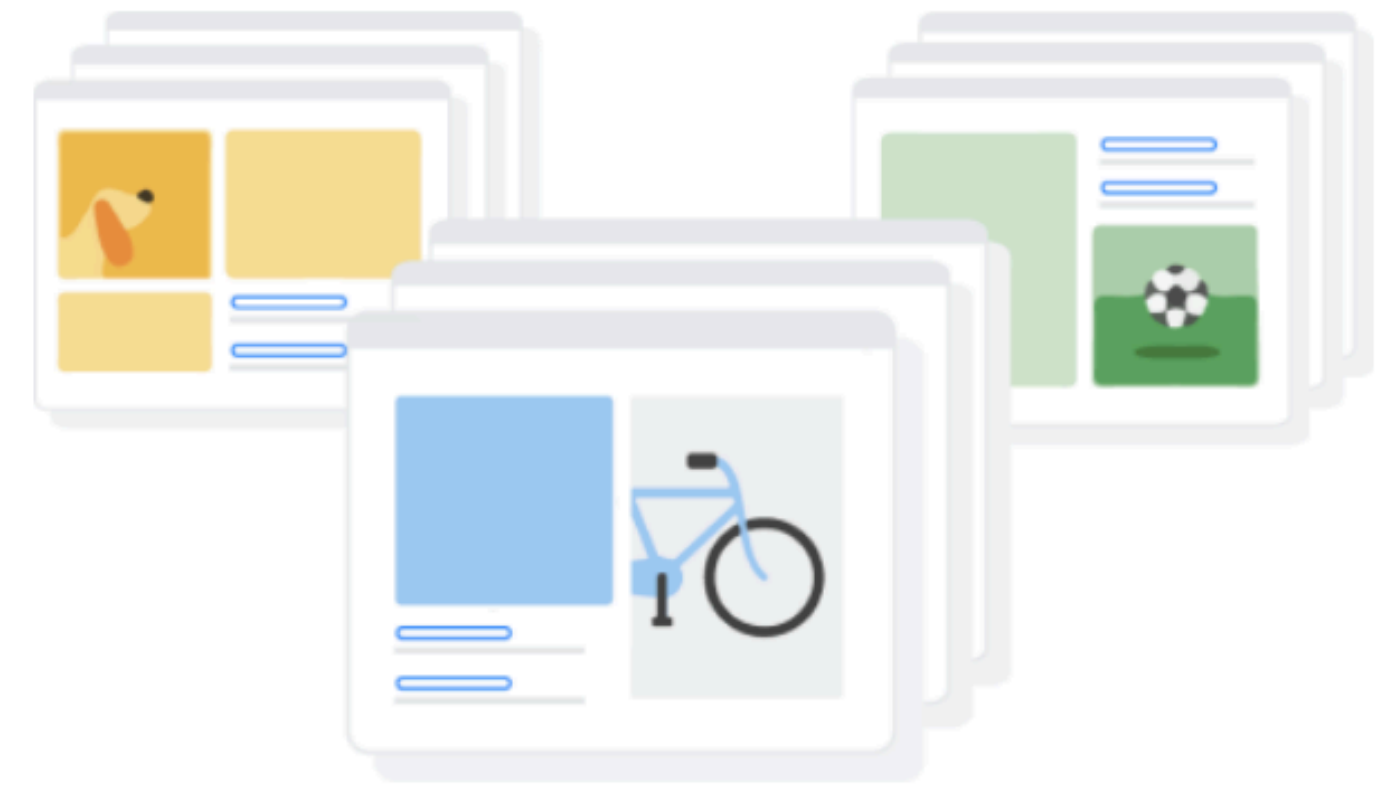


# Indeksowanie stron internetowych przez Google

Stanisław Greń

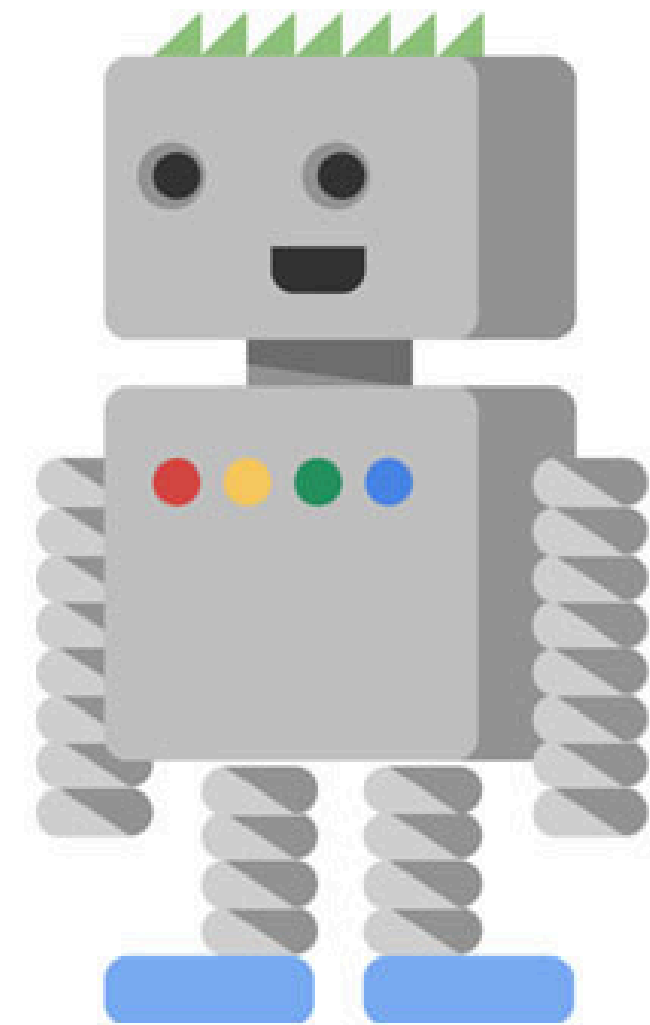
# Czym jest indeksowanie?

