민	010301		

## 8) 기타 명단

중학교 의무교육 대상자 명부						
학년 반 호	학생 성명	주민등록번호	거 주 지	보 호 자		비 고
				성명	주민등록번호	
0101	김지민	940121-	수동 주공이	김지민	641022-	남
0102	김지민	940317-	수동 364-9	김지민	600929-	
0103	최지민	940405-	수동 현대이	최지민	690510-	
0104	윤지민	940412-	수동 대수공	윤지민	710917-	
0105	이민	940523-	수동 현대이	이민	680306-	
0107	김지민	940608-	수동 주공이	김지민	661225-	
0108	김지민	940628-	수동 현대이	김지민	620603-	
0109	김지민	940702-	수동 주공이	김지민	671105-	
0110	송지민	940825-	수동 397-3	송지민	630130-	
0111	노지민	940915-	수동 주공이	노지민	671104-	



## 참고 2    로봇배제 표준 적용 방법

### 가) 로봇배제표준 이란?

구글 등 검색 로봇이 정보수집을 위해 사이트를 방문할 경우를 대비하여, 검색로봇에 대한 웹사이트의 디렉토리 및 파일들에 대한 검색조건을 명시해 놓은 국제 규약입니다.

로봇의 배제와 관련된 내용을 적어 놓은 **robots.txt** 파일을 사이트의 최상위 주소(서버 루트 디렉토리 예 : [www.kissp.co.kr/robots.txt](http://www.kissp.co.kr/robots.txt))에 저장하면 됩니다. robots.txt를 서버 디렉토리에 놓는 경우 효력이 없다는 사실에 주의해야 합니다.

- 정상적인 검색로봇은 사이트 방문시 가장 먼저 웹사이트의 로봇배제 표준파일(robots.txt)를 확인하고, 파일에 명시된 사항을 준수하면서 웹사이트의 콘텐츠를 수집합니다.
- 단, 로봇 배제 표준은 방지기술이 아닌 사이트 및 검색로봇 운영자간의 단순한 약속(규약)이므로 악의적인 로봇이 로봇 배제 표준을 무시하고 사이트의 콘텐츠를 수집할 수도 있습니다.

### 나) robots.txt 파일 생성 방법

robots.txt 파일의 내용은 크게 2가지 부분으로 나뉘어집니다. 로봇의 이름을 적는 부분(User-agent), 방문을 허용하지 않을 디렉터리를 적는 부분(Disallow)으로 구분됩니다.

#### ○ robots.txt 문법

User-agent:	*은 User-agent이고, 특정 User-agent를 지정할 경우 로봇 이름을 지정
Allow or Disallow:	Disallow에서 /는 모든 하위 문서에 대해 배제를 의미하고 아무것도 지정하지 않으면 허가를 의미

