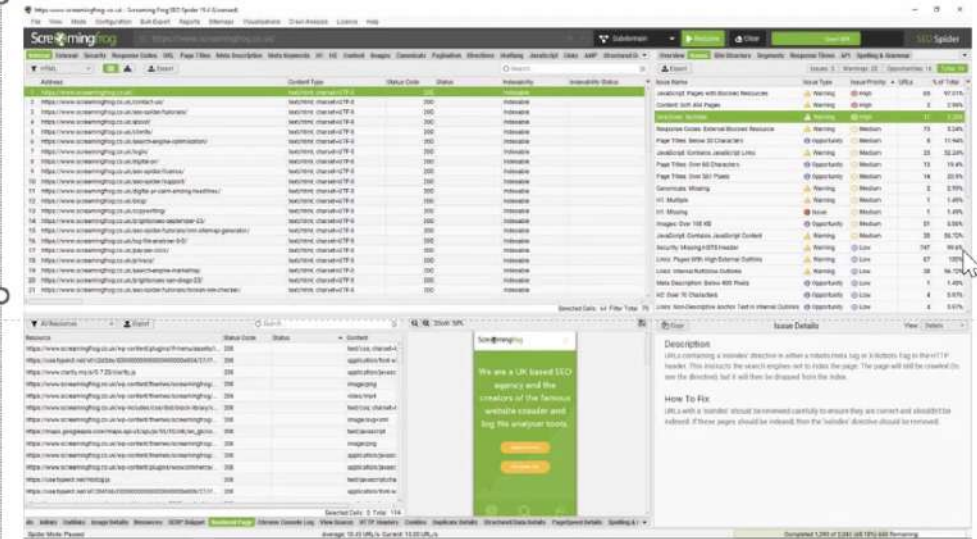


■ 사이트 전체에 대한 요소

	중요도	페이지
• <u>암호화 통신 대응 https 도입 사이트의 암호화 통신(SSL) 대응</u>	중	4
• <u>크롤 최적화 사이트맵 송신 동적 갱신 사이트 맵 작성·송신</u>	고~저	6
• <u>크롤 최적화 사이트맵 송신 동적 갱신 사이트맵과 RSS/ATOM 피드 병용</u>	중~저	7
• <u>크롤 최적화 사이트맵 동적 업데이트 사이트맵의 적절한 동적 생성</u>	중	8
• <u>크롤 최적화 robots.txt robots.txt에 의한 크롤링 제어</u>	중	9
• <u>크롤 최적화 noindex noindex를 사용한 검색 인덱스 등록 블록</u>	고	10
• <u>크롤 최적화 URL 기본 구조 중복 회피 URL 기본 구조 중복 회피</u>	중	11
• <u>크롤 최적화 리다이렉트 적절한 리다이렉트 설정</u>	중	12
• <u>크롤 최적화 크롤 통계 정보 오류 해소 GSC의 크롤 통계 정보 오류 중지</u>	저~중	13
• <u>크롤 최적화 hreflang 태그 설정 페이지의 현지화 버전을 Google에 전달</u>	고	15
• <u>크롤 최적화 무한 스크롤 스크롤 하단의 콘텐츠를 구글에 인식시키기</u>	저	16
	중	17



- ✓ GSC(구글 서치 콘솔)로 http 및 https 여부를 확인한다
- ✓ mixed contents(예시 중 하나로, 이미지의 URL이 http 인 경우) 존재 여부를 확인한다.
- ✓ (추천) SEO 분석 도구인, **Screaming Frog**로 내부 링크 기술 확인한다.
- 일반적인 리다이렉트 시 문제 없으나, 만약 302 리다이렉트인 경우, 이전 URL이 계속 잔류할 가능성이 있다.