- Jest to proces w którym robot indeksujący zbiera informacje o stronie oraz zapisuje ją do bazy (indeksu)
- Proces ten umożliwia późniejsze wyświetlanie strony w wynikach wyszukiwania
- Pozwala na ocenę jakości strony, aby później odpowiednio ją pozycjonować



# Googlebot

- Robot indeksujący Google (nazywany też botem, lub pajakiem) ustala priorytety stron.
- Dzieli się na dwa typy: desktopowy oraz mobilny
- Przechodzi z jednej na drugą stronę, używając odnośników, tak tworzy sieć połączeń
- Przeszukuję miliardy stron dziennie, popularniejsze z nich, nawet kilkanaście razy dziennie



# Jakie są etapy indeksowania?

**01**

**Crawling**  
(Skanowanie)

**02**

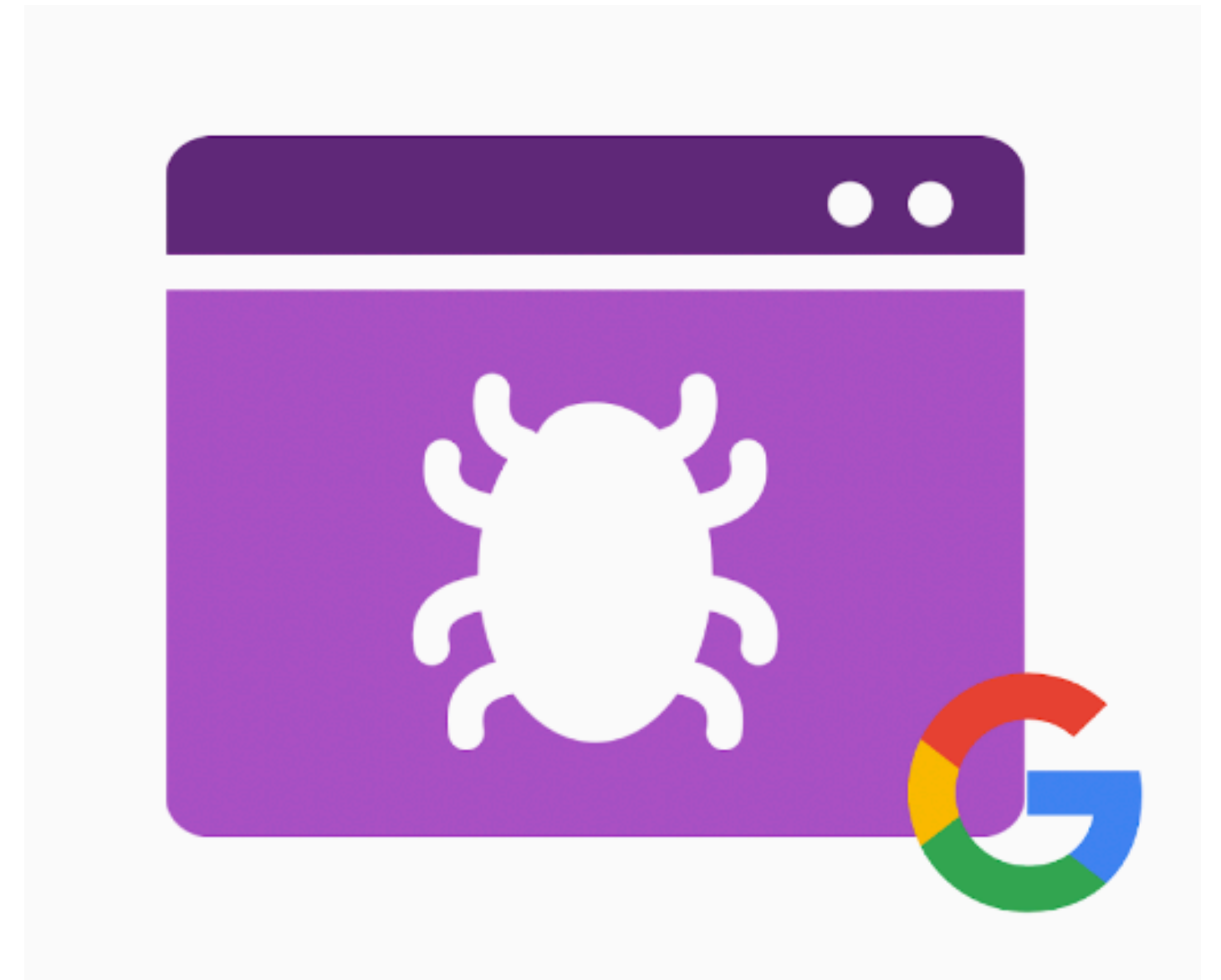
**Indexing**  
(Analiza i zapis do indeksu)

