

# Indeksowanie stron internetowych przez Google

Hanna Gancarz  
Marcin Kurdziel  
Grzegorz Porwoł



# “Dlaczego?” indeksowania

- Porządkowanie treści w Internecie
- Dopasowanie wyniku do wyszukiwania
- Zapewnienie jakości wyników wyszukiwania



# Plan prezentacji

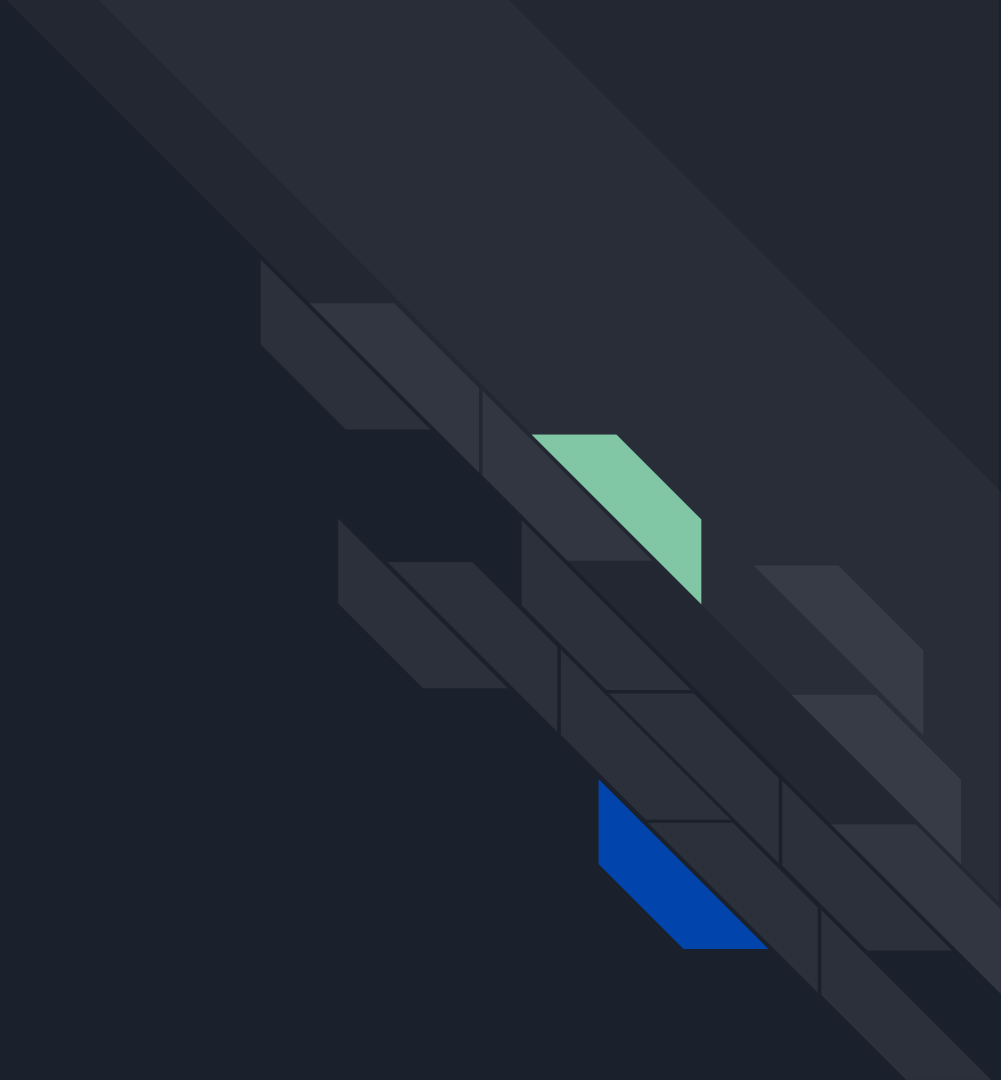
- Index Googla
- Algorytmy wyszukiwania
- Search console
- Indexing API



Polecana strona

<https://www.google.com/search/howsearchworks/>

Index Googla





# Ideologia indeksowania przez Google

- Odseparowanie reklam od wyników wyszukiwania
- Brak możliwości zakupu lepszej pozycji w wynikach wyszukiwania



Najważniejsza informacja

Wyszukiwanie przeszukuje zbudowany wcześniej indeks



# Budowanie indeksu

- Właściciele stron -> dostarczają mapy witryn
  - mapa witryny - plik z powiązaniem oraz informacjami
- Algorytmy Google -> przeszukuje publicznie dostępne strony za pomocą robotów indeksujących
  - robot indeksujący
    - analizuje strony internetowe
    - otwiera linki
    - przesyła dane o stronach do Google





# Gdy robot indeksujący znajdzie stronę

System Google skanuje treść strony jak przeglądarka, zwracając uwagę na m.in.

