

# 구글에서 웹 사이트 검색 차단하기

작 성 자 : 강 태 우

작 성 일 : 09. 01. 29

소 속 : SPIDER

지금부터 구글에서 개인 정보 유출에 따른 사이트 검색 차단 방법에 대해서 소개 해드리겠습니다. 이 방법은 자신의 계정에 관리하는 홈페이지들을 등록시킨 후 검색을 허용할 사이트와 차단해야할 사이트를 직접 작성하는 방법으로 웹상에 자신의 홈페이지가 있고, 현재 구글에서 홈페이지가 검색된다는 전제하에 설명해드리겠습니다.

1. 우선 [www.google.co.kr](http://www.google.co.kr) 홈페이지에 접속하셔서 자신의 계정으로 로그인 합니다. 그러면 오른쪽 상단에 내 계정이라는 부분이 생기는데 이를 클릭해줍니다.



2. 계정에 들어오셨으면 오른쪽 중간에 웹마스터 도구가 있습니다. 이 곳은 웹 계정에 대한 여러 가지 설정을 할 수 있는 부분입니다. 클릭해줍니다.



3. 웹마스터 도구에 가시면 대시보드가 나오게 됩니다. 여기에다가 자신이 관리하는 사이트를 추가해줍니다.

## Google 웹마스터 도구



4. 개요 부분입니다. 이 단계에서 추가 정보를 눌러 기본적인 설정을 해보도록 하겠습니다.

### 개요

law.chonbuk.ac.kr

**다음 단계**  
사이트 확인: 사이트를 확인하면 페이지에 대한 종합적인 통계와 크롤링 오류 정보에 액세스할 수 있습니다.

색인생성	
홈페이지 크롤링:	✓ Googlebot이 귀하의 홈페이지에 액세스했습니다. <b>추가 정보</b>
색인 상태:	✓ 사이트의 페이지가 Google 색인에 포함되어 있습니다. 색인 통계를 참조하세요. ?

도움말 센터에서 다음을 포함한 자세한 답변 찾기:

- 내 사이트의 순위를 높이려면 어떻게 해야 하나요?
- 특정 키워드에 대해 내 사이트가 나타나지 않는 이유는 무엇인가요?
- 검색결과에서 내 페이지의 순위가 낮아졌습니다. 왜 그런가요?

5. 추가정보를 누르면 다음과 같이 확인이 되어있지 않다고 나올 것입니다. 그 이유는 아직 홈페이지에 메타태그를 추가해주지 않았기 때문인데요. 자신의 홈페이지에 가서 메타 태그를 추가해줍니다.

### 사이트 확인 ?

law.chonbuk.ac.kr

**확인 상태: 확인되지 않음**  
마지막으로 2009-1-27에 시도됨: 확인 메타 태그를 찾을 수 없습니다.

사이트 소유자임을 확인한 다음 사이트의 페이지에 대한 종합적인 통계 및 오류 정보를 제공해 드리겠습니다. 본인이 사이트 소유자임을 확인할 수 없는 경우에도 계속해서 웹마스터 도구를 사용하여 Sitemap을 제출할 수 있으며 사용자의 사이트에 대한 기본 정보 및 제출된 Sitemap에 대한 자세한 정보를 볼 수 있습니다.

인증하는 방법은 두 가지가 있습니다. 하나는 Google이 지정한 이름으로 HTML 파일을 업로드하는 방법이고, 다른 하나는 사용자 사이트의 색인 파일에 메타태그를 추가하는 방법입니다. 아래에서 원하는 방법을 선택하세요. ?

메타태그 추가

아래에 있는 메타태그를 복사하여 사이트 홈페이지의 첫 번째 <body> 섹션 앞의 첫 번째 <head> 섹션에 붙여넣으세요. ?

```
<meta name="verify-v1" content="tXU0I7+5yZrJbfTPfsOAYHw+LJA54p18RR0adThjfTY=" />
```

▶ 예제 보기

http://law.chonbuk.ac.kr/의 홈페이지에 메타태그를 추가했습니다.