**03**

**Wyświetlenie  
Wyników**

# Crawling

- Robot wyszukuje stronę, poprzez odwiedzanie kolejnych linków (wewnętrznych oraz zewnętrznych),
- Strona jest pobierana przez robota,
- Robot renderuje stronę, w oparciu o najnowszą wersję przeglądarki Chrome



# Indexing

- Rozpoznanie tematyki strony,
- Google analizuje strone w oparciu o tagi takie jak: title, meta (noindex, nofollow), nagłówki, czy atrybuty tagów img (alt, text),
- Google określa czy strona nie jest duplikatem, czy jest kanoniczna,
- Decyzja o zapisaniu strony w indeksie.

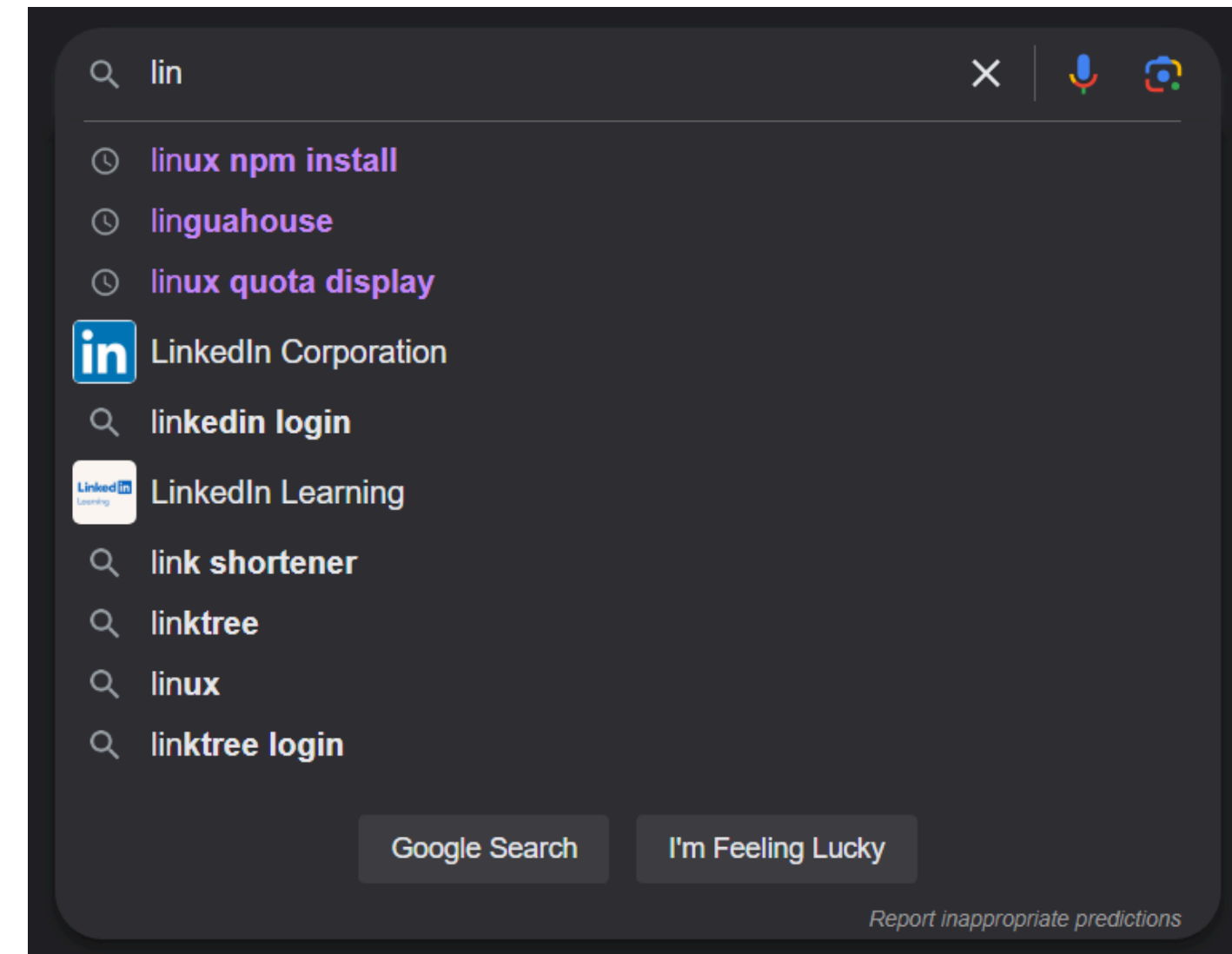
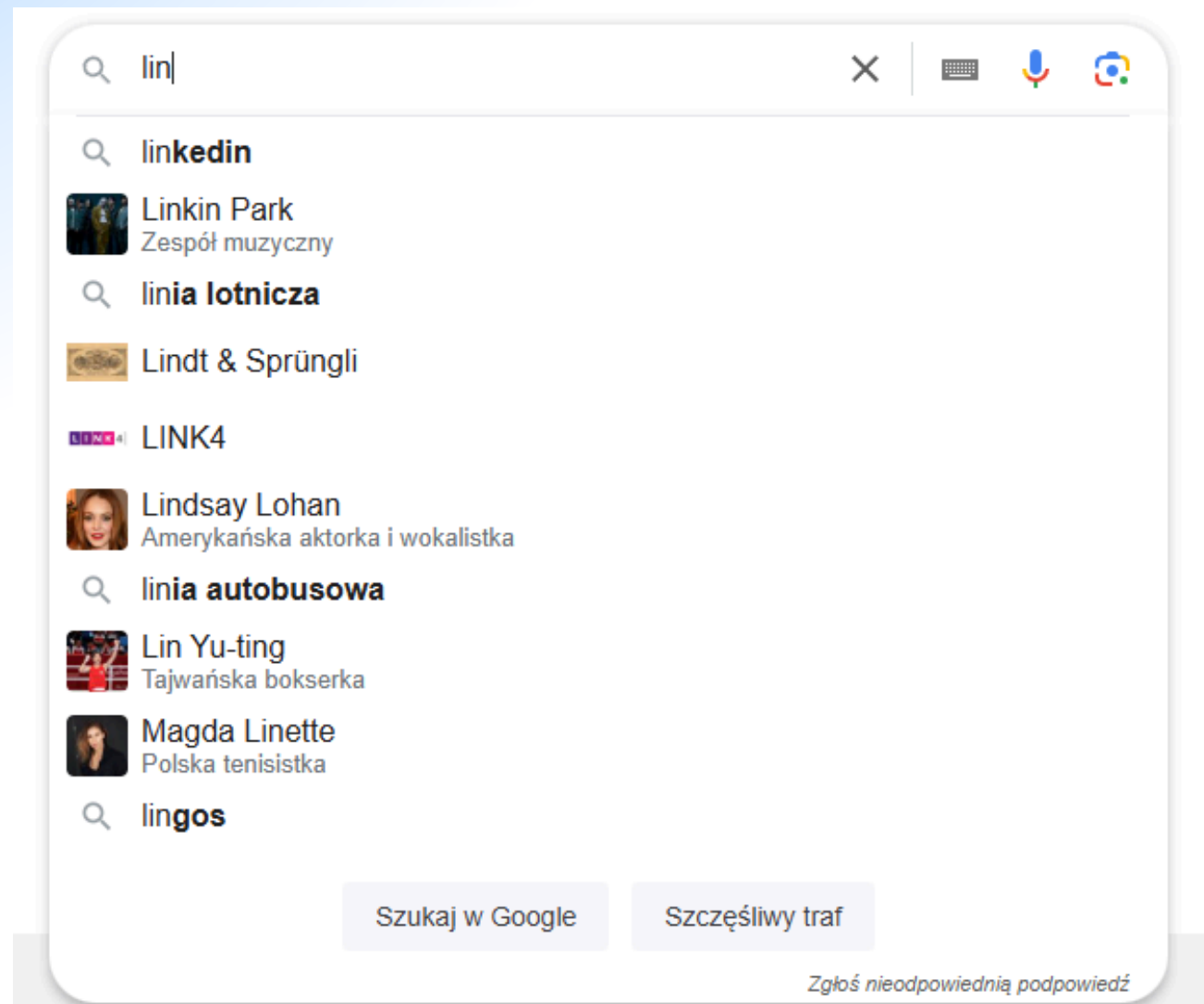


# Wyświetlanie wyników

- Użytkownik po wpisaniu konkretnego hasła w wyszukiwarce otrzymuje wyniki, na podstawie indeksu.
- Pozycjonowanie strony zależy od wielu czynników takich jak:
  - Lokalizacja
  - Język
  - Jakość



# Porównanie dwóch wyszukiwarek





# Kluczowe elementy techniczne



## robots.txt

Plik instruujący robota, które obszary witryny mogą być przeszukiwane.



## sitemap.xml

Plik zawierający listę podstron serwisu, jest mapą serwisu dla robota.