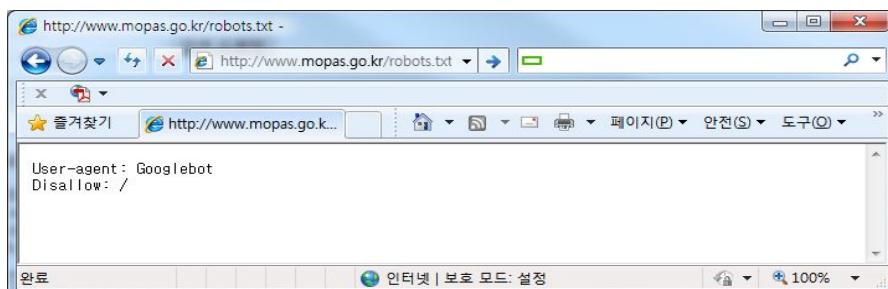
## ○ 로봇배제표준 적용 범위

User-agent: * Disallow: /	홈페이지 전체 내용을 모든 검색엔진에 노출 방지
User-agent: * Disallow:	홈페이지 전체 내용을 모든 검색엔진에 노출 허용
User-agent: * Disallow: /cgi-bin/ Disallow: /temp/ Disallow: /private/	홈페이지의 디렉토리가 cgi-bin, tmp, private(실행파일 모음)등 인 것에 대한 노출 방지
User-agent: kkkkkk Disallow: / User-agent:* Disallow:	특정 검색엔진의 접근만을 거부 (ex: kkkkkk)
User-agent: kkkkkk Disallow: User-agent:* Disallow: /	특정 검색엔진의 접근만을 허용 (ex: kkkkkk)
User-Agent:* Disallow:/directory/a.html	모든 검색엔진에 대해 모든 홈페이지 접근을 허용하지만 /directory/a.html 의 접근을 거부
User-Agent:* Disallow: / Allow:/information	모든 검색엔진에 대해 모든 홈페이지 접근을 거부하지만 디렉토리명이 information인 것에 대한 접근을 허가
User-agent: * Disallow: /*.php	파일 확장자로 차단 (ex: 사이트의 모든 PHP프로그램은 검색로봇이 접근을 거부)

## 다) 로봇배제표준 적용 유무 확인 방법

### ○ 홈페이지에 적용된 robots.txt 파일과 URL 확인

홈페이지에 robots.txt를 적용한 이후에 적용된 robots.txt 파일이 잘 적용되었는지 여부를 파악하기 위해 직접 URL을 통해 점검을 해봅니다.



## ○ Robots.txt Checker를 통한 적용 유무 확인

홈페이지에 적용된 robots.txt 파일의 URL 확인을 한 후 robots.txt 파일 적용이 문법적으로 틀리지 않고 잘 적용되었는지 여부를 확인합니다.

robots.txt 파일이 잘 적용되었는지 여부 확인은 Robots.txt Checker 사이트에서 확인 할 수 있습니다.

- 각 포털사이트에서 Robots.txt Checker 사이트를 찾습니다.
- Robots.txt Checker 실행 후 화면에서 입력란에 웹서버에 적용된 robots.txt 파일의 전체 URL을 입력합니다.

**robots.txt checker**

This tool validates **/robots.txt** files according to the robots exclusion de-facto standard.

Checks are done considering the original 1994 document [A Standard for Robot Exclusion](#), the 1997 Internet Draft specification [A Method for Web Robots Control](#) and [nonstandard extensions](#) that have emerged over the years.

The robots.txt validator checks the syntax and structure of the document, is able to spot typos and can be used to test if a specific crawler is permitted access to a given URL.

**Check URI :**  
   
Can be any document on your website; will automatically download http://yoursite/robots.txt

**User Agents (optional):**  
  
Tests if the user agent may access the URL above.  
Comma separated list; in example Googlebot, Yahoo Slurp, Myrobot

- robots.txt 파일의 적용 여부 확인

**Testresults for http://mopas.go.kr/robots.txt**

robots.txt: <http://mopas.go.kr/robots.txt>  
302 Found 1.92s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/ZfeVZ/robots.txt>  
302 Found 3.83s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/robots.txt>  
302 Found 1.92s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/ZfeVZ/robots.txt>  
302 Found 3.83s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/robots.txt>  
302 Found 1.92s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/ZfeVZ/robots.txt>  
302 Found 3.83s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/robots.txt>  
302 Found 1.92s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/ZfeVZ/robots.txt>  
302 Found 3.83s - 0 Bytes - 🚫

Redirects To: <http://mopas.go.kr/robots.txt>  
302 Found 1.92s - 0 Bytes - 🚫

1 Error 1 Warning 0 Notices

**Relative URI in HTTP Location Header!**

HTTP/1.1 requires the HTTP Location header field to state the redirect target as an absolute URI. While some clients accept relative URIs, most will not.

**Recurring redirects detected; possible redirect loop.**

**Access Test to /robots.txt**

User-agent	Access?	Reason
*	yes	no Disallow directive found <i>This is the default access policy for any robot without a matching User-agent field.</i>

## 라) 로봇배제표준 대안

### ○ HTML의 meta 태그 이용

html 문서의 <head>와</head> 사이에 입력합니다.

```
<meta name="Robots" content="Noindex, Nofollow" />
```

※ 이러한 방법은 일반적인 방법이 아니고, 아직까지 일부의 로봇만이 지원

### 【 검색엔진 LIST 】

검색엔진	검색로봇
다음	Daumoa
구글	Googlebot
야후	Yahoo! Slurp
네이버	Naverbot
Bing	Bingbot
Microsoft	Msnbot