확인    나중에 확인

6. 다음과 같이 <head>부분에 메타태그를 추가시켜줍니다.

```

<html>
<head>
<title>:::전북대학교 법과대학::: </title>
<meta http-equiv="Content-type" content="text/html; charset=euc-kr">
<meta name="verify-v1" content="tXU0I7+5yZrJbfTPfsOAYHw+LJA54p18RR0adThjfTY=" />
<link href="new_style.css" rel="stylesheet" type="text/css">

<style type="text/css">

```


7. 다시 사이트 확인 창으로 넘어와서 제일 밑에 있는 확인 버튼을 눌러줍니다.


### 사이트 확인

law.chonbuk.ac.kr

**확인 상태: 확인되지 않음**  
마지막으로 2009-1-27에 시도됨. 확인 메타 태그를 찾을 수 없습니다.

사이트 소유자임을 확인한 다음 사이트의 페이지에 대한 종합적인 통계 및 오류 정보를 제공해 드리겠습니다. 본인이 사이트 소유자임을 확인할 수 없는 경우에도 계속해서 웹마스터 도구를 사용하여 Sitemap을 제출할 수 있으며 사용자의 사이트에 대한 기본 정보 및 제출된 Sitemap에 대한 자세한 정보를 볼 수 있습니다.

인증을 하는 방법은 두 가지가 있습니다. 하나는 Google이 지정한 이름으로 HTML 파일을 업로드하는 방법이고, 다른 하나는 사용자 사이트의 색인 파일에 메타태그를 추가하는 방법입니다. 아래에서 원하는 방법을 선택하세요. 

메타태그 추가 

아래에 있는 메타태그를 복사하여 사이트 홈페이지의 첫 번째 <body> 섹션 앞의 첫 번째 <head> 섹션에 붙여넣으세요. 

```
<meta name="verify-v1" content="tXU0I7+5yZrJbFTPfsOAYHw+LJA54p18RR0adThjfTY=" />
```

▶ 예제 보기

http://law.chonbuk.ac.kr/의 홈페이지에 메타태그를 추가했습니다.

**확인**    나중에 확인

8. 확인을 누르면 구글에서 홈페이지에 대한 크롤링 정보를 보여주게 됩니다. 만약에 구글에서 검색한 내용중에 오류가 있다면 아래의 웹크롤링 오류 부분에 에러 화면이 뜰 것입니다.

### 개요

law.chonbuk.ac.kr

✓ http://law.chonbuk.ac.kr/이(가) 성공적으로 확인되었습니다.

색인생성   인기 검색어 >	
홈페이지 크롤링:	✓ Googlebot이 귀하의 홈페이지에 액세스했습니다. 추가 정보
색인 상태:	✓ 사이트의 페이지가 Google 색인에 포함되어 있습니다. 색인 통계를 참조하세요.   Sitemap이 제출되지 않았습니다. Sitemap을 제출하면 Google이 페이지를 찾는 데 도움이 됩니다. 사이트의 URL 목록이 나와 있는 Sitemap을 작성하여 제출하면 Google에서 각 페이지에 대해 색인을 생성하는 방법에 관한 데이터를 제공합니다. <a href="#">추가 정보</a>

웹크롤링 오류		
HTTP 오류	✓ 0	--
Sitemap의 URL 오류	✓ 0	--
URL을 추적할 수 없습니다.	✓ 0	--
URL이 시간초과되었습니다.	✓ 0	--
robots.txt에 의해 제한된 URL	✓ 0	--
연결할 수 없는 URL	✓ 0	--
찾을 수 없습니다.	✓ 0	--
합계:	0	

도움말 센터에서 다음을 포함한 자세한 답변 찾기:

- 내 사이트의 순위를 높이려면 어떻게 해야 하나요?
- 특정 키워드에 대해 내 사이트가 나타나지 않는 이유는 무엇인가요?
- 검색결과에서 내 페이지의 순위가 낮아졌습니다. 왜 그런가요?

9. 이제 [대시보드] > [도구]에서 설정을 하도록 하겠습니다. 이 메뉴는 홈페이지에 robots.txt 파일을 넣는 것과 구글에서 검색되지 않아야 할 URL 제거 등을 나타내는 부분입니다. 우선 웹에서 취약한 페이지 삭제 요청을 하기 위해서 URL 제거를 클릭해줍니다.