## Jakość

Unikalność, oraz świeżość treści zawartych na stronie pozytywnie wpływa na indeksowanie



## Linki

Pomagają robotą poruszać się po naszej stronie, oraz tworzyć mapę powiązań

# robots.txt

 robots.txt - Notepad

File Edit Format View Help

---

```
User-agent: *
```

```
Disallow: /wp-admin/
```

```
Disallow: /wp-includes/
```

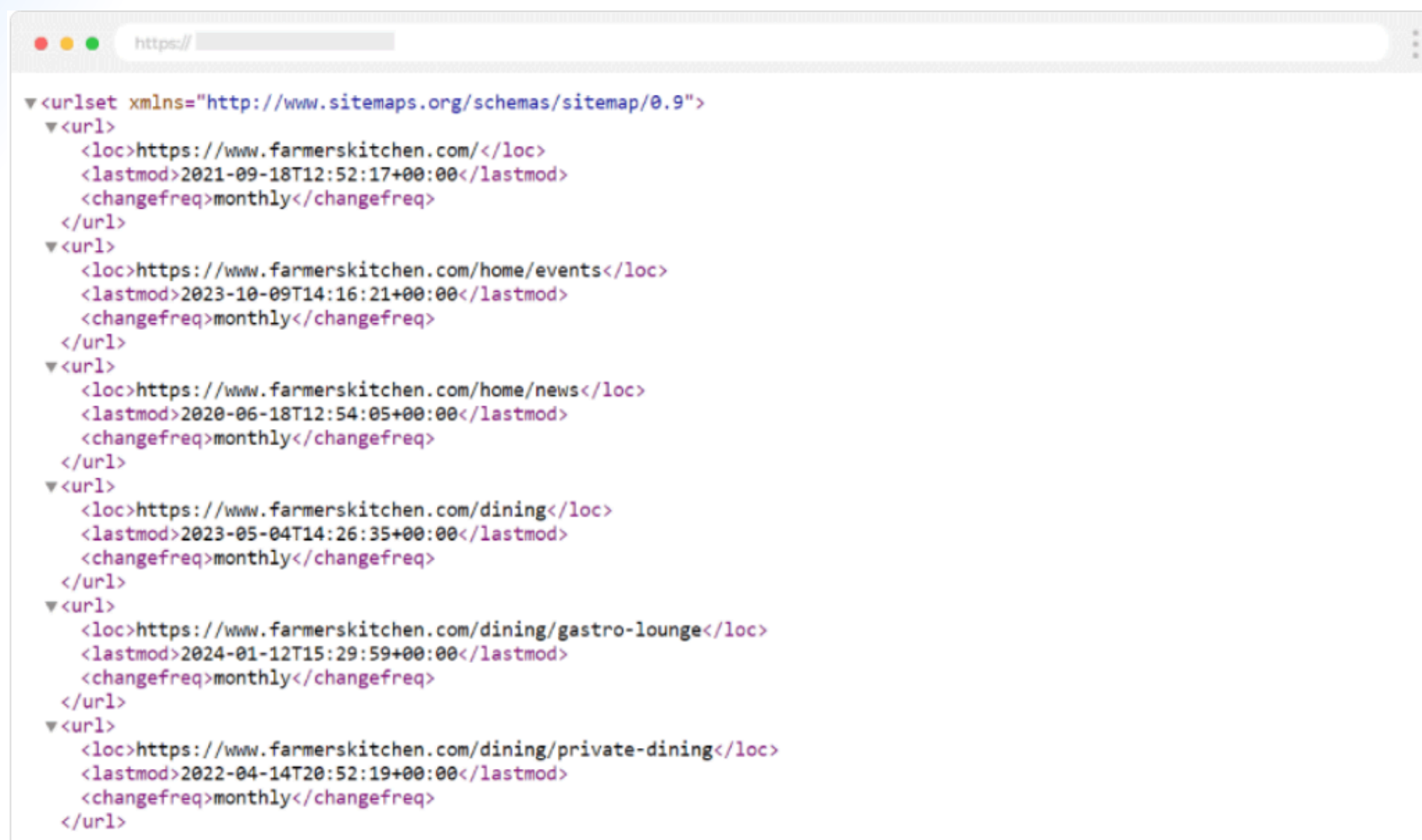
```
Disallow: /wp-content/plugins/
```

```
Disallow: /wp-content/themes/
```

```
Sitemap: http://www.example.com/sitemap.xml|
```

© GeeksForGeeks

# sitemap.xml



```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://www.farmerskitchen.com/</loc>
    <lastmod>2021-09-18T12:52:17+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>https://www.farmerskitchen.com/home/events</loc>
    <lastmod>2023-10-09T14:16:21+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>https://www.farmerskitchen.com/home/news</loc>
    <lastmod>2020-06-18T12:54:05+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>https://www.farmerskitchen.com/dining</loc>
    <lastmod>2023-05-04T14:26:35+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>https://www.farmerskitchen.com/dining/gastro-lounge</loc>
    <lastmod>2024-01-12T15:29:59+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>https://www.farmerskitchen.com/dining/private-dining</loc>
    <lastmod>2022-04-14T20:52:19+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
</urlset>
```