#### ☞ 로봇배제표준 적용 시 유의할 점

로봇배제표준 적용을 홈페이지 전체 내용에 대해 검색엔진에 노출되지 않도록 설정할 경우, 일반 검색 포털사이트에서 기관의 홈페이지가 검색되지 않을 수 있으므로, 외부에 비공개되어야 할 정보를 포함하고 있는 폴더(또는 페이지) 혹은 인증절차를 거쳐 접근되는 폴더(또는 페이지) 등에 제한적으로 로봇배제표준을 설정하여 사용

<b>참고 3</b>	<b>게시판 글 등록 시 주민번호 등 개인정보 필터링 적용방법</b>
-------------	--

홈페이지 게시판에 글을 등록할 경우 개인정보가 포함되어 있는지를 점검할 수 있도록 구축할 수 있습니다. 이때 정규 표현식을 통해 개인정보로 간주되는 항목을 찾을 수 있습니다. 정규 표현식은 각각의 버전마다 지원하는 함수 및 문법이 상이할 수 있으므로, 사용하는 프로그램 환경에 맞는 정규 표현식을 사용하시기 바랍니다.

### 가) 정규 표현식(正規 表現式, Regular Expression)

#### ○ ‘정규 표현식’이란?

특정한 규칙을 가진 문자열의 집합을 표현하는 데 사용하는 형식 언어입니다. 정규 표현식은 많은 텍스트 편집기와 프로그래밍 언어에서 문자열의 검색과 치환을 위해 지원하고 있으며, 특히 Perl과 Tcl은 언어 자체에 강력한 정규 표현식 구현을 내장하고 있습니다.

정규 표현식은 컴퓨터 과학의 정규 언어로부터 유래하였으나 구현체에 따라서 정규 언어보다 더 넓은 언어를 표현할 수 있는 경우도 있으며, 심지어 정규 표현식 자체의 문법도 여러 가지가 있습니다. 이 중 표준화된 것으로는 POSIX의 확장 정규 표현식이 있으며, 표준화되지는 않았지만 Perl의 정규 표현식과 그 대체 구현인 PCRE도 널리 사용됩니다.

정규 표현식은 주로 텍스트 탐색과 문자열 조작에 쓰입니다. 정규 표현식은 하나의 문자와 일치(match)하거나, 혹은 문자열의 일부분(substring)이나 전체 문자열인 문자 집합들과 일치하게 됩니다.

## ○ 문법 (기본적인 정규 표현식 문법)

몇몇 문자는 특정한 논리적 관계를 나타내는 기호로 쓰입니다. 이를 제외한 나머지 문자는 일반적인 문자를 나타냅니다.

- 선택 기호 : "|" 기호는 여러 식 중에서 하나를 선택합니다.

예를 들어, "abc|adc"는 abc라는 문자열과 adc라는 문자열을 모두 포함합니다.

- 묶기 기호 : "("와 ")"로 여러 식을 하나로 묶을 수 있습니다. "abc|adc"와 "a(b|d)c"는 같은 의미를 가집니다.

- 개수 기호 : 문자 뒤에 붙어 문자의 개수를 나타냅니다.

- "\*" : 0개 이상. "a\*b"는 "b", "ab", "aab", "aaab"를 포함
- "+" : 1개 이상. "a+b"는 "ab", "aab", "aaab"를 포함하지만 "b"는 포함하지 않음
- "?" : 0개 또는 1개. "a?b"는 "b", "ab"를 포함
- "{m, n}" : m개 이상 n개 이하. "a{1,3}b"는 "ab", "aab", "aaab"를 포함하지만, "b"나 "aaaab"는 포함하지 않음

ex) "(fa|mo|b?o)ther"는 "father", "mother", "bother", "other"를 나타냅니다.

- 많은 프로그래밍 언어에서는 이를 확장한 문법을 가지고 있습니다. 이 중 일반적으로 사용되는 연산자는 다음과 같습니다.