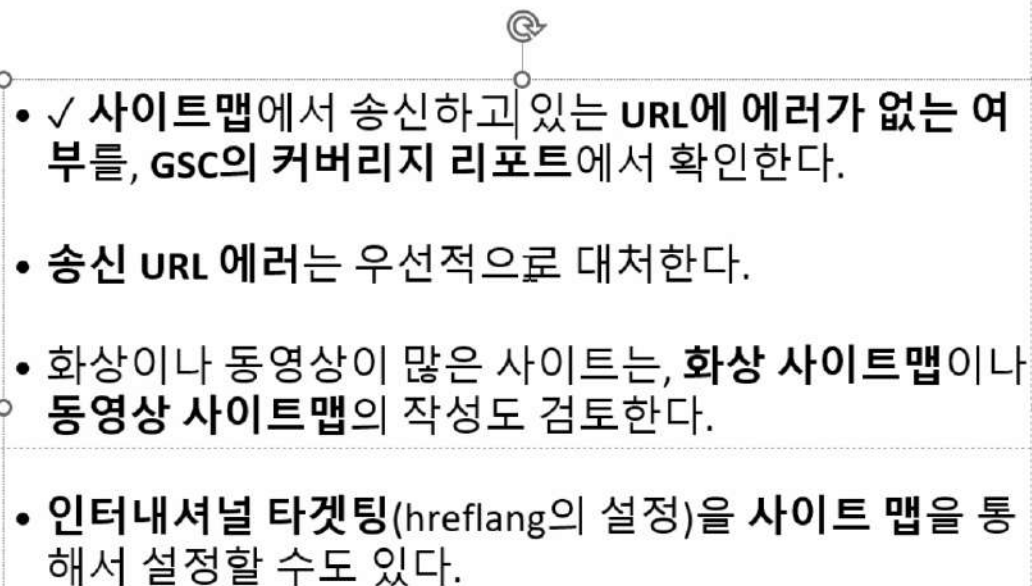
• 참고 자료

- (유튜브) [Streaming Frog 소개 영상](#)

- ✓ **GSC [인덱스] → '사이트맵' 항목을 확인한다.**
 - ※ 만약 정상적으로 송신되고 있는가?
 - ※ 오류는 없는가?
- ✓ **domain/sitemap.xml로 존재 유무 확인한다.**
 - ※ 사이트맵 파일명은 사이트마다 다를 수 있으므로, 크롤러가 못찾는 경우가 종종 있다.
- **robots.txt에 기술되어 있고, 또 GSC에서 송신하고 있는 사이트맵을 확인한다.**
- ✓ **사이트맵 인덱스 파일과 XML사이트맵 속성을 확인한다.**
 - **lastmod** → 올바른 갱신 날짜(실제 갱신 날짜)로 기술되어 있는가?
 - 일본어 및 한국어는 **utf-8 인코딩**되어 있는가?
 - 특수문자는 **이스케이프**되어 있는가?
- **priority와 changefrep은 Google에서 중요하지 않은 태그이므로, 삭제 대응도 가능하다.**

참고 자료

- [\(구글 search central\) 이미지 사이트맵](#)
- [\(구글 search central\) 페이지의 현지화 버전에 대해 Google에 알린다.](#)

- 
- ✓ 사이트맵에서 송신하고 있는 URL에 에러가 없는 여부를, GSC의 커버리지 리포트에서 확인한다.
 - 송신 URL 에러는 우선적으로 대처한다.
 - 화상이나 동영상이 많은 사이트는, **화상 사이트맵**이나 **동영상 사이트맵**의 작성도 검토한다.
 - **인터내셔널 타겟팅**(hreflang의 설정)을 사이트맵을 통해서 설정할 수도 있다.