© htmlburger.com

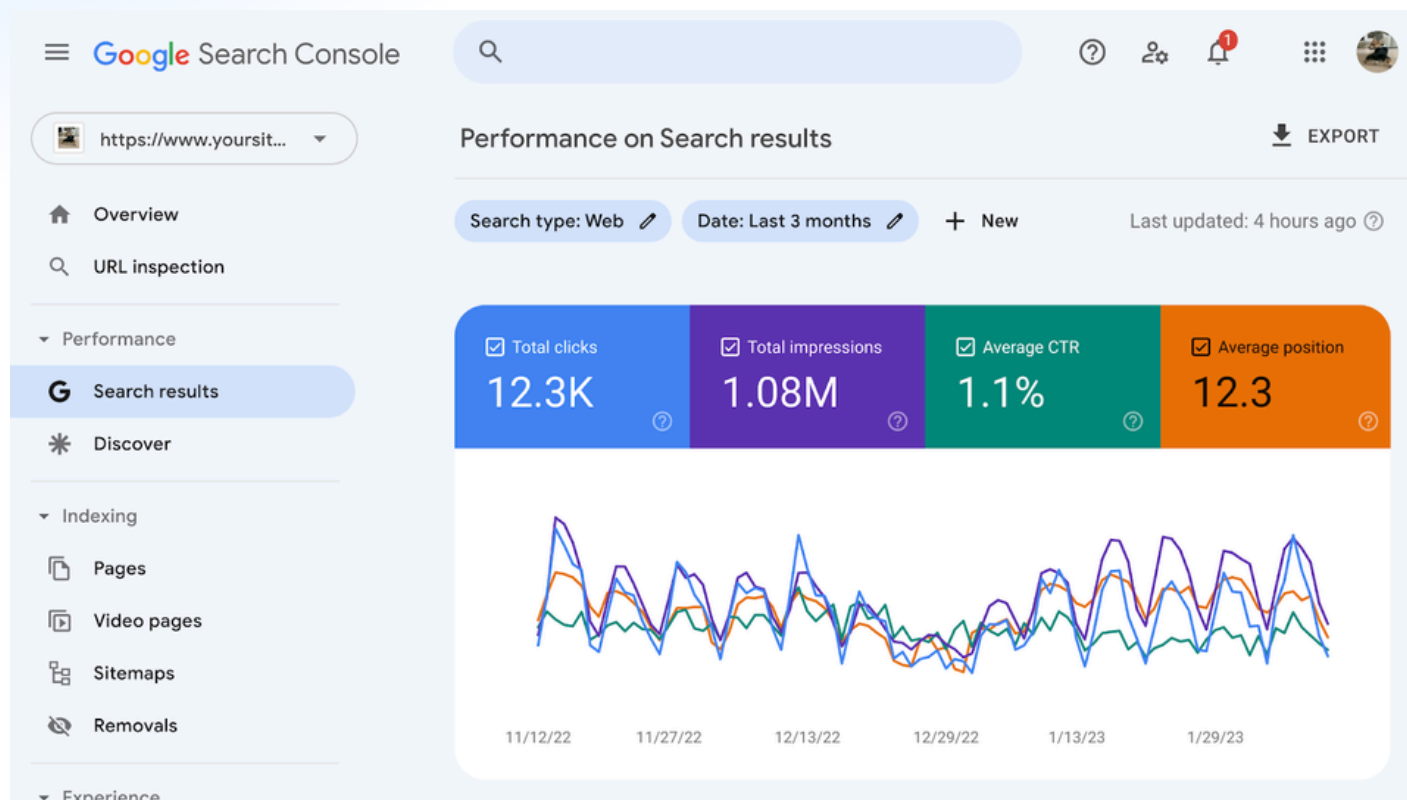
# Wyzwania przy indeksowaniu

- Treści generowane dynamicznie - Treści generowane przez JavaScript mogą opóźniać proces
- Duplikaty treści - Strona może być potraktowana jako spam
- Problemy techniczne - Brak kluczowych elementów wpływających na indeksowanie
- Wydajność