개요	<b>도구</b>
설정	law.chonbuk.ac.kr
진단	<a href="#">robots.txt 분석</a>
통계	robots.txt 파일이 특정 URL을 차단하는지 여부와 수정내용을 확인합니다. Google의 다양한 user-agent도 테스트합니다.
링크	
Sitemap	
<b>도구</b>	
개요	
robots.txt 분석	
robots.txt 생성	<a href="#">robots.txt 생성</a>
사이트 확인 관리	<a href="#">사이트 확인 관리</a>
<b>URL 제거</b>	<a href="#">URL 제거</a>
404 페이지 기능 향상	<a href="#">404 페이지 기능 향상</a>
가젯	<a href="#">가젯</a>

**사이트 확인 관리**  
해당 사이트에 대해 확인된 모든 소유자에 대해 알아보고 필요한 경우 다시 확인합니다.

**URL 제거**  
제거 계획을 포함하며, Google 색인에서 콘텐츠를 제거합니다.

**가젯**  
내 홈페이지에서 Google 웹마스터 도구를 사용해 보세요. 오류를 찾고 사이트 통계 및 세부정보를 볼 수 있습니다. iGoogle에 대해서는 더욱 많은 정보를 알아볼 수 있습니다.

10. URL 제거 부분입니다. 3개를 선택가능할 수 있는데요. 우리는 정보가 공개된 페이지만 삭제되도록 요청하기 위해서 제일 위에 있는 개별 URL:웹페이지,이미지 또는 기타 파일을 선택 후 다음을 클릭합니다.

### URL 제거 ?

law.chonbuk.ac.kr

#### Google 검색결과에서 무엇을 제거하시겠습니까?

삭제 상태는 90일 동안 유지됩니다. 언제든지 삭제를 취소할 수도 있습니다. 90일이 지난 후 페이지가 차단되어 있거나 존재하지 않으면 페이지가 Google 검색결과에 표시되지 않습니다.

- 개별 URL: 웹페이지, 이미지 또는 기타 파일** 시한이 지났거나 차단된 웹 페이지, 이미지 및 기타 파일이 Google 검색결과에 나타나지 못하게 합니다.
- 해당 사이트의 디렉토리 및 모든 하위 디렉토리 사이트의 특정 디렉토리에 있는 모든 파일 및 하위 디렉토리가 Google 검색결과에 표시되지 않도록 제거합니다.
- 전체 사이트 귀하의 사이트를 제거하며 Google 검색결과에 나타나지 않도록 합니다.
- Google 검색결과에 캐시된 사본 시한이 지난 페이지 또는 사용자가 noarchive 메타태그를 추가한 페이지의 캐시된 사본 및 설명을 제거합니다.

참고: law.chonbuk.ac.kr 에 없는 URL을 제거하려면 [공개 URL 제거 도구](#)를 사용해 보시기 바랍니다.

11. 제거할 웹페이지 URL을 지정하는 곳입니다. 이곳에 URL을 제거할 콘텐츠를 추가사키는 곳입니다.

- ① 제거할 URL을 입력합니다.
- ② 추가를 누르면 아래 제거할 콘텐츠 부분에 URL이 삽입 됩니다.
- ③ 제거할 URL을 다 작성하였으면 삭제 요청 제출을 클릭합니다.

## URL 제거 ?

[law.chonbuk.ac.kr](http://law.chonbuk.ac.kr)

**제거할 웹페이지 URL을 지정하십시오.**

해당 요청을 수행하려면 제출된 모든 URL은 제거가 가능해야 합니다. [자세히 알아보기](#)

검색결과에 표시되는 URL:

①

그 깨진 링크가 어디에 표시됩니까?

Google 웹 검색결과

Google 이미지 검색결과

②

---

**제거할 콘텐츠 (최대: 100)**

③

12. 삭제 요청을 제출하면 자신이 요청한 웹페이지가 나오게 됩니다. 일반적으로 삭제 요청을 처리하는 데는 3~5일정도 소유 됩니다.