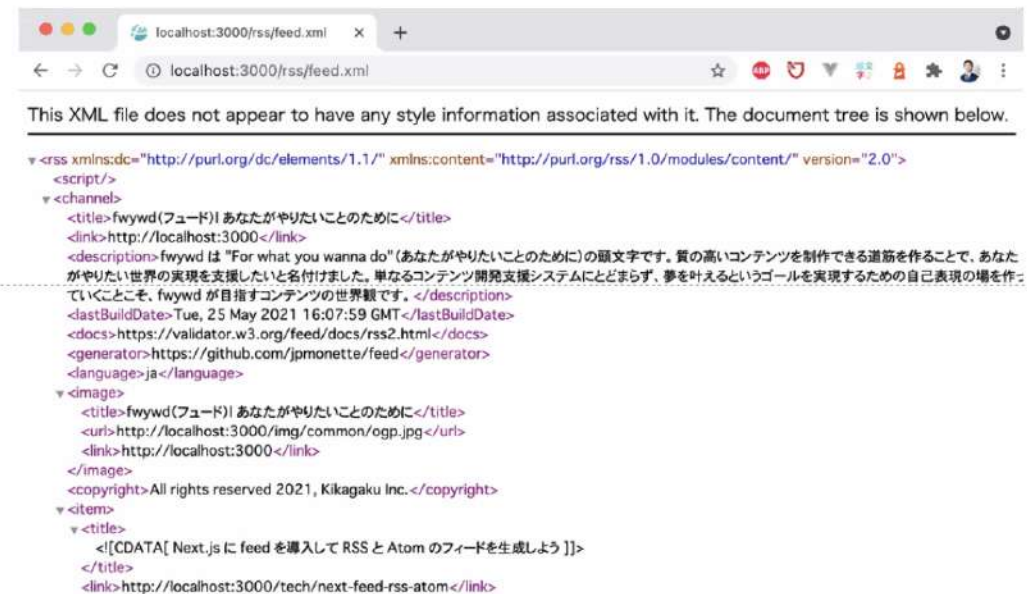
• 참고 자료

- [\(구글 search central\) 이미지 사이트맵](#)
- [\(구글 search central\) 페이지의 현지화 버전에 대해 Google에 알린다.](#)

- (구글 SEO 정책상) **noindex** 대응의 페이지를 Google에 인식시키는 경우,
우선적으로 크롤러 순회를 원한다면, **사이트맵 송신**과 병행해 **RSS/ATOM 피드**의 검용을 제안하는 것을 권장한다.

참고 자료

- [\(구글 search central - 일본어\) XML 사이트 맵과 RSS/Atom 피드의 베스트 프랙티스](#)



- 반드시 동적일 필요는 없지만, **사이트의 갱신 빈도가 높은 경우는 동적인 생성/갱신을 검토한다.**
- 추가로, **RSS/ATOM 피드와 겸용도 검토한다.**

참고 자료

- [\(구글 search central - 일본어\) XML 사이트 맵과 RSS/Atom 피드의 베스트 프랙티스](#)

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<urlset xmlns="https://www.sitemaps.org/schemas/sitemap/0.9" xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance">
  <url>
    <loc>https://www.geo-code.co.jp/</loc>
    <priority>1.0</priority>
    <changefreq>daily</changefreq>
  </url>
  <url>
    <loc>https://www.geo-code.co.jp/enterprise/</loc>
    <priority>0.8</priority>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>https://www.geo-code.co.jp/enterprise/internet_advertisement/</loc>
    <priority>0.8</priority>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>https://www.geo-code.co.jp/enterprise/web_integration/</loc>
    <priority>0.8</priority>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>https://www.geo-code.co.jp/enterprise/internet_asp/</loc>
    <priority>0.8</priority>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>https://www.geo-code.co.jp/enterprise/internet_media/</loc>
    <priority>0.8</priority>
    <changefreq>weekly</changefreq>
  </url>
</urlset>
```

- ✓ domain/robots.txt의 기술을 확인하고 과부족이 없는지 확인한다.
- 특히, 화상이나 CSS, JS를 블록 하거나 하고 있는 경우, 페이지의 정확한 읽기에 필요한 경우도 있으므로, 자세히 알아본다.
- ✓ GSC의 커버리지 리포트의 경고나, 제외의 「차단되었습니다」도 확인한다.
- ✓ robots.txt에 기술하더라도, 다른 사이트로부터 링크 되고 있는 URL은 「경고」 취급이 되어, 인덱스 되어 버린다.
- ✓ 즉, 경고 취급이 되는 경우에는 차단하는 것이 아니라 noindex로 하는 등의 대처가 필요하다.

참고 자료

- [\(구글 search central\) robots.txt 파일에 대하여](#)
- [\(구글 search central\) Robots.txt 사양](#)
- [\(구글 search central\) 인덱스 커버리지 리포트](#)

- ✓ 로그인 계열 페이지의 헤드 태그에는 **noindex** 처리한다.

I

- ✓ **noindex**와 **canonical**이 겸용으로 사용되지 않도록 **noindex**를 사용하여 검색 인덱스 등록을 차단한다.

- 참고 자료

- [\(구글 search central\) noindex로 검색 색인 생성 차단하기](#)

noindex 구현

noindex 는 <meta> 태그 및 HTTP 응답 헤더의 두 가지 방법으로 구현할 수 있습니다. 두 방법의 효과는 동일하며 사이트에 더 편리하고 콘텐츠 유형에 적절한 방법을 선택하면 됩니다. Google은 robots.txt 파일에서 noindex 규칙을 지정하는 것을 지원하지 않습니다.

noindex 규칙을 색인 생성을 제어하는 다른 규칙과 결합할 수도 있습니다. 예를 들어 nofollow 힌트를 noindex 규칙인 <meta name="robots" content="noindex, nofollow" /> 와 결합할 수 있습니다.

<meta> 태그

noindex 규칙을 지원하는 모든 검색엔진이 사이트 페이지에서 색인을 생성하지 않도록 하려면 다음 <meta> 태그를 페이지의 <head> 섹션에 배치하세요.