- słowa kluczowe
- aktualności strony
- linki
- pliki

Informacje te są rejestrowane w indeksie wyszukiwania



Czy wiesz, że...

Indeks wyszukiwarki Google ma rozmiar ponad  
100 000 000 gigabajtów

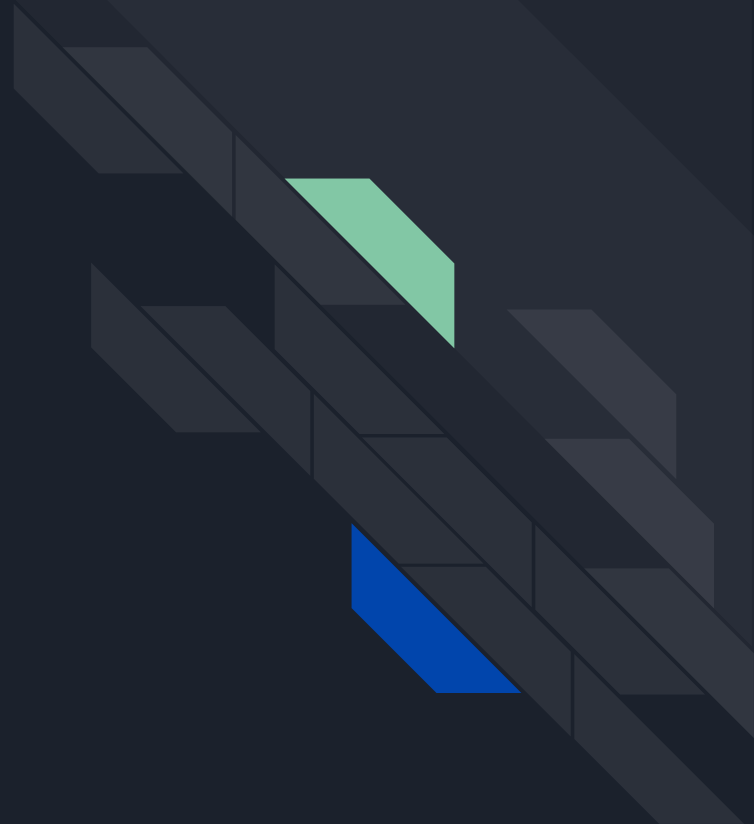


# Nie tylko strony internetowe

Poza stronami, Google porządkuje także m.in:

- Miliony książek z dużych bibliotek
- Rozkłady jazdy lokalnych środków komunikacji publicznej
- Dane ze źródeł publicznych (np. z Banku Światowego)

# Algorytmy wyszukiwania





## Główna idea algorytmu wyszukiwania

W sieci jest za dużo informacji - bez wcześniejszego uporządkowania nie można byłoby znaleźć tego, czego potrzebujemy

Najpierw porządkowanie, potem wyszukiwanie



# System rankingowy

- Wiele algorytmów sprawdzających wiele czynników
- Sprawdzają m.in.
  - treść zapytania
  - trafność i przydatność stron
  - wiarygodność źródła
  - lokalizacja
  - czas zamieszczenia
- Waga czynnika może być uzależniona od treści zapytania



# Zapewnienie wysokiego standardu wyników

- Ciągłe usprawnianie algorytmów
- Testy na żywo przez zewnętrznych testerów
- Rygorystyczne wytyczne testowania



# Główne czynniki wpływające na wyniki zapytania

- Analiza słów w zapytaniu
  - interpretacja literówek
  - system synonimów
  - określenie kategorii informacji
  - czy potrzebne są aktualne informacje?
  
