**robots.txt를 효과적으로 활용하기**

- robots: 웹사이트를 크롤링하기 위해 들어오는 시스템

- robots.txt: robot들의 접근제어, 권한부여 등을 위해 필요한 정보

naver.com/robots.txt

- User-agent: \* > robots, 크롤러

- Disallow: / > naver에서 제작된 contents는 타 검색엔진에서 접근 불가

- Allow : /$

- Sitemap: /sitemap > 해당 사이트의 지도, 구조를 기계가 이해하기 쉬운 형태로 제공

google.com/robots.txt

- User-agent: \*

- Disallow: /search > google.com/search

> 정보 보호, 접근할 정보 존재 X

- Allow: /search/about > google.com/search/about

1) robots.txt를 보안 도구로 사용하지 않기

- 정중한 요청의 방식 🡪 악의적 목적의 검색엔진은 이를 무시

2) 민감한 콘텐츠는 보다 안전한 방법 사용하기

Sitemap

- 확장자: xml; 기계가 더 이해하기 용이

- sitemap.xml file에 웹사이트의 내의 존재하는 page를 <url></url>내에 나열 🡪 검색엔진의 이해 명확

- Sitemap: /sitemap.xml