```
<meta name="robots" content="noindex">
```



- ✓ http→https
- ✓ /index.html ▪ /index.php→/ I
- ✓ **www**의 유무 액세스 시, **리다이렉트 체크 툴** 등을 사용하여 리다이렉트 상황을 확인한다.
- ✓ **www** 있음/없음 등 이외에 개발 환경 등이 존재하는 경우도 종종 있으므로, 어떤 관련 도메인이 있는지는 확인하는 것이 좋다.
- ✓ 가능하다면 관련 사이트 **GSC**(구글 서치 콘솔)를 보면 **리다이렉트**가 처리되고 있는지 등의 이슈를 추가로 확인할 수 있다.
- ✓ 만약 통합하려면 **301 리다이렉트**가 이상적이지만, **canonical**로 대체하는 것도 가능하다.

- 참고 자료
 - [\(일본어\) 리다이렉트 체크 툴 예시](#)

- ✓ 현재는 사용하지 않는 이전 페이지가 존재하는 경우, 현재 페이지로 제대로 리다이렉트 하고 있는지 확인한다.
- 방법은, ScreamingFrog → Report → RedirectChain 을 확인한다.
- ✓ GSC의 리다이렉트 에러나, 리다이렉트 되고 있는 페이지군을 체크한다.
- ✓ Screaming Frog에서 리다이렉트 리포트를 확인한다.
- ✓ 리다이렉트가 여러 개 얹히거나 302 리다이렉트 → 404페이지 등의 경우는 제외된 '크롤 에러'로 분류되는 경우도 있다.
- ※ 404 페이지에 301 리다이렉트 하는 것은 Google도 인정하고 있는 방법이다. (직접 404 반환이 베스트 프랙티스)

참고 자료

- [\(구글 search central\) 자바스크립트 검색 엔진 최적화의 기본사항 이해하기](#)

- ✓ 기본 리다이렉트는 301이지만, 아래 경우에는 302 리다이렉트를 사용한다.

- ✓ 모바일 버전과 PC 버전의 사이트를 동시에 가진다.

- ✓ User Agent에 따라 PC로부터의 액세스를 모바일 버전의 사이트로 리다이렉트 시키거나 할 경우에는, 기본은 리다이렉트는 301이지만, 특정 경우는 302 리다이렉트를 사용한다.

- 모바일 버전과 PC 버전의 사이트가 있으며, User Agent에 따라 PC로부터의 액세스를 모바일 버전의 사이트로 리다이렉트 시키거나 하는 경우

- 참고 자료

- [\(구글 search central\) 다른 URL](#)

- A/B 테스트에서 원래 URL에서 패턴 URL로 사용자를 리다이렉트 시 301 리다이렉트가 아닌, 302 리다이렉트 사용을 구글은 권장하고 있습니다.

- 참고 자료

- [\(구글 search central\) Google 검색에서 웹 사이트를 테스트하는 경우 추천 방법](#)