## URL 제거

[law.chonbuk.ac.kr](http://law.chonbuk.ac.kr)

URL을 삭제하기 전에 Google 검색결과에서 삭제하려는 콘텐츠를 Google 및 다른 검색엔진에서 크롤링하지 않는지 확인해야 합니다.

URL을 삭제하려면 각 페이지에서 [404](#) 또는 [410 HTTP 상태 코드](#)가 반환되거나 [robots.txt 파일](#) 또는 [noindex 메타 태그](#)를 사용하여 크롤러가 콘텐츠에 접근하는 것을 차단해야 합니다.

전체 사이트 또는 디렉토리 제거를 요청하려면 robots.txt 파일을 사용하여 크롤러가 이 콘텐츠에 액세스하는 것을 차단해야 합니다.

[Google 색인에서 콘텐츠 제거에 대해 자세히 알아보기](#)

삭제 요청을 처리하는 데에는 업무일 기준 3~5일 정도 소요되나 규모가 큰 URL의 경우 시간이 더 걸릴 수 있습니다.

http://law.chonbuk.ac.kr/sub04\_1.php(가) 제거 목록에 추가되었습니다.

**진행 중인 요청 (2)**    제거된 콘텐츠

요청됨	URL	제거 유형	상태
2009. 1. 28.	<a href="http://law.chonbuk.ac.kr/sub04_1.php">http://law.chonbuk.ac.kr/sub04_1.php</a>	웹페이지 제거	보류 <input type="button" value="취소"/>
2009. 1. 27.	<a href="http://law.chonbuk.ac.kr/">http://law.chonbuk.ac.kr/</a>	사이트 제거	취소됨

[개인정보 보호정책](#)

13. 삭제가 되었더라도 구글봇은 다시 사이트를 크롤링하여 등록 시킬 수 있습니다. 이를 방지하기 위해서는 robots.txt 파일을 삽입해서 특정 페이지가 검색이 되지 않도록 해야됩니다. robots.txt 파일은 웹사이트의 가장 최상위 폴더에 설치해서 로봇이 검색할 시에 조건을 만들어 주게 됩니다.

1번은 모든 로봇의 검색을 허용할지 차단할지 결정하는 곳입니다.

2번은 직접 차단 또는 허용할 디렉토리나 파일을 직접 지정하는 곳입니다.

작업 : 검색 허용, 검색 차단

로봇 : 어떤 로봇에 대해서 검색을 허용할지 차단할지 결정

파일 또는 디렉토리 : 어떤 파일이나 디렉토리에 대해서 위의 작업을 허용할지 결정

위의 세가지를 지정하고 추가를 누르면 3번에 조건에 대한 파일을 생성도록 해줍니다.

3번에는 2번에 대한 규칙을 추가 시켜주는 곳입니다. 조건에 대해 다 추가 해주었으면 다운로드를 눌러서 파일을 자신의 웹사이트의 최상위 디렉토리에 저장시켜줍니다.

### robots.txt 생성

law.chonbuk.ac.kr

robots.txt 파일을 사용하여 사이트 크롤링을 허용하지 않을 로봇과 크롤링 대상에서 제외할 파일 또는 디렉토리를 지정할 수 있습니다. [자세히 알아보기](#)

#### 1: 기본 로봇 액세스 선택

- 모든 로봇 허용(권장) — 다음 단계에서 이 규칙을 세부적으로 조정할 수 있습니다.)
- 모든 로봇 차단

#### 2: 추가 규칙 지정(선택사항)

작업	로봇	파일 또는 디렉토리	
차단	모든 로봇	/sub04_1.php	추가
작업...			
허용			
차단			

이 규칙을 모든 파일 및 디렉토리에 적용하려면 /를 입력하세요. 개별 파일 또는 디렉토리를 지정하려면 해당 경로를 한 줄에 하나씩 입력하세요. 예:  
/directory/file.html  
/directory/subdirectory/