# Pomocne narzędzia

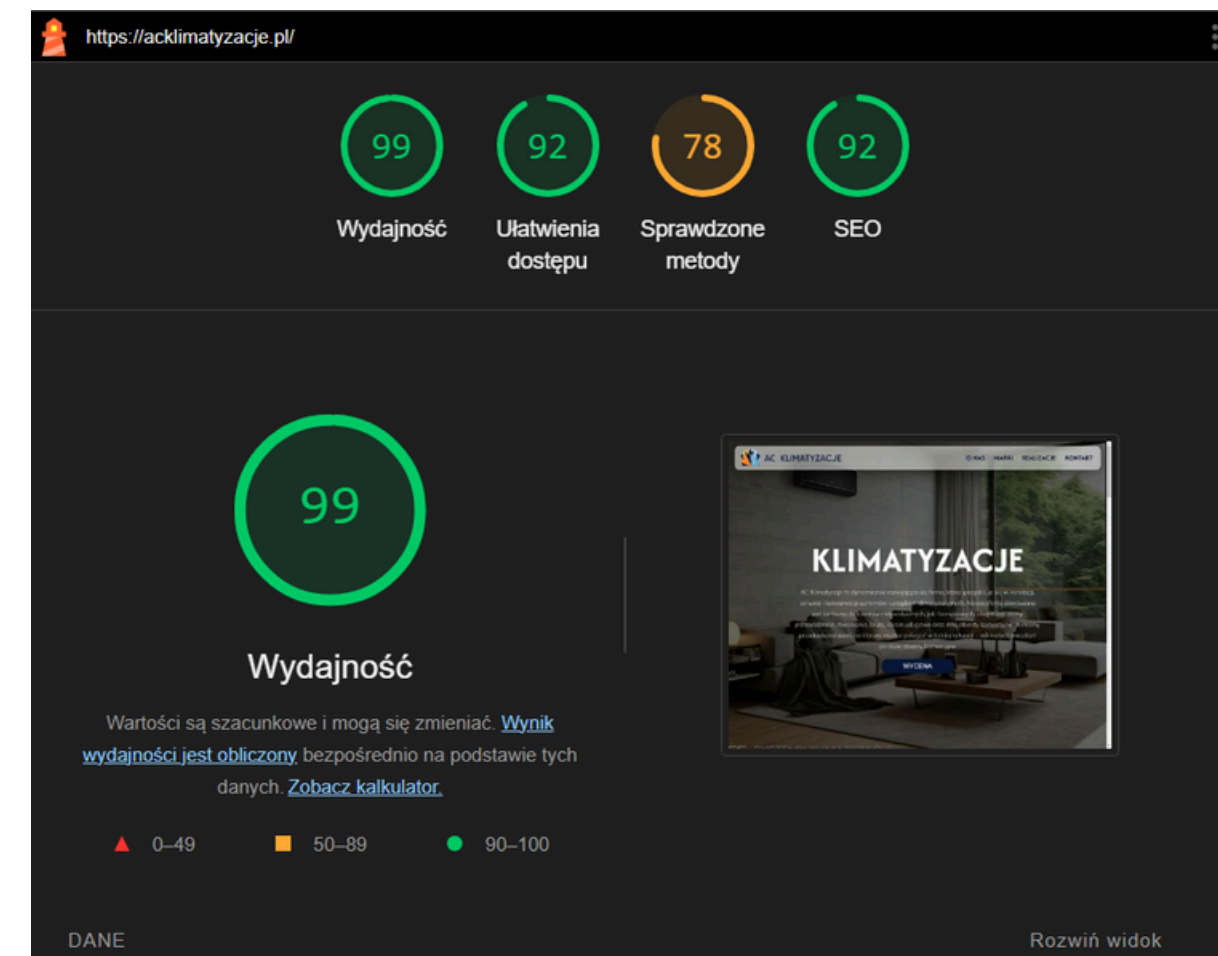
- Google Search Console

Pozwala monitorować statystyki, oraz dostarcza informacje o indeksowaniu.