- Dopasowanie do zapytania
  - wyszukiwanie słów z zapytania w indeksie
  - częstotliwość wystąpień słów kluczowych w tytule, nagłówkach i treści
  - sprzężenie zwrotne (anonimowe)
  - brak analizy pojęć subiektywnych

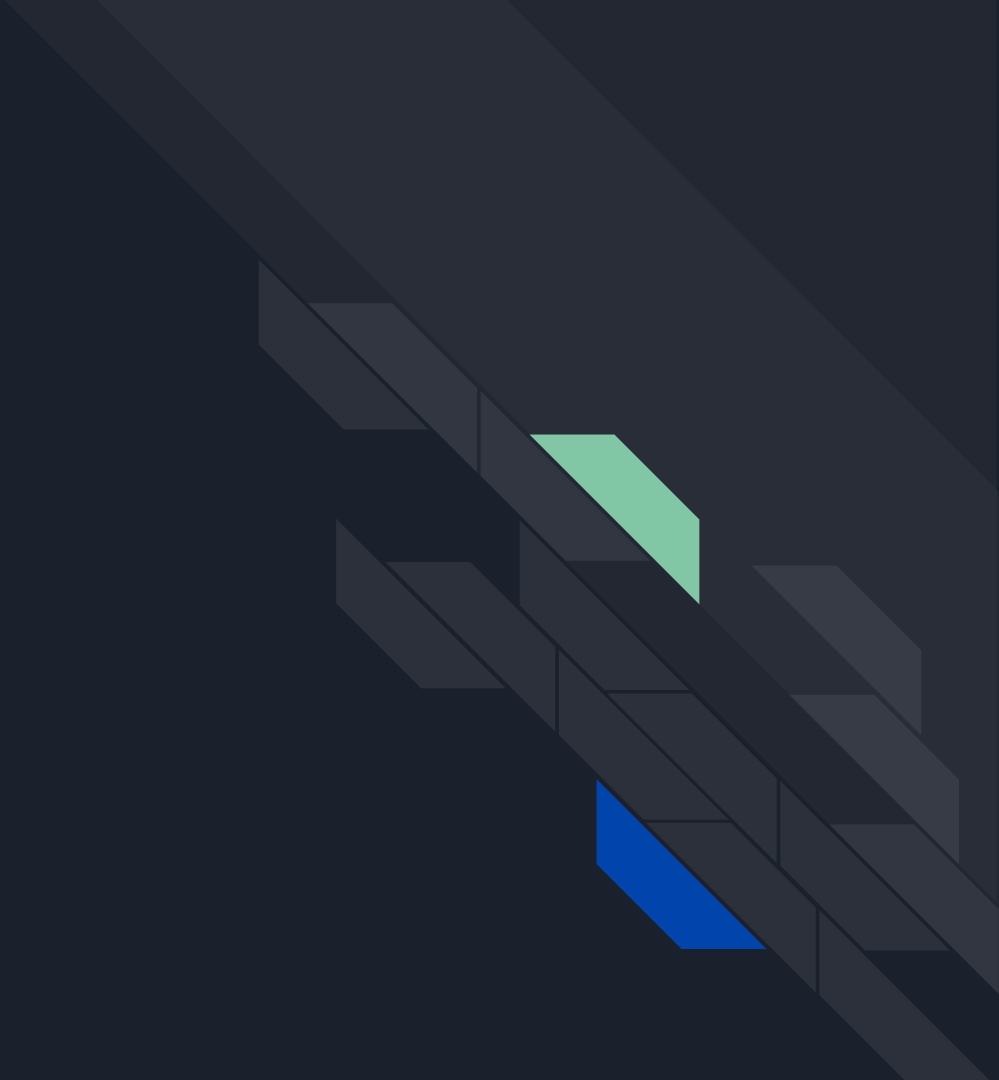




# Główne czynniki wpływające na wyniki zapytania

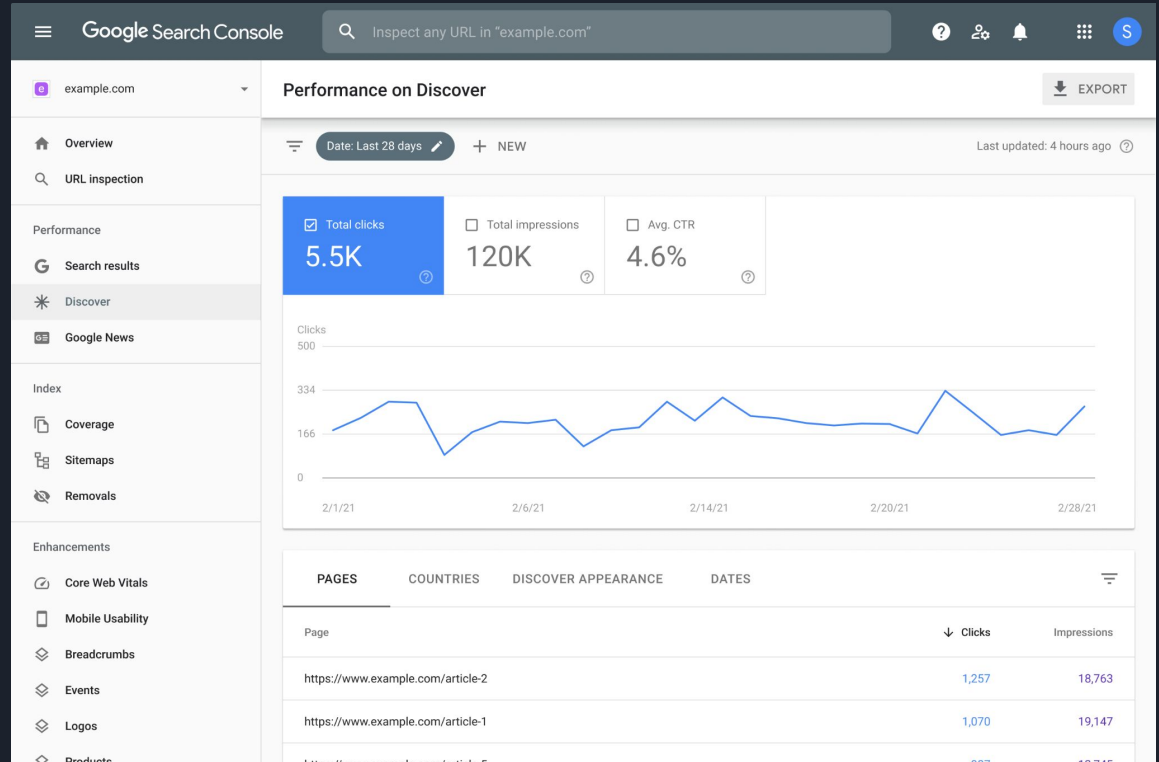
- Ranking przydatnych stron
  - sortowanie od najbardziej do najmniej trafnych
  - rozpoznawanie spamu
- Pokazywanie najlepszych wyników
  - ocena powiązań między wynikami - zróżnicowanie informacji
  - poprawność wyświetlania na różnych urządzeniach
  - czas wczytywania
- Dostrzeganie kontekstu
  - lokalizacja, historia wyszukiwań oraz ustawienia wyszukiwarki
  - dopasowanie wyników do konkretnego użytkownika

