

Indeksowanie stron internetowych przez Google

Michał Cichocki

Adrian Cygan

Szymon Czajkowski

semestr 7, Informatyka, Wydział Informatyki i Telekomunikacji

Spis treści

- Proces indeksowania
- Jak działa Google?
- Przeszukiwanie
- Indeksowanie
- Google Search Console
- Ranking
- Wykorzystywane algorytmy
- Czynniki wpływające na indeksowanie
- Podsumowanie

Proces indeksowania

Na czym polega?

**Dlaczego jest kluczowy dla
widoczności wyszukiwania?**

Roboty sieciowe

Jak działa Google?

Przeszukiwanie

Indeksowanie

Ranking

Jak działa Google?

Przeszukiwanie



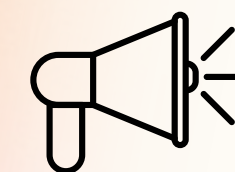
Zbieranie informacji poprzez Google Boty

Boty wchodzą w interakcję ze stroną jak zwykli użytkownicy



Zasady działania Google botów

Na co zwracają uwagę Google Boty?



Plik robots.txt

Wskazuje zakres strony jaki jest dostępny dla botów

pk.edu.pl/robots.txt

```
User-agent: *  
Disallow: /administrator/  
Disallow: /bin/  
Disallow: /cache/  
Disallow: /cli/  
Disallow: /components/  
Disallow: /includes/  
Disallow: /installation/  
Disallow: /language/  
Disallow: /layouts/  
Disallow: /libraries/  
Disallow: /logs/  
Disallow: /modules/  
Disallow: /plugins/  
Disallow: /tmp/
```

Jak działa Google?

Indeksowanie



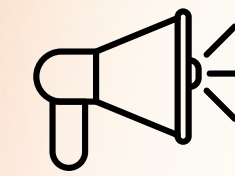
Analiza i przechowywanie w indeksie – bazie danych

Badana zawartość podlega ocenie na podstawie struktury i wartości użytkowej



Właściwości indeksu wyszukiwarki Google

Rozmiar bazy i nauka na nieustających zmianach



Zakres wiedzy indeksu Google

Informacje pochodzące z książek oraz z tzw. grafu wiedzy

Google Search Console

Narzędzie ułatwiające indeksowanie treści

- Jak Google widzi i indeksuje stronę?
- Wyświetlenie danych związanych z ruchem
- Linki do strony w innych witrynach
- Raportowanie i rozwiązywanie problemów związanych z indeksowaniem

Ranking - uporządkowanie pytania do odpowiedzi

Znaczenie

- Rozpoznawanie intencji słów, rozumienie rodzaju poszukiwanych informacji

Trafność

- Czy treści zawierają odpowiedź na pytanie użytkownika

Jakość

- Nadawanie priorytetu na podstawie jakości

Użyteczność

- Dane techniczne

Kontekst

- Informacje o lokalizacji, historii wyszukiwania, oraz ustawień

Wykorzystywane algorytmy



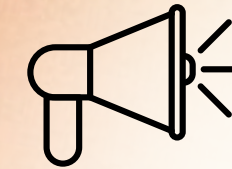
RankBrain

Pierwszy system uczenia głębokiego wykorzystywany przez wyszukiwarkę Google. Kojarzy słowa z ideą jaka za nimi idzie. Nadal używany.



PageRank

Nadaje stronom internetowym wartość od 0 do 10. Jakość tekstu jest proporcjonalna do tego ile tekstów się na niego powołuje.



BERT

Model językowy stworzony w 2018 roku. Analizuje zadane pytania w całym ich kontekście, a nie słowo po słowie.

Czynniki wpływające na indeksowanie

Poprawny kod HTML

Strona powinna być napisana zgodnie z semantyką i zasadami języka HTML.

Responsywność

Strona powinna być przystosowana do urządzeń mobilnych.

Czas ładowania

Wolno ładujące się strony mogą być trudniejsze do pełnego przeszukania.

Pytania i Odpowiedzi

Dziękujemy za uwagę