- 리다이렉트가 5회 이상인 경우도, 크롤 하지 않는 것은 아닌 것 같다.

- 일단 크롤을 멈추고 나중에 재 크롤을 시도할 것.

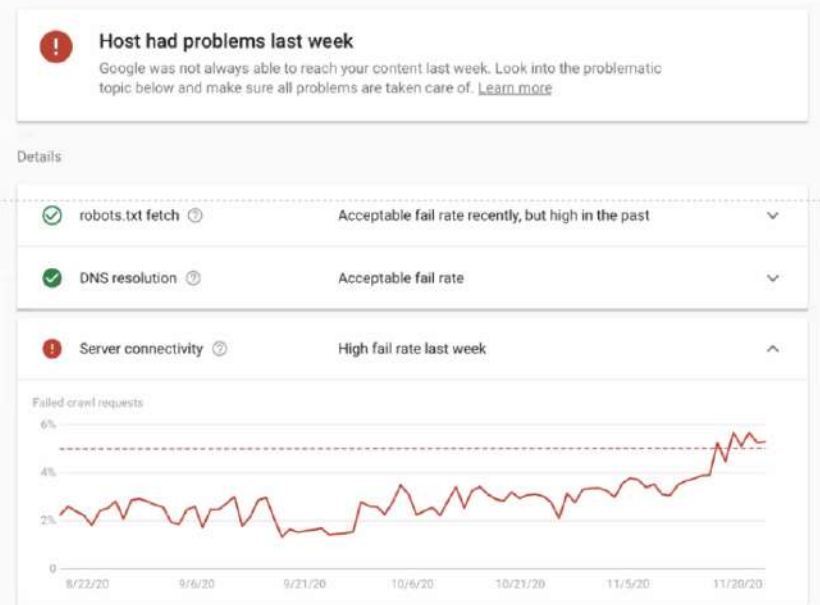
- ✓GSC → 설정 → 크롤 통계 정보를 확인한다.
- 평균 응답시간이 200밀리 초 이하가 이상적이지만, 극단적으로 과도한 수치(1000 밀리초 이상)만 아니면 문제없다.



- ✓호스트 상태에 문제가 있는지 확인한다.
- 도메인의 모든 호스트의 성능을 한 곳에서 평가한다.
- 사이트 크롤의 최신 통계 정보를 확인한다.
- 참고 자료
 - (구글 search central) 새롭게 개선된 사이트 크롤링 통계

호스트 상태 문제에 관한 전체적인 개요 및 세부정보

보고서의 호스트 상태 세부정보를 통해 최근 90일 동안 Google에서 확인한 사이트의 일반적인 가용성을 볼 수 있습니다.



- ✓ 로컬라이즈 페이지를 분석하기 위해서, 크롤러는 hreflang 태그를 참조한다.
- 가능한 모든 대상 페이지에 hreflang 태그를 사용한다.
- 참고 자료
 - [\(구글 search central\) 페이지의 현지화 버전에 대해서](#)

HTTP 헤더

페이지의 GET 응답을 통해 HTTP 헤더를 반환하여 Google에 페이지의 모든 언어 및 지역 버전을 알릴 수 있습니다. PDF와 같이 HTML이 아닌 파일에 유용합니다.


다음은 헤더의 형식입니다.

Link: <url1>; rel="alternate"; hreflang="lang_code_1", <url2>; rel="alternate"; hreflang="lang_code_2", ...

구문

<url_x>	연결된 hreflang 속성에 할당된 언어 문자열에 해당하는 대체 페이지의 정규화된 URL. URL은 < 및 > 표시로 둘러싸여 있어야 합니다. 예: <https://www.google.com>
lang_code_x	페이지의 이 버전에서 타겟팅하는 <u>지원되는 언어/지역 코드</u> 또는 페이지의 hreflang 태그에 의해 명시적으로 나열되지 않은 모든 언어에 해당하는 x-default.



-  무한 스크롤 하부 콘텐츠를 구글이 인식하기 쉬워진다.
- 링크를 따라가면 검색엔진은 특정 페이지에 대해 명시적으로 접근하기 쉬워진다.
- 스크롤 후 URL이 변화하고 있는지 여부를 확인한다.
(?page=x : 파라미터 포함 URL 부여 여부)
- 각 페이지의 중복 여부를 확인한다.
- 참고 자료
 - [\(구글 search central- 일본어\) 무한 스크롤 검색 친화적 추천방법](#)