- "[]" : "["과 "]" 사이의 문자 중 하나를 선택합니다. "|"를 여러 개 쓴 것과 같은 의미를 가짐

예를 들면, [abc]d는 ad, bd, cd를 뜻함. 또한, "-" 기호와 함께 쓰면 문자의 범위를 지정할 수 있습니다. "[a-z]"는 a부터 z까지 중 하나, "[1-9]"는 1부터 9까지 중의 하나를 뜻함

- "[^]" : "[^"과 "]" 사이의 문자를 제외한 나머지 하나를 선택

예를 들면, [^abc]d는 ad, bd, cd는 포함하지 않고 ed, fd 등

포함. `[^a-z]`는 알파벳 소문자로 시작하지 않는 모든 문자를 나타냄

- `"^"`, `"$"` : 각각 문자열의 처음과 끝을 나타냄.

## 나) 개인정보 패턴 예시 (linux/unix C - regex 버전)

- 주민등록번호  
`(([01][0-9]{5})[[:space:]],~-[1-4][0-9]{6})[2-9][0-9]{5}[[:space:]],~-[1-2][0-9]{6})`
- 여권번호  
`[a-zA-Z]{2}[~.[:space:]][0-9]{7}`
- 운전면허번호  
`[0-9]{2}[~.[:space:]][0-9]{6}[~.[:space:]][0-9]{2}`
- 핸드폰번호  
`01[016789][~.[:space:]][0-9]{3,4}[~.[:space:]][0-9]{4}`
- 신용카드번호  
`[34569][0-9]{3}[~.[:space:]][0-9]{4}[~.[:space:]][0-9]{4}[~.[:space:]][0-9]{4}`
- 건강보험번호  
`[1257][~.[:space:]][0-9]{10}`
- 계좌번호  
`(([0-9]{2}[~.[:space:]][0-9]{2}[~.[:space:]][0-9]{6})[0-9]{3}[~.[:space:]]([0-9]{5,6}[~.[:space:]][0-9]{3})[0-9]{6}[~.[:space:]][0-9]{5}[0-9]{2,3}[~.[:space:]][0-9]{6})[0-9]{2}[~.[:space:]][0-9]{7}[0-9]{2}[~.[:space:]][0-9]{4,6}[~.[:space:]][0-9]{5}[~.[:space:]][0-9]{3}[~.[:space:]][0-9]{2}[0-9]{2}[~.[:space:]][0-9]{5}[~.[:space:]][0-9]{3}[0-9]{4}[~.[:space:]][0-9]{4}[~.[:space:]][0-9]{3}[0-9]{6}[~.[:space:]][0-9]{2}[~.[:space:]][0-9]{3}[0-9]{2}[~.[:space:]][0-9]{2}[~.[:space:]][0-9]{7}[0-9]{4}[~.[:space:]]([0-9]{3}[~.[:space:]][0-9]{6})[0-9]{2}[~.[:space:]][0-9]{6}[~.[:space:]][0-9]{5}[~.[:space:]][0-9]{2}[~.[:space:]][0-9]{6}[0-9]{6}[~.[:space:]][0-9]{2}[~.[:space:]][0-9]{5,6})`

## ○ 개인정보 패턴 적용 예시 (Linux/Unix C - regex 버전)

```
int main(int argc, char *argv[])
{
    //주민등록번호 정규표현식
    char *ssn_regex="([01][0-9]{5})[[:space:]],~-[1-4][0-9]{6})[2-9][0-9]{5}[[:space:]],~-[1-2][0-9]{6})";
    //테스트용 문자열
    char *sample="테스트 중인 주민등록번호는 810514-1234567 입니다.";
    char *p=NULL;
    regex_t *re;
    regmatch_t match;
    int ret;

    ret = regcomp(re, ssn_regex, REG_EXTENDED);
    if(ret) return 0;

    while(sample && *sample) {
        //정규표현식으로 주민번호 패턴에 맞는 문자열이 존재하는지 검사
        ret = regexec(re, sample, 1, &match, 0);
        //정규표현식에 매치되는 문자열이 없으면 중단
        if(ret) break;

        //검색한 문자열을 출력
        p = (char *)calloc(match.rm_eo-match.rm_so+1, sizeof(char));
        memcpy(p, sample+match.rm_so, match.rm_eo - match.rm_so);
        p[match.rm_eo - match.rm_so] = 0x00;

        printf("주민등록번호: %s\n", p);
    }
}
```