Search console



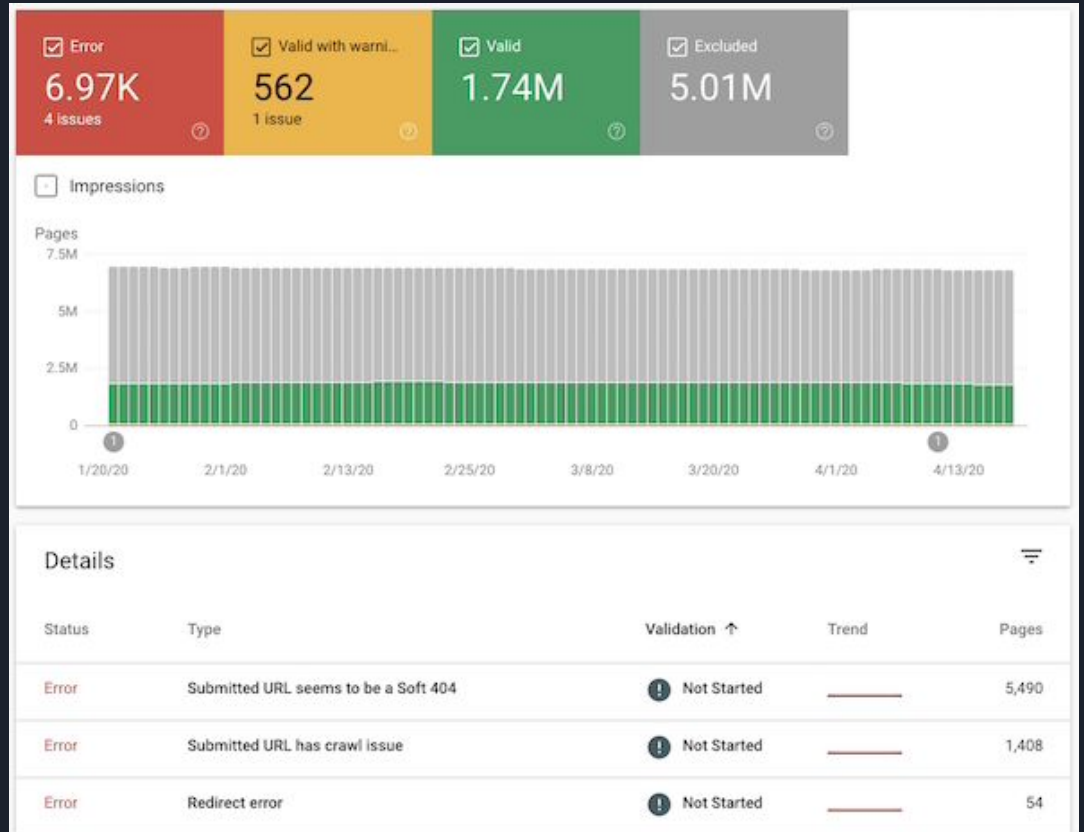
# Search console

1. Nauka w jaki sposób Google wykonuje crawling, indeksowanie i przeszukiwanie
2. Identyfikacja i naprawa błędów
3. Zatwierdzanie zaktualizowanej treści
4. Obserwacja i optymalizacja wydajności szukania



# Index Coverage Report

1. Error
2. Valid and warnings
3. Valid
4. Excluded



# Indexing API





# Indexing API

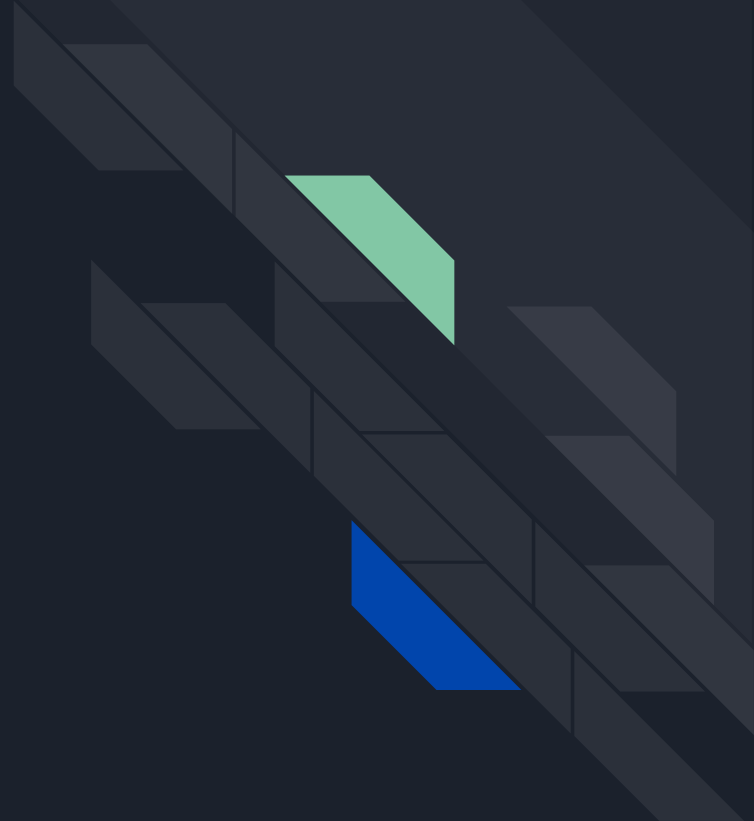
- aktualizacja URL

```
Send the following HTTP POST request to https://indexing.googleapis.com/v3/urlNotifications:publish
{
  "url": "https://careers.google.com/jobs/google/technical-writer",
  "type": "URL_UPDATED"
}
```

- usunięcie URL

```
Send the following HTTP POST request to https://indexing.googleapis.com/v3/urlNotifications:publish
{
  "url": "https://careers.google.com/jobs/google/technical-writer",
  "type": "URL_DELETED"
}
```

# Podsumowanie





# Podsumowanie

- Indeksowanie służy porządkowaniu treści
- Dzięki indeksowaniu wyszukiwanie jest szybsze i skuteczniejsze
- Na ocenę strony w kontekście zapytania wpływa wiele czynników
- Search Console pomaga zoptymalizować indeksowanie witryny
- Indexing API przyspiesza aktualizację Indexu Google



Dziękujemy za uwagę :)