#### 3: robots.txt 파일 저장

robots.txt 파일을 저장하려면 다운로드를 클릭하거나 아래의 텍스트를 .txt 파일에 붙여넣습니다. 파일의 이름을 robots.txt로 지정하고 사이트의 상위 수준 디렉토리에 저장합니다. [자세히 알아보기](#)

```
User-Agent: *
Allow: /
```

다운로드

<규칙 생성 전>

## robots.txt 생성

law.chonbuk.ac.kr

robots.txt 파일을 사용하여 사이트 크롤링을 허용하지 않을 로봇과 크롤링 대상에서 제외할 파일 또는 디렉토리를 지정할 수 있습니다. [자세히 알아보기](#)

### 1: 기본 로봇 액세스 선택

- 모든 로봇 허용(권장 mdash; 다음 단계에서 이 규칙을 세부적으로 조정할 수 있습니다.)
- 모든 로봇 차단

### 2: 추가 규칙 지정(선택사항)

작업	로봇	파일 또는 디렉토리	수정	삭제
차단	모든 로봇	/sub04_1.php		
<input type="button" value="규칙 추가"/>				

### 3: robots.txt 파일 저장

robots.txt 파일을 저장하려면 다운로드를 클릭하거나 아래의 텍스트를 .txt 파일에 붙여넣습니다. 파일의 이름을 robots.txt로 지정하고 사이트의 상위 수준 디렉토리에 저장합니다. [자세히 알아보기](#)

```
User-Agent: *
Disallow: /sub04_1.php
Allow: /
```

<규칙 생성 후>

14. 분석에 가보시면 구글에서 robots.txt 파일에 대한 분석결과를 나타내 주게 됩니다.

## robots.txt 분석

law.chonbuk.ac.kr

robots.txt URL	<a href="http://law.chonbuk.ac.kr/robots.txt">http://law.chonbuk.ac.kr/robots.txt</a>
마지막 다운로드	1시간 전
상태	200 (성공)
홈페이지 접근	Googlebot 차단 사이트: <a href="http://law.chonbuk.ac.kr/">http://law.chonbuk.ac.kr/</a>

### 이 robots.txt 파일을 테스트한 후 변경해 보십시오.

텍스트 <http://law.chonbuk.ac.kr/robots.txt>

캐시된 robots.txt 파일의 기본 텍스트는 다음과 같습니다. 다운로드한 후 해당 파일을 변경했다면 제시된 텍스트가 최신 버전과 다를 수 있습니다. Google은 거의 하루에 한 번씩 새 robots.txt 파일을 확인합니다.

캐시된 robots.txt 파일이 특정 URL에 대한 접근을 차단하는지 여부를 확인하려면 해당 텍스트를 그대로 유지하고, robots.txt 파일을 변경할 경우 특정 URL에 대한 접근 설정이 어떻게 변경되는지 확인하려면 해당 파일을 수정하십시오. robots.txt 구문에 대한 자세한 내용을 보려면 [robotstxt.org](http://robotstxt.org)를 참조하십시오.

이 텍스트를 수정할 때 사이트에 있는 robots.txt 파일을 변경하는 것은 아닙니다. robots.txt 변경사항을 테스트한 후 사이트에 있는 해당 파일을 적절하게 변경하십시오.

```
User-Agent: *
Disallow: /
```

### 이 robots.txt 파일에 대한 URL 테스트

첫 번째를 추가했습니다. 확인하려는 URL을 한 줄에 하나씩 추가하세요.

```
http://law.chonbuk.ac.kr/
```

### user-agent 선택

Google은 몇 가지 사용자 에이전트를 사용합니다. 사용자는 이 robots.txt 파일에서 특정 사용자 에이전트에 대한 액세스를 제한할 수 있습니다. Google은 Google 검색 로봇을 사용하여 정기적인 웹 크롤링을 수행하기 때문에 자동으로 Google 검색 로봇을 검사합니다. 테스트할 사용자 에이전트를 추가로 선택할 수 있습니다. robots.txt 파일을 사용하여 크롤러가 사이트에 액세스하는 것을 제한하는 방법에 대해 [자세히 알아보세요](#).

- Googlebot - 웹 색인 및 Google 뉴스용 페이지 크롤링
- 검사할 추가 user-agent 선택