# 목차

**수집 로봇 관련 설정** ..... 3
***robots.txt*** ..... 3
***이미 수집된 항목의 삭제*** ..... 3

| | | |
|---|---|---|
| Last update: 2020/12/13 06:17 | 수집로봇관련설정 | https://atl.kr/dokuwiki/doku.php/%EC%88%98%EC%A7%91_%EB%A1%9C%EB%B4%87_%EA%B4%80%EB%A0%A8_%EC%84%A4%EC%A0%95 |

# 수집 로봇 관련 설정

## robots.txt

robots.txt 파일을 사용하면 크롤러가 사이트에 요청할 수 있는 페이지/파일과 요청할 수 없는 페이지/파일을 검색엔진 크롤러에 알려 줄 수 있습니다. 이 파일은 주로 요청으로 인해 사이트가 오버로드되는 것을 방지하기 위해 사용하며, 로봇으로부터 웹페이지를 숨기기 위한 메커니즘이 아닙니다. 웹페이지가 표시되지 않도록 하려면 밑에 설명한 이미 수집된 항목의 삭제 방법을 사용하거나 비밀번호로 페이지를 보호해야 합니다.

예제

```
# Example 1: Block only Googlebot
User-agent: Googlebot
Disallow: /

# Example 2: Block Googlebot and Adsbot
User-agent: Googlebot
User-agent: AdsBot-Google
Disallow: /

# Example 3: Block all but AdsBot crawlers
User-agent: *
Disallow: /
```

- https://developers.google.com/search/docs/advanced/robots/intro?hl=ko

## 이미 수집된 항목의 삭제

검색엔진에 이미 수집된 항목들은 robots.txt에 모두 거부를 한다 해도 삭제되지 않습니다. 해당 항목에 대해 삭제를 원할경우 검색엔진쪽에 직접 요청하거나 해당 페이지에 대해 정보삭제를 요청하는 메타태그를 추가하여 다음에 수집해가는 시점에 반영되도록 해야 합니다.

예제

```
<!DOCTYPE html>
<html><head>
<meta name="robots" content="noindex" />
(…)
</head>
<body>(…)</body>
</html>
```

- https://developers.google.com/search/reference/robots_meta_tag?hl=ko

| | | |
|---|---|---|
| Last update: 2020/12/13 06:17 | 수집로봇관련설정 | https://atl.kr/dokuwiki/doku.php/%EC%88%98%EC%A7%91_%EB%A1%9C%EB%B4%87_%EA%B4%80%EB%A0%A8_%EC%84%A4%EC%A0%95 |

From:
https://atl.kr/dokuwiki/ - **AllThatLinux!**

Permanent link:
**https://atl.kr/dokuwiki/doku.php/%EC%88%98%EC%A7%91_%EB%A1%9C%EB%B4%87_%EA%B4%80%EB%A0%A8_%EC%84%A4%EC%A0%95**

Last update: **2020/12/13 06:17**