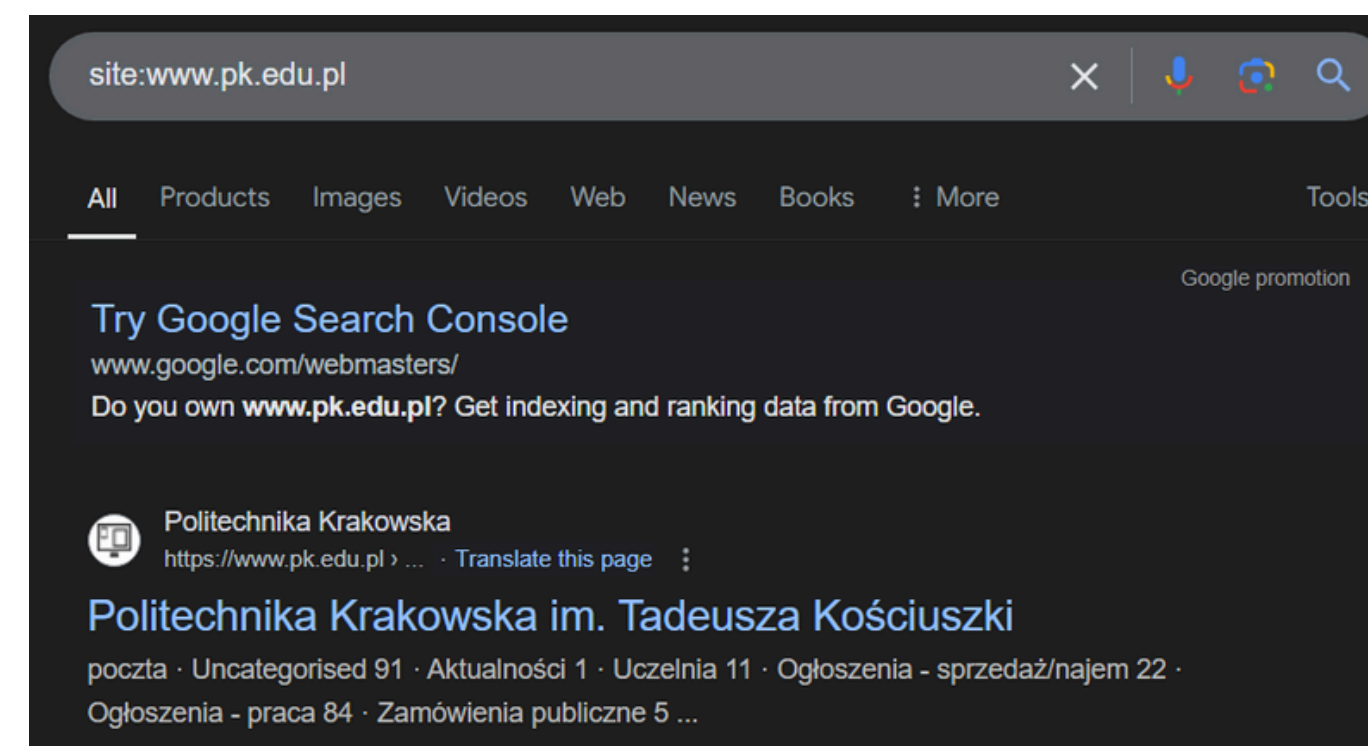
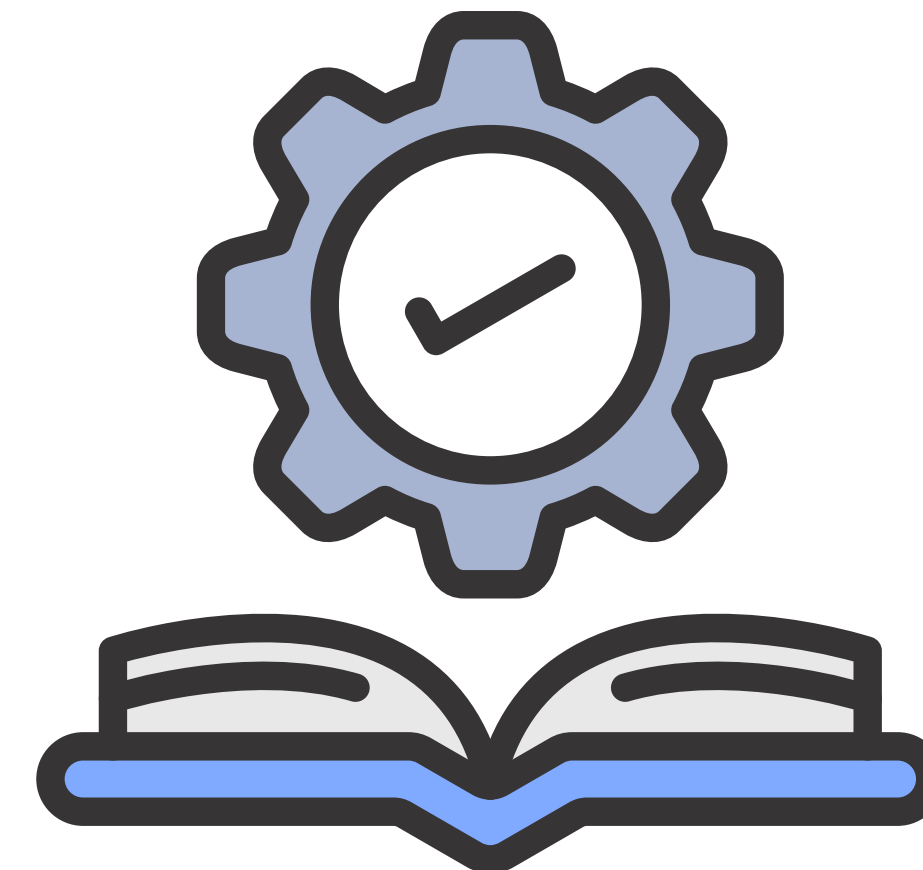
- Lighthouse

Testuje naszą stronę pod kątem wydajności oraz przyjazności robotom.



# Dobre praktyki

- Unikanie nadmiernego blokowania treści w pliku robots,
- Aktualizowanie mapy strony,
- Prerendering / Server-side rendering
- Optymalizacja techniczna
- Sprawdzenie czy nasza strona została zaindeksowana



**Dziękuję za uwagę**

# Bibliografia

- In-depth guide to how Google Search work:  
<https://developers.google.com/search/docs/fundamentals/how-search-works?hl=pl>
- Organizing Informations: <https://www.google.com/search/howsearchworks/how-search-works/organizing-information/>
- Co to jest skanowanie stron – crawling: <https://www.sempire.pl/co-to-jest-skanowanie-stron-crawling.html>
- <https://edgemes.com/blog/how-google-crawls-your-website-and-indexes-your-pages>
- <https://www.semrush.com/blog/google-search-console/>